

**Analysis of Consumer Research Evidence on the
Impact of Plain Packaging for Tobacco Products**

(Updated to 2014)

Prepared by:

Professor Timothy M. Devinney

3 January 2014

Table of Contents

1. Report Introduction	3
2. Criteria for Evaluation.....	5
3. Selection of the Studies	19
4. Review of the Studies.....	19
Classification	19
Detailed analysis of the Studies.....	26
Pechey et al. (2013).....	26
Wakefield et al. (2013).....	34
Moodie and Mackintosh (2013).....	42
Hammond et al. (2013).....	47
Hammond, White et al. (2013)	52
Borland and Savvas (2013).....	58
Ramunno et al. (2012).....	64
Ford, et al. (2013).....	66
Dewe et al. (2013).....	73
Lund and Scheffels (2013)	77
Scheffels and Lund (2013)	82
Focus Group Studies	88
‘Public Support’ Studies.....	97
5. Conclusion	101
6. Exhibit One – Qualifications of Professor Devinney.....	115
7. Exhibit Two – Resume	117
(a) Examples of Research Support Received	122
(b) Professional Consulting Experience (Selection)	124
8. Exhibit Three – Selected Publications	125
9. Exhibit Four – Other Materials Considered	131
Studies which are related to plain packaging as a regulatory initiative but which do not generate any consumer survey evidence in respect of plain packaging	131
Studies which appear to be related to the issue of plain packaging as a regulatory initiative, but for which survey analysis does not appear to be publicly available	132

1. REPORT INTRODUCTION

1.1 My full name is Professor Timothy M. Devinney. I am a University Leadership Chair at the University of Leeds in the UK. In addition, I am a Conjoint Professor in the Faculty of Medicine at the University of New South Wales, Australia and a Visiting Professor at the Institute of Management at Humboldt, Berlin. I am an academic trained in the areas of Psychology, Public Policy, Economics, Statistics and Management. I have extensive experience in the conduct and evaluation of consumer research studies, both from an academic and commercial perspective. I have been involved most recently in an extensive set of research projects examining the degree to which social aspects of consumption influence behaviour.

1.2 Exhibits One to Three of this report set 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 am an expert in consumer survey research, experimental methods and associated statistical analysis. I have specialised knowledge in assessing the methodology of consumer survey research to determine the extent to which it provides credible, methodologically and empirically sound evidence (which I refer to in this report as “reliable evidence”) in support of stated conclusions. Although not limited to this area, I have extensive experience with these issues in the context of consumer goods. I have been requested to prepare this report for Japan Tobacco International (“JTI”) (I describe in further detail below the basis on which I have prepared this report).

1.3 I am the author of the following reports on plain packaging previously prepared at the request of JTI:

- (a) “Analysis of Consumer Research Evidence on the Impact of Plain Packaging for Tobacco Products” dated 30 November 2010 (the “2010 Report”); and

- (b) “Analysis of Consumer Research Evidence on the Impact of Plain Packaging for Tobacco Products” (Updated to 2012) dated 29 June 2012 (the “2012 Report”).

1.4 In the 2010 and 2012 Reports, I examined the publicly available consumer surveys and experiments purporting to be in support of plain packaging and concluded that they do not provide reliable evidence that plain packaging would be effective in changing actual smoking behaviour.

1.5 Since completing my 2012 Report, further studies purporting to be in support of plain packaging have been published or have been otherwise made available for review (which I refer to below as “the Studies”). I have been asked to review the Studies using the same criteria employed in my 2010 and 2012 Reports,¹ to address the extent to which it provides reliable evidence that plain packaging would be effective in achieving the public policy goals (identified by various regulators) of changing actual smoking behaviour, namely in:

- (a) reducing smoking uptake (also known as smoking initiation) among minors;
- (b) reducing smoking among minors and/or adults; or
- (c) increasing smoking cessation among minors and/or adults.

1.6 In summary, and as discussed further in Sections 4 and 5 of this report, it is my expert opinion that none of the Studies, including those I reviewed in my 2010 and 2012 Reports, provides reliable evidence that plain packaging would be effective in achieving the public policy goals of changing actual smoking behaviour as set out in paragraph 1.5 above. In short, there is not a single study which I have reviewed in this report, or in my 2010 and 2012 Report, which provides a “line of sight” between their

¹ See Section 2 of my 2010 and 2012 Reports.

attitudinal and perceptual measures and actual purchasing behaviour, nor any of the measures outlined in paragraph 1.5 above.

1.7 Insofar as the body of research as a whole, and while there are now a number of studies purporting to determine the potential impact of plain packaging in achieving these goals, a close examination of them reveals that: (i) the body of research has not examined any specific aspect of market related behaviour directly, instead relying exclusively on attitudes, opinions and self reports of possible behaviours; (ii) the body of research has continued to apply the same flawed methodological approaches over and over again; and (iii) the body of research has failed to take into account important related research that could improve and inform both the scientific and policy debate.

1.8 For ease of exposition, I repeat the evaluation criteria employed in my 2010 and 2012 Reports in this report in Section 2 below.

2. CRITERIA FOR EVALUATION

2.1 In evaluating the Studies a number of factors come into play in determining their validity with respect to addressing the questions being set. As in my 2010 and 2012 Reports, I build on the criteria outlined by Dr. Keegan.² For ease of exposition, I repeat the outline I presented in my 2010 and 2012 Reports as way of a summary of the criteria discussed in detail by

² As discussed in paragraph 1.3 of my 2010 and 2012 Reports, I have reviewed the following documents prepared by Dr Warren Keegan, which review, amongst other things, publicly available consumer survey evidence cited in support of plain packaging for tobacco products: (i) “Analysis of Consumer Survey Evidence Relevant to the UK Department of Health Consultation on the Future of Tobacco Control” dated 2 September 2008; (ii) “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; (iii) “Analysis of Consumer Survey Evidence Relevant to DG SANCO’s Proposal to Increase the Size of Health Warnings on Tobacco Packaging” dated 24 November 2010. In addition, I have also reviewed the document prepared by Dr Keegan entitled “Analysis of Consumer Survey Evidence Relevant to the Display Ban Requirement in England” dated 28 April 2010. In this report, I refer to these four documents collectively as “the Reports”.

Dr Keegan. This table also provides a brief definitional description of the criteria.

2.2 I would also note that in line with my 2010 and 2012 Reports and Dr Keegan's Reports, I have applied the same professional standards for researchers as outlined in the various codes of conduct from groups such as the UK-based Market Research Society (MRS),³ the American Association for Public Opinion Research (AAPOR),⁴ the World Association for Public Opinion Research (WAPOR),⁵ the Council of American Research Organizations (CASRO),⁶ the International Statistical Institute (ISI),⁷ the Canadian Market Research and Intelligence Association (MRIA),⁸ and the European Society for Opinion and Marketing Research (ESOMAR).⁹

2.3 Because many of the plain packaging studies utilise experimental or quasi-experimental approaches to consumer research, I believe that a number of additional criteria are relevant. The relevance of these additional criteria is due to the nature of the proposals for the introduction of plain packaging. These criteria were outlined in my 2010 and 2012 Reports and are repeated here for ease of exposition. Specifically:

- (a) Because plain packaging has only recently been adopted in Australia, and does not exist anywhere else in the world, and consumers are being asked to state an 'intention' relating to purchasing

³ Market Research Society (2005), *MRS Code of Conduct*. <http://www.mrs.org.uk/standards/downloads/code2005.pdf>.

⁴ American Association for Public Opinion Research (2005), *AAPOR Code of Professional Ethics & Practices*. <http://www.aapor.org/aaporcodeofethics>.

⁵ World Association for Public Opinion Research, *WAPOR Code of Professional Ethics and Practices*. <http://www.unl.edu/WAPOR/ethics.html>.

⁶ Council of American Research Organizations (2008), *Code of Ethics and Standards for Survey Research*. <http://www.casro.org/pdfs/CodeVertical-FINAL.pdf>.

⁷ International Statistical Institute (1985), *Declaration on Professional Ethics*. <http://isi.cbs.nl/ethics.htm>.

⁸ Market Research and Intelligence Association (2007), *Code of Conduct for Members*. <http://www.mria-arim.ca/STANDARDS/CODE2007.asp>.

⁹ European Society for Opinion and Marketing Research, *ICC/ESOMAR International Code on Market and Social Research*. http://194.38.169.84/uploads/pdf/professional-standards/ICESOMAR_Code_English_.pdf.

circumstances that are not available in most countries, they are being forced to speculate about a specific behaviour that may or may not arise in new circumstances. Hence, researchers must be cognisant of the degree to which the experimental task creates outcomes that can be linked specifically to behaviour.

- (b) Because plain packaging has embedded within it a social outcome – i.e., the sheer existence of the unbranded packaging is based on the belief that it will make a product category less attractive and therefore will change actual smoking behaviour – individuals will most likely know the intent of the investigator. Hence, the researcher must be cognisant of the degree to which the study itself enhances artificially the salience of the factors being studied.
- (c) Ultimately, the goal of policy related research is to examine the efficacy of a change in policy on a change in behaviour. Changes in behaviour themselves require the affected individuals to either use different decision models or different criteria within their existing models. Hence, it is critical for researchers to have as complete an understanding as possible of the parameters of consumers' decision models.

2.4 Points (a) and (b) in paragraph 2.3 relate specifically to what is known as the attitude-behaviour gap or the difference between 'stated' intentions and 'revealed' or actual purchases.¹⁰ Good research will attempt to

¹⁰ This is an extensive area of research. Some representative publications include: Sheerana, P. (2002), Intention-Behaviour Relations: A Conceptual and Empirical Review, *European Review of Social Psychology*, 12: 1-36; Carrigan, M. and A. Attalla, (2001), The Myth of the Ethical Consumer – Do Ethics Matter in Purchase Behaviour? *Journal of Consumer Marketing*, 18: 560-578; Ajzen, I. and M. Fishbein (2002), The Influence of Attitudes on Behaviour, in Albarracín, D. Johnson, B. and M. Zanna (Eds.), *The Handbook of Attitudes*, Abingdon, UK: Routledge, 173-221.

reduce this problem by focusing on three factors, which I will add to my list of criteria:¹¹

(a) **Incentive compatibility.** Incentive compatibility addresses the extent to which the methodology used by the researcher allows (or makes) subjects to reveal their true behaviour (if they currently engage in an activity, such as purchasing or not purchasing a product that is currently available) or what that behaviour would be if they were given the opportunity (in situations where there is no opportunity to reveal that behaviour, such as in the case of a new product). For example, it is well understood that forcing individuals to make a trade-off or asking them to pay a real price creates incentives that align better with their actual purchasing behaviour. Incentive compatibility is a particularly critical issue when asking:

- a) ‘intention’ questions – e.g., “if faced with these alternatives, which would you choose?”;
- b) ‘speculative’ questions – e.g., “how do you think a person faced with these alternatives would behave?”; and
- c) most types of ‘self-report’ questions – e.g.’ “how likely is it that you do/would engage in a behaviour?”.

(b) **Inference of salience.** Inference of salience addresses the degree to which the sheer addition of a factor that would otherwise not be part of the consumer’s decision is all of a sudden added into the mix. For example, asking individuals about newly added attributes to existing products – i.e., aspects of a product that consumers know do not currently exist in what is offered in the market – heightens the

¹¹ An overview of the importance of these issues is given in Devinney, T., Auger, P. and G Eckhardt (2010), *The Myth of the Ethical Consumer*, Cambridge UK: Cambridge University Press, pp. 56-59.

salience of the new information making it more likely that the consumers will over-react to the new aspects of the product.

- (c) **Context.** Context addresses the degree to which the decision individuals are being asked to make is outside the context in which it might normally be made. For example, it is quite common to find that individuals, when asked in a survey the degree to which they will act in a pro-social way – such as purchasing ‘green’ products, volunteering or donating to a charitable cause – will overstate very significantly the likelihood that they will do so. Part of this is related to salience and incentive compatibility but it is also the case that most social behaviours are context driven, meaning that it is the context that drives behaviour. The context of answering something in a survey is different from the context of opening one’s wallet or sacrificing one’s time.

2.5 Point (a) in paragraph 2.3 forces us to ask very specifically how an individual chooses to do (or not do) something. Given the policy goals outlined in paragraph 1.5 we can reframe this to read:

- (a) Do the Studies have an effective statistical decision model of smoking uptake (initiation) among minors?
- (b) Do the Studies have an effective statistical decision model that explains a reduction in smoking consumption among minors and/or adults?
- (c) Do the Studies have an effective statistical decision model of smoking cessation among minors and/or adults?

2.6 What these questions ask collectively is: do the Studies effectively tell us something about the cognitive process that people go through when making a decision with regard to smoking-related behaviour(s)?

2.7 The point of this criterion is the link between the structure and design of the study and the operative criteria that a consumer would be using in realistic purchasing circumstances. In other words, does the study **appropriately model decision making** when the decisions are smoking initiation, smoking reduction and smoking cessation?

2.8 The experimental studies examined in this report were attempting to determine what decision an individual would make in “what if” circumstances. What this implies is that studies that can most effectively mimic the decision making process and criteria used by the individual in realistic circumstances will be the most valid. Hence, an additional criterion that we need to consider is the degree to which the experimental approach was designed in a manner that allowed the researcher to model the decision making process consumers would be using in realistic purchasing environments.

2.9 When attempting to determine what an individual’s decision model is via the choices that they can make in experiments it is important that the decision model and the experimental structure are aligned. The most accepted method for doing this is via the application of what are known as discrete choice models (or its variant conjoint analysis). This approach requires that individuals: (a) choose amongst a set of alternatives, (b) rank a set of alternatives, and/or (c) rate a set of alternatives.

2.10 The set of alternatives – known as a ‘choice set’ – must be structured according to an experimental design that has certain specific properties that we also consider as part of our evaluative criteria.

(a) First, they must be **statistically efficient**. Statistical efficiency implies that the structure of the experiment allows the researchers to recreate the decision model in use, either by the individual or by a

group of individuals.¹² Statistical efficiency is important because most experiments cannot possibly have all individuals look at every possible combination of a product's features that might be potentially on offer in a market. Hence, the extent to which an experiment is efficient is the degree to which the choice sets allow the researcher to have confidence that they have enough information to say that their results are a realistic representation of how an individual or group of individuals would behave.

- (b) Second, it is important when looking at the choices that individuals are being asked to consider that the attributes and features presented exhibit (as closely as possible) **orthogonality**.¹³ In product choice experiments the features of the product are broken into 'attributes' – such as brand, colour or price – and the attributes into 'levels'. For example, the 'levels' of the attribute *price* might be €1, €3, €5, €7 and the attributes of *colour* might be red, green and blue. Orthogonality implies that the experiment examining *price* x *colour* be set up with $4 \times 3 = 12$ alternatives as shown in Table 1 and that the individual be presented with a series of price x colour alternatives – as represented by the cells A-L below – such that the effects of each combination can be evaluated independently.¹⁴ This is achieved by the fact that

¹² Mathematically, efficiency is a comparison of the design used to the 'optimal' or best possible design (the one that ensures that the decision model is estimated with the greatest degree of statistical precision). See, e.g., Street, D. and L. Burgess (2007). *The Construction of Optimal Stated Choice Experiments*, Hoboken, NJ: Wiley, 85-86.

¹³ Perfect orthogonality is sometimes difficult to achieve and there are classes of non-orthogonal experimental designs. However, these are normally reserved for quite complex experiments and none of the Studies examined here would, in my expert opinion, be classified as being so complex as to require anything other than a basic experimental approach. See, e.g., Kuhfeld, W., Tobias, R. and M. Garrett (1994), Efficient Experimental Design with Marketing Research Applications, *J. of Marketing Research*, 31: 545-557.

¹⁴ More specifically this is important when using linear regression models for estimation purposes. As noted by Kuhfeld W., Tobias, R., and M. Garrett (1994), "a linear model is fit with an orthogonal design, the parameter estimates are uncorrelated, which means each estimate is independent of the other terms in the model. More importantly, orthogonality, usually implies that the coefficients will have minimum variance" (p. 545). The first part of this statement implies that we can readily make general statements about the importance of an attribute independent of the other attributes. For example, we can talk of the effects of

any combination of an attribute level in the experiment below will appear 1/12th of the time and Red, Green and Blue will appear in 1/3rd of the options presented.¹⁵ A lack of orthogonality makes the determination of the effect of specific attributes and levels difficult and will also generally imply less than efficient statistical estimation.

		Colour		
		Red	Green	Blue
Price	€1	A	B	C
	€3	D	E	F
	€5	G	H	I
	€7	J	K	L

- (c) Third, it is important when looking at the choices that individuals are being asked to consider that the experiment exhibits **balance**.¹⁶ This means that the attribute levels are appearing across the choice sets an equal number of times. For example, in Table 1, it is possible that each individual would see the four prices two or three times each. However, an improper design would have them see €1 three times and €3, €5, and €7 only once. The lack of balance has two possible effects. First, the subject may notice the imbalance and focus on the

price independent of the effects of the different colours. The second part of the statement implies that the estimates of that importance are the ‘best’ estimates we can achieve.

¹⁵ Technically, orthogonality is achieved when “the joint occurrence of any two levels of different attributes appear in options with frequencies equal to the product of their marginal frequencies”. Hubert, I. and K. Zwerina (1996), The Importance of Utility Balance in Efficient Choice Designs, *J. Marketing Research*, 33: 307-317.

¹⁶ Haaijer, R. and M. Wedel (2007), Conjoint Choice Experiments: General Characteristics and Alternative Model Specifications, in Gustafsson, A., Herrmann, A. and F. Huber (Eds), *Conjoint Measurement: Methods and Applications*, 4th Edition, Heidelberg, Germany: Springer. Note that sometimes balance and orthogonality are in conflict, particularly when some of the options presented might not make logical sense to the subject of the experiment. Such an example might be a car that had low price and lots of luxuries or very high engine power but also very high fuel efficiency. However, this again is not an issue relevant to the Studies being examined in this report as there is no indication that the authors considered such factors when setting up their studies. In the case where such an issue exists, the research would generally use a non-orthogonal design of the types discussed by Kuhfeld, Tobias and Garrett (1994).

levels (in this case the prices) that appear more/less frequently (we do not know in which direction the bias would go, just that it exists). Second, statistically we will have inefficient estimates of the effects of the levels because we have only one response for three of the prices and three responses for one of the prices.

2.11 Point (a) in paragraph 2.3 relates also to intentionality. Morowitz, Steckel and Gupta¹⁷ have shown that ‘intentions’ are most related to actual purchasing when:

- (a) They are for **existing products**;
- (b) They are for durable rather than **non-durable goods**;
- (c) They are for **short-term horizon** decisions rather than for long-term time horizon decisions;
- (d) Subjects are asked about purchase intentions for **specific brands** rather than for the product category in general;
- (e) Purchase intentions are measured as **‘trial’ rates amongst existing purchasers in the relevant product segment**, rather than being measured in terms of total market share; and
- (f) Purchase intentions are collected in a **comparative mode**, rather than monadically (e.g., a paired comparison versus asking the subject to evaluate a single alternative at a time).

2.12 As noted in paragraph 2.1 above, this report includes the evaluation criteria applied by Dr. Keegan in his Reports. Table 2 provides a short summary of the criteria applied by Dr Keegan and highlights which are most relevant to this report. Those criteria that are not applied here are excluded

¹⁷ Morowitz, V., Steckel, J. and A. Gupta (2007), When Do Purchase Intentions Predict Sales? *International Journal of Forecasting*, 23: 347-364.

only because they are not applicable to the research studies examined below. Specifically, the Studies are: (a) all relatively recent and hence do not suffer from ‘study age’ issues; and (b) are examinations within a single culture/country and hence do not suffer from issues of cross cultural bias.

Table 2: A Summary of the Evaluative Criteria Outlined by Dr. Keegan

Short Description of Criteria and Its Relevance	
<i>1. Standards Compliance</i>	
Compliance with International Standards	Complies with the standards outlined in paragraph 2.2 above.
<i>2. Age of Study</i>	
Study age	Consumer research has a limited lifespan; the older the study, the less likely its current applicability. The study must be reflective of current market conditions and regulatory environment. This criterion is not applied to the Studies in this report.
<i>3. Field Administration Protocol</i>	
Question design	Ensuring proper question design is a requirement that is reflected across internationally accepted research standards. Questions should not cue responses, i.e., ‘beg the answer’. Questions should not make assumptions about a respondent’s knowledge or experiences. Respondents should be given the

Short Description of Criteria and Its Relevance

	<p>opportunity to give a “don’t know” or “no opinion” answer.</p>
Interviewer response bias	<p>To the extent that it may bias the results, neither respondents nor persons responsible for the data collection should be informed as to the sponsor or purpose of the study.</p>
Researcher objectivity	<p>A researcher, whatever his/her views or opinions on a topic, must ensure that the study design is impartial and not designed to yield any particular result.</p> <p>To the extent that an author’s advocacy influences the study design, the study’s reliability and validity suffers.</p>
Response reliability	<p>Observing what people do is a better predictor of behaviour than recording how people respond to questions about what they think they will do, or what they think others will do, or what they report they have done.</p> <p>In consumer research, the gold standard is to get as close as one can to observing behaviour. The gradient of research reliability, from most reliable to least reliable, is generally as follows:</p>

Short Description of Criteria and Its Relevance

Reliability by Data Collection Method

Reliability	Data Collection Method	Research Type
Most reliable	Direct observation	Observed Behavioural
↓	Recent recall of behaviour	Observed Behavioural
	Recall of non-recent past behaviour	Self-Reported Behavioural
	Prediction of future behaviour	Opinion / Attitudinal
Least reliable	Prediction of others' future behaviour	Opinion / Attitudinal

4. Appropriateness of Sampling Frame

General appropriateness The sample should reflect the population relevant to the question at hand allowing for the greatest degree of generalisation.

Age of respondents Conducting research among minors presents particular issues that must be accounted for to ensure the reliability of the data collected.¹⁸ Young respondents are more likely to feel pressured during an interview situation; such pressure can result in answers that are inaccurate.

It is much more difficult to ask a minor a difficult policy question and have an acceptable degree of confidence that the information collected will have any resemblance to the effect that would be

¹⁸ Churchill, G. and D. Iacobucci (2005), *Marketing Research: Methodological Foundations, 9th Edition*. Orlando, FL: Dryden, 387-390.

Short Description of Criteria and Its Relevance

observed if the policy were actually enacted. For example, asking a minor “*Will young people buy fewer bus passes if fares are increased?*” is unlikely to generate reliable data.

Focus groups

Focus group studies are exploratory. They generate hypotheses rather than findings that can be generalised to a wider population.¹⁹

The reported findings of focus groups often have no statistical significance due to the small sample size and informal nature of the responses.

5. Analysis

Statistical significance

It is imperative that authors refrain from projecting results that are not statistically significant to general populations or markets.²⁰

It is widely recognised in the research community that statistical significance is a necessary pre-requisite in determining that a causal relationship is an observed result

¹⁹ Churchill, G. and D. Iacobucci (2005), *Marketing Research: Methodological Foundations, 9th Edition*. Orlando, FL: Dryden, 81-85.

²⁰ International Statistical Institute (1985), *Declaration on Professional Ethics*. <http://isi.cbs.nl/ethics.htm>, paragraphs 1.3 and 3.1.

²¹ Thompson, B. (1994), The Concept of Statistical Significance Testing, *Practical Assessment, Research & Evaluation*, 4.5.

Short Description of Criteria and Its Relevance

and not caused by chance, error or other factors.²¹

Unsupported results or conclusions

In interpreting study results, authors sometimes make ‘leaps’ between the data yielded by the study and the conclusion the author puts forth. It is concerning when an author draws conclusions that are not supported by the research.²²

False comparison

It is imperative that authors refrain from generating comparisons and drawing conclusions from comparisons that are not reflective of actual real life conditions.

Cross cultural applicability

Cultures can have unique characteristics that must be accounted for when designing a study, and specifically, a questionnaire. Cultural differences exist both between countries and, indeed, in many cases, within different geographical regions of one country.

This criterion is not applied to the Studies in this report.

²² American Association for Public Opinion Research (2005), *AAPOR Code of Professional Ethics & Practices*. <http://www.aapor.org/aaporcodeofethics>, paragraph I.A.2 and I.A.3.

3. SELECTION OF THE STUDIES

3.1 I have based my opinions and conclusions in this report on studies that presented original consumer research, and which purported to be in support of plain packaging.²³ Studies that did not generate or evaluate consumer evidence relating to the effectiveness of plain packaging in achieving the public policy goals given in paragraph 1.5 above were not considered in the formulation of my opinions.

3.2 The other documents I have considered in reaching my conclusions on the issues addressed in this report are listed in full in chronological order in Exhibit Four.

3.3 I have sought to identify all potentially relevant materials using the resources available to me. I have conducted the most objective review possible in accordance with the international research standards outlined.

4. REVIEW OF THE STUDIES

CLASSIFICATION

4.1 For ease of exposition, I set out below the studies which warranted a full review in my 2010 and 2012 Reports:

- (a) Hammond, D., Dockrell, M., Arnott, D., Lee, A. and A. McNeill (2009). Cigarette pack design and perceptions of risk among UK

²³ In this regard, I have also reviewed a study entitled “An analysis of smoking prevalence in Australia” by London Economics, dated November 2013. This study does not fall within my review parameters, as it does not purport to be in support of plain packaging. Accordingly, it is not necessary for me to review it in detail. However, I note the following three points. First, the data is said not to demonstrate that there has been a change in smoking prevalence in Australia among the adult resident population (youth smoking prevalence is not considered at all) caused by the introduction of plain packaging in December 2012. Second, the authors say that “*it is important to note it is not possible to assign a causal relationship between the changes in noticeability of health warnings or smoking prevalence and the introduction of plain packaging*”. Finally, the authors note that “*as more robust information becomes available, a range of econometric techniques could be adopted to disentangle the relative effect of plain packaging and larger health warnings on smoking behaviour*”.

adults and youth. *The European Journal of Public Health*, Volume 19, Issue 6, 631-637.

- (b) Hammond, D. and C. Parkinson (2009). The impact of cigarette package design on perceptions of risk. *The Journal of Public Health*, Sep; 31 (3): 345-53.
- (c) Moodie, C. and G. Hastings (2009). Making the pack the hero, tobacco industry response to marketing restrictions in the UK: Findings from a long-term audit. *International Journal of Mental Health Addiction*, Volume 9, Number 1; 24-38.
- (d) Germain, D., Wakefield, M. and S. Durkin (2009). Adolescent's perceptions of cigarette brand image: Does plain packaging make a difference? *Journal of Adolescent Health*, Volume 46, Issue 4; 385-392.
- (e) Doxey, J. (2009). Deadly in Pink: The impact of female-oriented cigarette packaging on brand appeal, beliefs about smoking, and risk perceptions among young women. Unpublished MS Dissertation. University of Waterloo.
- (f) Bansal-Travers, M., Hammond, D., Smith, P. and K. Cummings (ND). The impact of cigarette pack design, descriptors and warning labels on risk perceptions in the US. Unpublished Working Paper, Roswell Park Cancer Institute
- (g) Gendall, P., Hoek, J., Thomson, G., Edwards, R., Pene, G., Gifford, H., Pirikahu, G. and J. McCool (2011). Young adults' interpretations of tobacco brands: Implications for tobacco control. *Nicotine & Tobacco Research*, Advance Publication.

- (h) Moodie, C., Mackintosh, A., Hastings, G. and A. Ford (2011). Young adult smokers' perceptions of plain packaging: a naturalistic study. *Tobacco Control*. Vol. 20, 367-373.
- (i) Hammond, D., Doxey, J., Bansal-Travers, M. et al. (2011). Impact of Female-Oriented Packaging in the United States. *Nicotine & Tobacco Research*, E-Pub ahead of print, 13 April 2011.
- (j) Munafo, M., Roberts, N., Bauld, L. and U. Leonards (2011). Plain packaging increases visual attention to health warnings on cigarette packs in non-smokers and weekly smokers but not daily smokers. *Addiction*, Advance Publication.²⁴
- (k) Moodie, C., Hastings, G. and L. Joossens (2011). Young adult smokers' perception of illicit tobacco and the possible impact of plain packaging on purchase behaviour. *European Journal of Public Health*, Advance Publication.
- (l) Thrasher, J., Rousu, M., Hammond, D. and A. Navarro (2011). Estimating the impact of pictorial health warnings and "plain" cigarette packaging: Evidence from experimental auctions among adult smokers in the United States. *Health Policy*, Advanced Publication.
- (m) Gallopel-Morvan, K., Moodie, C., Hammond, D. et al. (2011). Consumer perceptions of cigarette pack design in France: a

²⁴ I note that another variant of this study has been recently published entitled "Visual Attention to Health Warnings on Plain Tobacco Packaging in Adolescent Smokers and Non-Smokers", Maynard, O.M., Munafo, M.R. and U. Leonards (2013). While the samples in the two studies are different (the original study sampled 43 'young adults' at the University of Bristol whereas the later study sampled 87 adolescents in three comprehensive secondary schools in Bristol) the methodology utilised in both studies is the same. I, therefore, repeat the conclusion I reached at paragraph 4.37 of my 2012 Report where I stated that "*the lack of a direct behavioural measure of usage leads me to conclude that it does not provide reliable evidence as to whether plain packaging would help achieve the public policy goals listed in paragraph 1.5 above*". It is also noted that in the case of this study there was an additional problem of a rather significant (14 of the original 101) subject attrition.

comparison of regular, limited edition and plain packaging. *Tobacco Control*, Online First, published on October 13, 2011.

- (n) Borland, R., Savvas, S., Sharkie, F. and K. Moore (2011). The impact of structural packaging design on young adult smokers' perceptions of tobacco products. *Tobacco Control*, Online First, published on December 13, 2011.
- (o) Moodie, C., Ford, A., Mackintosh, A. and G. Hastings (2011). Young people's perceptions of cigarette packaging and plain packaging: An online survey. *Nicotine & Tobacco Research*, Advance Access, published October 24, 2011.
- (p) M. Al-Hamdani, (2011). A survey on the effects of progressive removal of brand imagery elements from cigarette packs on the perceptions of adult university students. Unpublished Master of Health Administration Thesis, Dalhousie University, August 2011.²⁵
- (q) Moodie, C. and A. Ford (2011). Young adult smokers' perceptions of cigarette pack innovation, pack colour and plain packaging. *Australasian Marketing Journal*, 19: 174-180.
- (r) Hoek, J., Gendall, P., et al. (2012). Tobacco branding, plain packaging, pictorial warnings, and symbolic consumption. *Qualitative Health Research*, 22 (5): 630-639.

²⁵ This study was ultimately published in a slightly different form with an update analysis (Al-Hamdani, M. (2013) "The Effect of Cigarette Plain Packaging on Individuals' Health Warning Recall", *Health Care Policy*, 8 (3): 68-77). I reviewed the original research in paragraphs 4.96 – 4.107 of my 2012 Report. I concluded in that case that "Overall, this study has a number of serious methodological flaws, some of which no doubt arise from the nature of the purpose of the work itself. As discussed, the author's analysis is quite poorly executed and there are reasons to be concerned about the quality of the measurement, the sophistication of the analysis and the nature of the sampling. It would be imprudent to consider the results of the study as a basis of any effective policy conclusions given these limitations." Despite the updated analysis in Al-Hamdani (2013), the fundamental methodological flaws remain, and the validity of the study as a guide to policy remains suspect. Because of this, a full review of the published version of the work would be redundant to what I have already concluded in my 2012 Report.

- (s) McCool, J., Webb, L., Cameron, L. and J. Hoek (2012). Graphic warning labels on plain cigarette packs: Will they make a difference to adolescents? *Social Science & Medicine*, doi:10.1016/j.soscimed.2011.12.043.
- (t) Wakefield, M., Germain, L., Durkin, S., Hammond, D., Goldberg, M., and R. Borland (2012). Do larger pictorial health warnings diminish the need for plain packaging of cigarettes? *Addiction* (2012), doi:10.1111/j.1360-0443.2012.03774.
- (u) Cancer Research UK (2012). The packaging of tobacco products. The Centre for Tobacco Control Research.
- (v) Carter, S. and S. Chapman (2006). Smokers and non-smokers talk about regulatory options in tobacco control. *Tobacco Control* 2006; 15: 398-404. Hereinafter referred to as “Carter & Chapman (2006).”²⁶
- (w) Parr V., Tan B., Ell P. and K. Miller (2011). Market research to determine effective plain packaging of tobacco products. GfK Blue Moon, Sydney. Hereinafter referred to as “GfK Bluemoon (2011).”

4.2 My review in this report has focused on key primary research studies published or otherwise obtained since my 2010 and 2012 Reports that warrant a full review to determine whether they contain reliable evidence that plain packaging will achieve the public policy goals set out at paragraph 1.5 above. The Studies are listed below, with a study-by study analysis set out from paragraph 4.3 below.

- (a) Pechey, R., Spiegelhalter, D. and T.M. Marteau (2013). Impact of plain packaging of tobacco products on smoking in adults and

²⁶ As I discussed in my 2012 Report, this study is not specifically related to plain packaging, however, plain packaging was one alternative considered in the 11 options discussed. In addition, I note that this study was suggested by some as providing evidence in support of plain packaging (see, for example, “Plain Tobacco Packaging: A Systematic Review”, Moodie, C., et al. (2012), University of Stirling, Public Health Research Consortium).

children: an elicitation of international experts' estimates. Hereinafter referred to as "Pechey et al. (2013)."

- (b) Wakefield, M., Hayes, L., Durkin, S. and R. Borland (2013). Introduction effects of the Australian plain packaging policy on adult smokers: a cross-sectional study. *BMJ Open* (2013). Hereinafter referred to as "Wakefield et al. (2013)."
- (c) Moodie C. and A.M. Mackintosh (2013). Young adult women smokers' response to using plain cigarette packaging: a naturalistic approach. *BMJ Open* (2013). Hereinafter referred to as "Moodie and Mackintosh (2013)."
- (d) Hammond, D., Daniel, S. and C.M. White (2013). The effect of cigarette branding and plain packaging on female youth in the United Kingdom. *Journal of Adolescent Health*, Volume 52, (2013), 151-157. Hereinafter referred to as "Hammond et al. (2013)."
- (e) Hammond, D., White, C., Anderson, W., Arnott, D. and M. Dockrell (2013). The perceptions of UK youth of branded and standardized, 'plain' cigarette packaging. *European Journal of Public Health, Advance Access, published October 8, 2013*. Hereinafter referred to as "Hammond, White et al. (2013)."
- (f) Borland, R. and S. Savvas, (2013). The effects of variant descriptors on the potential effectiveness of plain packaging. *Tobacco Control*, Online First, published on February 22, 2013. Hereinafter referred to as "Borland and Savvas (2013)."
- (g) Ramunno, S., Mandeville, K.L. and K. Yarrow (2012). The effect of cigarette packaging on attention to health warnings, *The Lancet*, Volume 380, Page S5, 23 November 2012. Hereinafter referred to as "Ramunno et al. (2012)."

- (h) Ford, A., Mackintosh, A.M., Moodie, C., Richardson, S. and G. Hastings (2013). Cigarette pack design and adolescent smoking susceptibility: a cross-sectional survey. *BMJ Open* (2013). Hereinafter referred to as “Ford et al. (2013).”
- (i) Ford, A., Moodie, C., Mackintosh, A.M. and G. Hastings (2013). How adolescents perceive cigarette packaging and possible benefits of plain packaging. *Education and Health*, Volume 31, Number 2, 2013. Hereinafter referred to as “Ford, Moodie, Mackintosh et al. (2013).”
- (j) Ford, A., Moodie, C., Mackintosh, A.M. and G. Hastings (2013). Adolescent perceptions of cigarette appearance. *European Journal of Public Health* 1-5. Hereinafter referred to as “Ford, Moodie et al. (2013).”
- (k) Uppal, N, Shahab, L, Britton, J. and E. Ratschen (2013). The forgotten smoker: a qualitative study of attitudes towards smoking, quitting, and tobacco control policies among continuing smokers. *BMC Public Health*, 13:432. Hereinafter referred to as “Uppal et al. (2013).”
- (l) Van Hal, G., Van Roosbroek, S., Vriesacker, B., Arts, M., Hoeck, S. and J. Fraeyman (2012). Flemish adolescents’ perceptions of cigarette plain packaging: a qualitative study with focus group discussions., *BMJ Open* (2012). Hereinafter referred to as “Van Hal et al. (2012).”
- (m) Edwards, R., Peace, J., Russell, M., Gifford, H., Thomson, G. and N. Wilson (2012). Qualitative exploration of public and smoker understanding of, and reactions to, an endgame solution to the tobacco epidemic. *BMC Public Health*, 12:782. Hereinafter referred to as “Edwards et al. (2012).”

- (n) Dewe, M., Ogden, J. and A. Coyle (2013). The cigarette box as an advertising vehicle in the United Kingdom: A case for plain packaging. *Journal of Health Psychology*, Online First, published on October 22 2013. Hereinafter referred to as “Dewe et al. (2013).”
- (o) Lund, I. and J. Scheffels (2013). Young smokers and non-smokers perceptions of typical users of plain vs. branded cigarette packs: a between-subjects experimental survey. *BMC Public Health*, 13:1005. Hereinafter referred to as “Lund and Scheffels (2013).”
- (p) Scheffels, J. and I. Lund (2013). The impact of cigarette branding and plain packaging on perceptions of product appeal and risk among young adults in Norway: A between-subjects experimental survey. *BMJ Open* (2013). Hereinafter referred to as “Scheffels and Lund (2013).”
- (q) Rosenberg, M., Pettigrew, S., Wood, L., Ferguson, R. and S. Houghton (2012). Public support for tobacco control policy extensions in Western Australia: a cross-sectional study. *BMJ Open* (2012). Hereinafter referred to as “Rosenberg et al. (2012).”
- (r) Hoek, J and P Gendall (2012). Strong public support for plain packaging of tobacco products, *Australian and New Zealand Journal of Public Health*. Hereinafter referred to as “Hoek et al. (2012).”

DETAILED ANALYSIS OF THE STUDIES

Pechey et al. (2013)

4.3 This study uses an ‘expert elicitation’ approach. Using 33 tobacco control professionals from the UK, Australasia and North America, it attempts to estimate, via linearly pooling, the (1) most likely, (2) highest possible, and (3) lowest possible value for the percentage of (a) adult smokers and (b) children trying smoking, two years after the introduction of

plain packaging (all other things being constant) in a target country in the expert's region of residence. At best, it concludes that plain packaging regulations will not increase smoking prevalence. Absent that the panel of experts give a wide range of possible effects, with most believing it will have a very minor effect on youth uptake of smoking and increase the rate at which minors stop smoking.

Analysis

4.4 Although the use of expert elicitation methods is not uncommon, it is generally considered to be a less effective substitute compared to a comprehensive and more direct estimation of effects, such as meta-analyses. As I discuss further below, expert judgment can be a valid and useful method of analysis when there is a clear base model on which to anchor the judgments. However, the approach applied in this study is seriously flawed and its conclusions cannot be considered to have much, if any, validity.

Expert Judgment

4.5 First, there is considerable evidence that experts do not make judgments that are effectively any better than would arise either from random assessments or from more comprehensive and controlled statistical analyses. Studies on expert predictions show clearly that experts invariably (and overwhelmingly) underperform statistical models. This was the classic finding of Meehl,²⁷ which has been confirmed and reconfirmed by hundreds of studies.²⁸

²⁷ Meehl, P. (1954). *Clinical Versus Statistical Prediction*. Minneapolis, MN: Minnesota. Meehl, P. (1986). Causes and Effects of My Disturbing Little Book, *Journal of Personality Assessment*, 50: 370–375.

²⁸ Some classics in this area are: Goldberg, L. (1970). Man versus Models of Man, *Psychological Bulletin*, 73: 422–432; Einhorn, H. (1972). Expert Measurement and Mechanical Combination, *Organizational Behavior and Human Performance*, 7: 86–106; Dawes, R. and B. Corrigan (1974). Linear Models in Decision Making, *Psychological Bulletin*, 81: 97; Einhorn, H. and R. Hogarth (1975). Unit Weighting Schemas for Decision Making, *Organizational Behavior and Human Performance*, 13: 171–192. In the political

4.6 Second, the authors justify this ‘expert elicitation’ approach by stating that these approaches have been used in other areas of work. This is irrelevant. The veracity of any methodological approach needs to be based on the fact that it has been found to be superior in specific areas as the one being addressed. There is no indication as to: (a) what these other studies entailed, (b) how they match up structurally with the work that is being addressed in the paper, and (c) whether they actually did work because of the procedure or simply because of a confluence of lucky outcomes (in other words, the exercise was never repeated to determine if it would work again). Indeed, the examples given in the study – clinical trials, climate change, and volcanic eruptions – are more likely to be examples where there are more comprehensive and valid underlying studies to be examined and where there is more agreement as to the model driving the data being evaluated.

4.7 Third, we know from specific research on experts that expert judgment is less accurate than mechanical models,²⁹ but expert judgment operates using different processes and experts may, in specific contexts, be more efficient. The reason for this is that experts embody rules and structures that are developed uniquely over long periods of time. However, these “configurational rules” can be wildly wrong. This can occur for many reasons, including over-generalising from a small sample or from biased samples or by making inferences from inappropriate experience which, in this scenario, are studies that are poorly designed (as I discussed in my 2010 and 2012 Reports). However, perhaps the worst problem is related to “illusory correlations”. Illusory correlations have been subject to a vast literature and are not dissimilar to the pseudo opinions and sunspots that I discussed in my 2010 and 2012 Reports.³⁰ They arise when factors are

realm see Tetlock, P. (2006). *Expert Political Judgment: How Good Is It? How Can We Know?* Princeton, NJ: Princeton University Press.

²⁹ Dawes, R. (1993). Prediction of the Future versus an Understanding of the Past: A Basic Asymmetry, *The American Journal of Psychology*, 106(1): 1–24.

³⁰ Chapman J. and J. Chapman (1969). Illusory Correlation as an Obstacle to the Use of Valid Psychodiagnostic Signs, *Journal of Abnormal Psychology*, 46: 271–280; Hamilton, D.

believed to be correlated with some outcome when the evidence shows that there is no such relationship. This is most likely going to be significantly the case here as the sample selection is based on experts who have a vested interest in proving the veracity of their earlier published work and who are working off the same script. Indeed, the authors state that “*to ensure that all participants had the same summary of the most recent evidence relating to plain packaging*”, all the participants received a copy of the same systematic review.³¹

The Use of the Systematic Review

4.8 The systematic review that each participant received prior to undertaking the telephone interview is itself a source of potential bias.³²

4.9 First, it is noted at page 2 of the study that: “*Two systematic reviews of this indirect evidence have described three ways in which plain packaging may reduce smoking rates, particularly amongst children and young adults: first, by reducing the appeal of packs; second, by increasing the salience of health warnings; and third, by standardising pack colour, thus avoiding perceptions of this as an indicator of product harmfulness.*” However, as I noted in my 2010 and 2012 Reports, nearly all of the studies on plain packaging speculate that because of these factors there will be a reduction in smoking and smoking uptake. In reality, there is not a single study that actually tested the degree to which this is true. This fact is acknowledged by the authors of this paper: “*As this evidence is necessarily indirect, its*

and T. Rose (1980). Illusory Correlation and the Maintenance of Stereotypic Beliefs, *Journal of Personality and Social Psychology*, 39(5): 832–845; Dawes, R. (1989). Experience and Validity of Clinical Judgment: The Illusory Correlation, *Behavioral Sciences & the Law*, 7(4): 457–467.

³¹ Moodie, C., Stead, M., Bauld, L., McNeill, A., Angus, K., Hinds, K., Kwan, I., Thomas, J., Hastings, G. and A. O'Mara-Eves (2012). *Plain Tobacco Packaging: A Systematic Review*. London: Public Health Research Consortium.

³² A more valid approach would have entailed not just giving individuals the “pro” Systematic Review but a “con” alternative interpretation of the same literature. Such an approach would have at least countered part of the bias induced by having only one side of the debate being presented.

relevance has been questioned (for example, the strength of the relationship between thinking about quitting, a common outcome measure in this set of evidence, and actual quitting).” The authors also note that “*many experts mentioned that they were uncomfortable with providing a precise estimate for the impact, given **the lack of direct evidence**, and a few declined to give numerical estimates on this basis.*”

4.10 Second, the systematic review itself was originally meant to be a more structured and valid meta-analysis. However, this approach was abandoned and a less structured and highly speculative narrative systematic review was done. The review itself suffers from a host of methodological flaws, of which I note a few below:

- (a) The studies are included and compared without any real measure of the quality or importance of the studies. For example, Munafo et al. (2011) (which I reviewed in my 2012 Report) is included in the review as if it is a full fledge study yet even the authors of the study concede that it is little more than a pretest with limited validity.
- (b) The studies are assumed to be fully valid when the validity of the studies can be questioned based on many other criteria. For example, of the 23 survey studies, 18 used experimental designs that were shown (in every case) to be of seriously limited validity. A systematic review of fundamentally invalidly conducted research will amount to little more than “garbage in, garbage out”.
- (c) The studies are assumed to be valid even in the case where there is no way of determining their validity. For example, eight of the studies were little more than small sample focus groups and two other studies are interview based; yet these were considered to be equally valid to other studies in the discussion.

- (d) Although no study actually measured observed behaviour, it implies that behaviours were measured. As discussed my 2010 and 2012 Reports, upon reviewing the studies examined in the systematic review, not a single study actually measured behaviour.
- (e) The majority of the studies examined are related to a small group of researchers applying what are fundamentally overlapping methods of analysis. What this implies is that these studies are not independent pieces of information but repeated pieces of information. In other words, if two studies are 90% different and the findings are the same I am more confident in the findings than when the studies are 90% identical. Given the fact that many of these studies mimic the same empirical (and incorrect) approach, the consistency of the findings may be revealing very little.

The Methodological Issues

4.11 Finally, methodologically Pechey et al. (2013) is lacking in validity both from a sample size and a sampling perspective.

4.12 First, focusing on “tobacco experts” biases the approach. The experts are themselves likely to want to promote specific outcomes. But there is no real way of determining this without creating an alternative panel of non-experts or those who are experts in other areas (e.g., marketing modellers, industrial organisation economists, psychologists, and so on who would be considered to be experts in packaging, marketing, pricing, and regulatory responses).

4.13 Second, when attempting to glean estimates from individuals, even experts, one needs reasonably large sample sizes. In this case, the sample amounts to no more than 33 individuals (which provides no effective statistical validity). The reason one needs large samples is that one way of dealing with expert predictions is to engage in “pooling”. This is common

with many large-scale economic forecasts (e.g., the *Economist* uses a forecast of forecasts – which is nothing more than pooling). Pooling works when the “experts” are operating with the same basic decision model but that variation in their complete model adds both information and variance. For example, suppose that I want to forecast X. I get forecasts ($x_1, x_2, x_3 \dots x_n$) from many different forecasters who are all using a model that looks like $x_i = \alpha + \beta Z + \gamma_i S_i + \epsilon_i$. $\alpha + \beta Z$ is common to all models (e.g., economic growth is related to money supply; while $\gamma_i S_i + \epsilon_i$ is idiosyncratic to each forecaster. $\gamma_i S_i$ is their unique addition to the model and ϵ_i is their individual random error. By creating an overall forecast of $X = \alpha_1 x_1 + \alpha_2 x_2 + \dots \alpha_n x_n$ (where the weights all sum up to 1 and are determined statistically), the forecaster using this approach will, over time, average all the forecasters individually.

4.14 However, what is important here are two things. First, there needs to be a clear basic model in operation that adds a degree of stability; in other words, the participants have to agree on $\alpha + \beta Z$. What this implies is that you need to know very specifically what it is that you are asking the experts to judge – as might be the case with volcanic eruptions. As I have noted in my 2010 and 2012 Reports, none of studies on the impact of plain packaging have developed even the most basic behavioural model where (a) the parameters that the authors want the experts to make predictions about are known clearly and (b) the form of the model is understood. Again, what this means is that there is no clear base model ($\alpha + \beta Z$) on which everyone agrees and this is revealed by the reticence of many scholars to respond by providing estimates. In other words, most of what the study is testing is idiosyncratic responses (the $\gamma_i S_i + \epsilon_i$); which in this case may be little more than speculation as to effects of plain packaging, which may be biased by the nature of the sample as will be noted below. Unlike the case of volcanic eruptions, where there is a clear science, or clinical judgments, where causes are understood, the elicitation being asked in this study is little more than

pure guesswork as none of the existing studies on plain packaging actually measured the relevant variables. As noted in the paper:

A semi-structured telephone interview was used to elicit subjective judgments for the impact of plain packaging on (a) the prevalence of smoking in adults and (b) the percentage of children trying smoking. [...] During the interview, the interviewer provided the prevalence rates for the two outcomes of interest and asked participants to estimate the expected values of these two years after the introduction of plain packaging in their region, and the lowest and highest likely values, holding all other relevant factors.

4.15 However, there is not a single study on plain packaging that actually ever measured those prevalence rates – the most typical variable being either ‘attractiveness’ or ‘intent’. So what the researchers are asking their ‘experts’ to do is provide estimates about something they themselves have actually never measured but that they believe is related to something they have measured. By providing them with the base prevalence rates, they create an anchor around which the ‘expert’ gives what they believe are reasonable estimates, but, in reality, these are really no more than ill-formed guesses. As noted by Lovallo and Kahneman³³ this approach to expert judgment will lead to “*overly optimistic forecasts [that] result from the adoption of an inside view of the problem.*”

4.16 Further, the fact that some experts would not provide estimates is, in my view, telling. For example, Tetlock (2006) contends that the fox, “the thinker who knows many little things”, draws from an eclectic array of traditions, is better at prediction than the hedgehog, “who knows one big thing, toils devotedly within one tradition, and imposes formulaic

³³ Kahneman, D. and D. Lovallo (1993). Timid Choices and Bold Forecasts: A Cognitive Perspective on Risk Taking, *Management Science*, 39 (1): 17–31.

solutions”.³⁴ He notes an inverse relationship between the best scientific indicators of good judgment and the single-minded determination required to prevail in ideological combat. What this implies is that those with the most bias are more likely to actually make a prediction and that this will be worse the more they actually believe that they know what is going on. The better forecasters will actually be the ones who are least likely to want to give specific forecasts.

Conclusion

4.17 Overall, this study does not provide any effective or conclusive predictions as to the impact of plain packaging on smoking uptake or prevalence. Although the study itself is severely limited on a number of dimensions, its main limitation is that it uses as its basis a set of studies that have not addressed the primary behavioural issues related to smoking. Without a good base of scientific evidence, aggregation of scientific opinion is unlikely to provide any meaningful guidance.

Wakefield et al. (2013)

4.18 This study uses a simple telephone survey. Using a panel of 572 “current” cigarette smokers in Australia, the study queries them on their perceptions, attitudes and recall. The queries related to (a) exposure to the new plain packs, (b) ‘brand appeal’, (c) quitting thoughts and intentions, (d) opinions on health warnings, and (e) perceptions of cigarette ‘quality’. The key argued finding was that “*compared with smokers smoking from branded packs, smokers who were smoking from the new plain packs were more likely to perceive their tobacco as being lower in quality and tended to be lower in*

³⁴ Tetlock, P. (2006). *Expert Political Judgment: How Good Is It? How Can We Know?* Princeton, NJ: Princeton University Press.

*satisfaction, were more likely to think about and prioritise quitting, and more likely to support the plain packaging policy”.*³⁵

Bias

4.19 Before getting to the specifics of the research findings it is important to focus on a potential sources of bias in this study.

Response Bias

4.20 A very common source of bias in surveys arises when researchers create circumstances that induce the respondent to read intentionality into the exercise and respond in a manner in line with what they believe the researcher wants. In these circumstances it is extremely important for the researcher to hide the intent of the study, so as to induce the respondent to answer as truthfully as possible, and formulate the mixture of questions in a way that does not reveal intent. This is noted as a criteria of good survey design in section 2.12 of my 2010 and 2012 Reports and in the reports of Dr. Keegan (mentioned in Section 2 above). The response bias issue arises in this study at a number of points and in two forms.

4.21 First, the intent of the study is clearly related to smoking and plain packaging but it is unclear from the published material whether or not the study: (a) directly identified the fact that it was being conducted by the Cancer Council of Victoria and (b) identified itself as related to cigarette packaging and regulation directly. It did, however, identify itself with as being associated with “*attitudes and behaviours related to smoking*” and via its structure revealed its logic and intent.

4.22 Later in the survey two very leading questions were asked: “*Opinion about the new health warnings was determined by asking: ‘In order to strengthen the impact of the health messages on packs, the Federal*

³⁵ See page 10, Wakefield et al. (2013).

government requires that from December 1 2012, the size of graphic health warnings will be increased to cover 75% of the front of the cigarette pack. Do you support or oppose increasing the size of graphic health warnings to 75% of the front of the cigarette pack?’ and “Plain packaging opinions were assessed by asking: ‘Tobacco companies often use cigarette packaging for promotional purposes. Because of this, the Federal government requires that from December 1 2012, cigarettes will be sold in plain packaging. This will mean that the packs will all look the same with the logos and colours removed. Just the brand in plain text and the graphic health warnings will remain on the packs. Do you approve or disapprove of this?’”³⁶

4.23 Secondly, all of the questions in the survey hint at issues of ‘dissatisfaction’ and ‘harm’. For example, the first question – which asks whether or not the respondent is smoking from a plain or ordinary pack – immediately puts the respondent into a cognitive frame in which they recognise that the interviewer is very interested in this point. This is the case for two reasons. First, the question is asked early in the survey (at the very beginning). It effectively anchors and frames the later responses. Second, it is given before all the subsequent questions about smoking behaviour and intent and opinions. All of the later questions ask about consideration of ‘smoking’s harm’ or ‘quitting behaviour’, again signalling to the respondent a negative orientation and highlighting a potential for social response bias.

4.24 Hence, these two issues serve to create circumstances in which the responses to the questions may not be revealing truthfulness about the responses but a complex cognitive response in which the structure and design of the survey are inducing the results seen. Respondents pick up the subtle cues and use them to inform their responses.³⁷ As noted by Schaeffer and Presser (page 67): “As respondents hear or read a survey question, they

³⁶ Page 3, Wakefield et al. (2013).

³⁷ See various discussions of this and related issues in Schaeffer, N. and S. Presser (2003). *The Science of Asking Questions, Review of Sociology*, 29: 65–88.

construct a "pragmatic meaning" that incorporates their interpretation of the gist of the question, why it is being asked, and what constitutes an acceptable answer" (emphasis added). I will discuss how this is the case shortly.

Recall Bias

4.25 One of the two main conclusions of the study is based on two recall questions: 'quality' of the cigarettes compared to a year ago and 'satisfaction' compared to a year ago. Unfortunately, this information is not determined via a longitudinal panel but by asking the individual to compare today to what they thought they felt a year ago (and to do this in the context of a 12 minute telephone conversation). In my view, and consistent with good science, it would have been more appropriate to run a longitudinal panel where actual usage was tracked and the questions (biased though they were) were asked within the context of current and immediate behaviour.

4.26 Recall questions are common mainly because they are simple to execute (the alternative being to track what the same individual actually does or their responses over a long period of time). But that simplicity comes with a host of problems and these have been outlined in many studies.³⁸ For example, it is well known that anchor points (that to which the comparison is made), the amount of time permitted to do the recalling, the order in which recall is being asked (e.g., whether you are asked to recall first something far back and then more recent), how far back one is being asked to recall, whether or not the respondent believes the recall will be confirmed later with actual data, and so on all influence the response to recall questions. However, from the standpoint of this study what matters more is the

³⁸ See, e.g., Tanur, J. (1994). *Questions About Questions: Inquiries into the Cognitive Bases of Surveys*. New York: Russell Sage Foundation; Hassan, E. (2005). Recall Bias can be a Threat to Retrospective and Prospective Research Designs. *The Internet Journal of Epidemiology*. 3 (2).

cognitive model individuals use when engaging in recall. According to Pearson, Ross and Dawes³⁹ recall often involves a two step process:

First the individual begins by noting his or her present status... The present serves as a benchmark ... Construction of the past consists, in large part, of characterizing the past as different from or the same as the present. ... Second, people often invoke and implicit theory of stability or change to guide the construction of the past. (p. 68).

4.27 Which theory the individual invokes determines how they respond. This can be influenced by the circumstance in which the recall is embedded. When we add this concern with the response bias issue we can see where the two biases interact. The initial framing of the questions create a context in which the respondent must create a past from what cues they are receiving in the present. Given that the initial questions relating to plain packs imply a situation of change, it is entirely possible that the respondents look to respond in a manner that reflects differences over similarities because that is what the ‘cues’ coming from the survey push them towards.

Analysis

4.28 Taking into account the discussion in paragraphs 4.20 to 4.27 above about potential biases, there are additional methodological issues that create questions as to the validity of the conclusions drawn by the authors of this study.

4.29 First, as with many of the studies I have examined in this report, and in my 2010 and 2012 Reports, there is a problem with “one at a time” comparisons. For example, comparing those using plain packs with those using standard packs, by assuming that all other factors that may influence the outcomes are effectively controlled. I have criticised this approach before

³⁹ Pearson, R., Ross, M. and R. Dawes (1994). Personal Recall and the Limits of Retrospective Questions in Surveys, in Tanur, J. (1994). *Questions About Questions: Inquiries into the Cognitive Bases of Surveys*. New York: Russell Sage Foundation, 65–94.

and argued that a better approach is to utilise more sophisticated econometric approaches and to also create a sampling frame that is driven by the key comparisons that are being sought.⁴⁰ In this study, concerns about these controls are potentially more of an issue because the key comparison is between those using plain packs and those who are not and the latter group is getting smaller as the survey process continues. As evidenced in Table 1 of the study on page 4, the number of respondents falling into the ‘plain pack’ category increases (which I recreate below but using the actual number of respondents).

	Plain Pack	Regular Pack
Week 1	71	54
Week 2	94	48
Week 3	114	28
Weeks 4-5	126	19

4.30 To be able to make the sorts of comparisons given in Figure 1 of the study (which relates to ‘appeal’ outcomes by survey week for plain and branded pack smokers) and to conclude that they are not being driven by other factors, the controls utilised must be valid within each cell. What this means is that whenever a breakdown comparison is made, the Adjusted Log Odds have to be purified of all other factors, which will be impossible as the sample in each comparison group gets smaller (e.g., in the above beginning in week 3). In addition, what this also implies is that as the number of those using the regular packs gets smaller, the validity of any comparison to that group gets weaker (and less valid). The only real solution to the problem would have been to oversample those using regular packs as the survey

⁴⁰ See, for example, my review of Gendall et al. (2011) in paragraphs 4.3 to 4.9 of my 2012 Report, Hammond, Doxey et al. (2011) in paragraphs 4.18 to 4.30 of my 2012 Report, Gallopel-Morvan et al. (2011) in paragraphs 4.49 to 4.56 of my 2012 Report and Borland et al. (2011) in paragraphs 4.57 to 4.77 of my 2012 Report. For brevity, I have not sought to list all examples here.

process continued to ensure that there were enough respondents to make valid conclusions.

4.31 Hence, what we see are two facts. First, there are two critical sources of bias that will potentially induce individuals into a cognitive frame in which they will create false differences – believing that their cigarettes tasted different when the reality is that they are absolutely identical. Second, as the number of individuals using plain packs increases the number of individuals influenced by that bias increases while the comparator group (those using regular packs) becomes subject to more extraneous and uncontrolled effects. Together, these factors raise questions as to the validity of the responses to the questions relating to brand ‘quality’ and ‘satisfaction’.

4.32 There are additional issues with the ‘quality’ and ‘satisfaction’ questions relating to how the questions are asked. First, the fact that the questions are being asked (and asked together) implies that they are not just ‘important’ but that ‘consistency’ is demanded in the response (e.g., how many people responded that they were not satisfied but that the quality was the same or that the quality declined and they were satisfied?). Good survey design would require that these questions (which effectively are tapping similar facts) would be at different points in the survey. Second, the form of the questions – ‘higher’, ‘lower’ or ‘about the same’ – could easily have been done differently and would potentially have led to very different responses. For example, the use of the word “about” implies that a change is possible (e.g., why was it not ‘higher’, ‘lower’, or ‘the same’). Another alternative would be to ask “do you perceive any changes in the quality of the cigarettes you are smoking today as compared to when you were smoking the same brand 6 months ago? Yes or No?” and then if they answer “Yes” there are two alternatives: “The Quality is better, The Quality is worse.” Another alternative would be to ask the same question but give the individual the option to say that they do not recall or cannot remember. The point here is that the study implies that the answer to this question is a definitive outcome

with some potential “halo effects” (page 7) but the structure framing of the question could easily be the driver of the responses when it is most likely obvious to the respondents that the cigarettes themselves had not changed.

4.33 Moving away from the questions of the validity of the conclusions reached by the authors in respect of ‘quality’ and ‘satisfaction’, we can see that there are also concerns as to the validity of the ‘quit intention’ questions. First, it should be clear that the same bias issues discussed earlier would be relevant here. Individuals will be more likely to respond that they have thought about quitting and it is more important to them when they are using plain packs because the survey structure highlights the importance of the change. This is made even more salient by asking the harm questions before the intention to quit questions. In addition, there is no actual measure of whether or not the individuals engaged in any later behavioural attempts to quit; hence the actual validity of the responses cannot be confirmed.

4.34 Although it was not utilised as anything but a covariate in the study, the addition of a question posed to respondents on advertising recall highlights other aspects of methodological rigour in this study. Respondents were asked whether they could recall one of the three public service advertisements. A common approach in recall studies is providing fake alternatives (e.g., a fourth advertisement that was never shown to the public) or lie scales (repeating items in different formats). One can then potentially account for individuals who are simply responding to the survey without actually revealing anything that might be truthful.⁴¹ This would have been one effective way of accounting for at least part of the social desirability bias in the study.

⁴¹ See, for example, Davis, C., Thake, J. and N. Vihena (2010). Social Desirability Biases in Self-Reported Alcohol Consumption and Harms”, *Addictive Behaviors*, 35: 302–311; af Wählberg, A., Dorn, L. and T. Kline (2010). The Effect of Social Desirability on Self Reported and Recorded Road Traffic Accidents, *Transportation Research Part F*, 13: 106–114.

4.35 Finally, this social response bias is a potential concern with the support for plain packaging regulations on the part of those using plain pack cigarettes. As noted before, the respondent is reading the cues from the form and structure of the preceding questions, at the end they get a final question that highlights the intent of the regulations but also reinforces the purpose of the survey in the first place.

Conclusion

4.36 Overall, this study does not provide evidence that plain packaging will lead to any fundamental change in behaviour. From the standpoint of survey design, it is limited dominantly because of the use of recall questions on the key variables of interest ('quality' and 'satisfaction') and its structure and design are likely to have strongly induced social response bias in relation to 'quitting intention', and reported support for plain packaging legislation. In a more general sense, the study is also unable to come up with an effective behavioural model that relates its findings to outcomes.

Moodie and Mackintosh (2013)

4.37 This study is a replication of Moodie et al. (2011), which I discussed in paragraphs 4.10 to 4.17 of my 2012 Report. It seeks to examine how using plain packs influenced pack perceptions, feelings about smoking, salience of health warnings and self-reported cigarette usage amongst a sample of 301 adult female smokers (aged 18 to 35) in various cities and towns in Scotland. The queries related to (a) exposure to the new plain packs, (b) 'brand appeal', (c) quitting thoughts and intentions, (d) opinions on health warnings, and (e) perceptions related to cigarette 'quality'. As this study is identical in formulation and execution to Moodie et al. (2011), I will only highlight the major issues again, adding a few clarifications. The reader can refer to my earlier report for specifics and a comparison.

Sample Size

4.38 Before getting to the specifics of the study it is important to acknowledge that, unlike Moodie et al. (2011), this study has sufficient sample size and the discussion about the small sample size in that study is not relevant here. However, it is important to point out that like Moodie et al. (2011), this study also had a very significant participant drop off rate (only 187 of the 301 initial participants completed the study). Such a drop off rate can have serious consequences for both validity and interpretation as those remaining in the study clearly had their own reasons for doing so. This potentially influences their responses to the tasks given. In particular, if the reason for staying in the study is related to the desire to please the researcher, then it is likely that the responses will be significantly biased. It is noted that there was no attempt in this study to address the attrition effect other than a simple demographic comparison, which is acknowledged by the authors on page 4.⁴²

Analysis

4.39 This study, by utilising a more active field-based activity, is an improvement on survey-based perception studies common in this field. By asking individuals to transfer their cigarettes from branded to plain packages for a specific period of time (in this case two weeks), there is an attempt to get closer to the use environment in which smoking is done.

4.40 As noted in paragraph 4.11 of my 2012 Report, the initial set up of the study is done in a manner that immediately reveals to the subject that the intent of the study is a comparison between perceptions based on package usage. For example, the pack transfer task (where individuals move their

⁴² For issues related to subject attrition see, e.g., Goodman, J. and T. Blum (1999). Assessing the Non-Random Sampling Effects of Subject Attrition in Longitudinal Research, *Journal of Management*, 22 (4): 627–652; Das, M., Toepael, V., and A. van Soest (2011). Nonparametric Tests of Panel Conditioning and Attrition Bias in Panel Surveys, *Sociological Methods Research*, 40 (1): 32–56.

cigarettes to the Kerrod pack) immediately primes the subject and reveals to them that what matters to the researcher is a comparison between their survey responses based on the usage of the package. This could have been resolved by utilising a design that incorporated between and within subject effects. For example, each individual was given the study materials all at the same time and simply told which order to use the packs (hence all the comparisons were within subject). It would have been better to have two control groups – one that smoked using their normal packs and one that only used the Kerrod plain packs and used these to establish a benchmark. While not completely resolving the issue of bias created by the artificiality of the approach, it would have allowed for more substantive comparisons that would address issues of bias based on the design itself.

4.41 It is also the case that the study could have been improved upon dramatically by not using the transfer aspect of the task. A more effectively designed study would have simply provided the individual with their brand of cigarettes in a plain package and then monitored the actual usage. In this case, the subject's cigarettes would have already been in the appropriate pack as the investigator would have done the transfer and any additional cigarettes could have been provided to them when appropriate. This would have: (a) removed the artefact of transferring the cigarettes to a plain package, which could have biased the results by increasing the salience of the task on one dimension while also being a factor leading to the high attrition in the study; and (b) allowed for a direct behavioural usage measure that could have been compared to a calibration period that measured normal cigarette usage.

4.42 Although this study is field-based, it fails on two basic dimensions. First, there are no actual behavioural usage measures. All that is collected are perceptual self-reports. While the authors make a point that the plain pack cigarette usage is 'significantly' smaller the actual 'effect' size is small (14.9 vs. 15.5 and 15.7 vs. 16.7), this is well within a margin of error when considering the meaningfulness of the results and the fact that individual

recall adds an additional source of error. Hence, without any actual usage measures, the ability to draw any policy inferences from the results is lost.⁴³

4.43 The most statistically significant results from the study arise in the perceptual measures, but again one has no way of knowing whether or not these perceptions matter, as there are no direct behavioural measures to which they are even remotely linked. The fact that someone may perceive a plain pack as less perceptually ‘appealing’ than one with more colour or design characteristics is probably not surprising but as with all other studies utilising the same stated survey approach, the link to any behavioural outcome is not made. In addition, the fact that the participants in the study are clearly in a study aimed at generating comparisons one would expect to see bias creep into the study responses.

4.44 As I indicated in my discussion of Moodie et al. (2011) in my 2012 Report, a more effective approach would have been to: (a) monitor usage without plain packaging; (b) survey without plain packaging; (c) provide plain packaged cigarettes; (d) survey with plain packaging; (e) monitor usage with plain packaging and (f) repeat (a) – (b) in a post-test confirmation. Using this sort of approach would have allowed the researcher to: (1) establish a clear control usage group (point a); (2) examine the effect of survey salience arising by just asking questions of respondents (point b); examine the impact of points (1) and (2) with respect to plain versus regular packaging; and (3) get a clear measure of actual usage with plain packaging (point (e)). The post-test confirmation would then allow the researchers to examine longer-term effects.⁴⁴

⁴³ Issues with self-report questions are outlined by Schwartz, N. (1999). Self Reports: How the Questions Shape the Answers, *American Psychologist*, 54 (2): 93–105.

⁴⁴ The approach discussed here is not dissimilar to that given in several studies reported in Belli, R. F., Stafford, F. P., and D. F. Alwin (Eds.) (2009). *Calendar and Time Diary Methods in Life Course Research*. Sage: Thousand Oaks, CA. Perhaps the most famous is the Day Reconstruction Method, employed by Nobel Prize winner Daniel Kahneman and his colleagues; Kahneman, D., Krueger, A., Schkade, D., Schwarz, N. and A. Stone (2004), A Survey Method for Characterizing Daily Life Experience: The Day Reconstruction Method,

4.45 In addition, the use of self-administered responses where the respondent determines when and how to respond can be improved upon by random response approaches. Such approaches “buzz” the individual at random times and query responses at that time (normally by SMS messaging). Hence, one could monitor product usage not by asking the participants to recall “what they may or may not have done” but specifically focusing on certain outcomes at a moment in time. For example, “have you had a cigarette in the last hour?”, “Did you give a thought to quitting in the last hour?”, and so on. Such responses are found to be more revealing as they are taken in real time.

Conclusion

4.46 In my view, there is no indication from this study that plain packaging actually impacts smoking behaviour. Although this study represents a more realistic examination of smoking perceptions than other studies I have reviewed, it is still significantly removed from actual behaviour. This last fact is important, as actual smoking behaviour was never observed or measured, despite the fact that this could have been done if more effort had been put into the study. In addition, the method utilised together with the lack of direct observation and reliance on self-reporting by the subjects are all potential sources of bias that are unaccounted for. This is further exacerbated by the fact that only 62% of the individuals starting the experiment went on to complete the study and no attempt was made to address attrition bias in the structure of the study itself; a fact that is interesting given that the authors clearly knew that attrition bias was an issue since it appeared dramatically in their previous research (Moodie et al. (2011)).

Science, 306 (5702): 1776–1780. Today, most such studies employ SMS messaging to randomly ‘buzz’ respondents to measure specific behavioural activities at random points in time.

Hammond et al. (2013)

4.47 This study is an online survey in which 947 16 – 19 year old females in the United Kingdom rated a series of cigarette packs, where the packs varied in design (i.e., branded versus plain packs and various different cigarette pack shapes). The study compared perceptions of ‘brand appeal’, ‘attractiveness’, ‘taste’, ‘harm’, ‘tar’ and ‘image’ across individuals. The study’s structure, and many of the questions, are identical to a number of studies discussed in my 2010 and 2012 Reports from the same lead author (Professor David Hammond of the University of Waterloo, Canada) and hence the reader can refer to my earlier reports and commentary on those articles as almost all of the major limitations of those studies remain here.⁴⁵ Overall, this study argues that: *“marketing in the form pack branding remains a potent tool for increasing the appeal of tobacco products to young women. The findings provide empirical support for plain cigarette packaging regulations in Australia”*.⁴⁶

Analysis

4.48 This study suffers from a number of critical limitations reflective of this stream of research on plain packaging (discussed in this report, and in my 2010 and 2012 Reports).

4.49 First, there is no underlying behavioural model into which the analysis fits. As with many other studies I have evaluated, all that is done is individuals are asked to make a series of inferences based upon a sequence of different packs presented to them. As noted this immediately signals to the subject that the point of the experiment is to discover rationales for the comparisons they are being asked to make. In addition, this bias is

⁴⁵ For ease, examples include Hammond, Doxey et al. (2011), Gallopel-Morvan et al. (2011) and Wakefield, Germain et al. (2012) discussed in my 2012 Report and Hammond, et al. (2009), Hammond and Parkinson (2009), Doxey (2009) and Bansal-Travers et al. (ND) in my 2010 Report.

⁴⁶ See page 6, Hammond et al. (2013).

exacerbated when individuals are queried on an item for which there is no natural logic for what it is that they are being queried on and where they have no prior knowledge. For example, questions relating to “tar delivery” (“*How much tar do you think these cigarettes would have compared to other cigarette brands?*”) and “health risks” (“*How would the health risks of these cigarettes compare to other cigarette brands?*”) are: (a) assuming that the individual is competent to understand the meaning of “tar delivery”; and (b) define “health risks” in a manner that is comparable between individuals. In addition, the comparator of “other cigarette brands” throws in an arbitrary base case against which the subject is supposed to compare. However, there is no guarantee that one person’s “other cigarette brands” is the same as another person’s, or that many individuals would know little or nothing about the tar yield of these “other cigarette brands” and what that might mean, if anything, in terms of comparative “health risks” of the different brands. As noted with respect to numerous studies using the approach seen here, subjects will attempt to generate what they consider to be consistent opinions, and be more likely to generate “pseudo opinions” that they believe reflect what it is that the investigator is seeking.

4.50 This issue of perceptions is made worse by the artificiality of the question design and measurement construction. The problems with this aggregation can be seen in a few examples which show that there is no theoretical or logical justification for such an approach:

- (a) “Brand appeal” ratings were determined by a question: “*How appealing is this brand of cigarettes compared to other brands on the market?*” The ratings were then aggregated so that “a lot more appealing” and “a little more appealing” received a score of “1” and “no difference”, “a lot less appealing”, “a little less appealing” received a score of “0”. The same approach was utilised for all the other perception ratings – e.g., “*Compared to other cigarette brands on the market, would these cigarettes be...less/more harmful?*” In

essence, this measurement approach assumes a clear dichotomous cognitive model for which no justification is given.

- (b) Unlike prior studies by the research team (see, for example, my comments on Hammond, Doxey et al. (2011) in paragraph 4.22 of my 2012 Report), the authors now exclude the “don’t know” option to their questions. I was highly critical of how they used that option in prior work. The solution that they chose here was simply not to have it as an option, forcing all the individuals to give a response when “don’t know” might be the most informative response that they could give.⁴⁷ This is a very disturbing outcome as it reveals that rather than attempting to build a meaningful cognitive response model that accounts for those that “don’t know” the researchers choose to solve a problem in their design simply by deleting a problematic part. They also make no statement that they are using exactly the same construct names (e.g., “health risks”) as in earlier studies but now creating them via different means. This is a serious violation of good research practice, the standards of publication and can confuse and bias those reading the studies and those attempting to replicate them.
- (c) The smoker image measure shows a potential common method bias as it is fairly clear that there is no effective differentiation amongst the 7 traits.

4.51 These arbitrarily aggregated measures are then further confused by the creation of an index in which each package in the experiment is added up to create a 1-10 scale that is meant to represent an overall “Brand” rating index. This creates what amounts to a theoretic formative measure for “Brand” that supposedly measures the overall perception of the ten packs

⁴⁷ A good example of this in the policy arena is given by Luskin, R. C. and J. G. Bullock (2011). ‘Don’t Know’ Means ‘Don’t Know’: DK Responses and the Public’s Level of Political Knowledge, *Journal of Politics*, 73 (2): 547–557.

seen by the subject. Again this is an approach that has been used in other studies. I noted previously that the meaning of these indexes was totally unclear and violated the logic of formative index design outlined by Rossiter and others (as noted also in paragraph 4.41 of my 2010 Report). However, in this study the index they created is even more confused. In Hammond, Doxey et al. (2011), the authors created four indices – one each for ‘Appeal’, ‘Taste’, ‘Tar’ and ‘Health’. In this case they actually aggregate these measures to create one “brand” index. However, there is no justification theoretically why this is appropriate or why it is that they deviated so dramatically from an approach they had utilised before and indeed cite in the current study. Again, good research practice requires that changes in construct structures be justified and also that where their measurement structure deviates from what was used in other work by the authors this fact must be noted clearly so that other researchers understand the extent to which constructs and measures are comparable across studies.

4.52 The authors also argue that they now have a “behavioural task” in their design. However, it is not clear what behaviour they are measuring with this task. What they have is not an actual incentive compatible choice or set of choices being made in realistic environments but simply a task in which they ask the individual to make a choice amongst four different packs shown on the screen. In the end the individual does not actually receive the pack nor is there anything they are giving up to get the choice that in the end they never receive! A more effective decision scenario would be to: (a) vary the packs seen via a factorial experimental design and (b) then give the individual a choice of taking a pack or money. This would at least add a degree of incentive compatibility to the structure of the final task. In the end, it is not measuring any specific realistic behaviour. What is also quite perplexing about this task is that the key argument being made by the whole study is that individuals are overwhelmingly negatively affected by the plain packs. Yet when they are then given a choice of taking a plain or regular

pack, with no price penalty, the ratio of those accepting the “branded” vs. “plain” pack was $51.8/44.6 = 1.16$. While the authors argue this is ‘significant’, it is quite counter intuitive and implies that individuals are not playing according to the model that the author(s) argue. If plain packs are so “off putting” to consumers (as is alleged in a number of the studies I have reviewed in the context of plain packaging), one would expect that none of the individuals measured in the study would choose the plain pack. Hence what matters is not that smokers and non-smokers make different choices, but that when everything else is held constant (e.g., ‘quality’, price, etc.), no one would choose the plain pack since it has nothing additional to offer. This should be seen even more so in the case of the smokers surveyed in this study, as the authors suggest that they are even more “influenced” by the branding and packaging of tobacco products (e.g., the authors of the study stated that “*smokers were also significantly more likely to rate packs as appealing*”). If anything these results hint that individuals may not really be taking the task seriously particularly given they were only paid the equivalent of CAN\$2 to be involved in the study. This amount is extraordinarily small and there is considerable research hinting that incentives can impact on who participates, data quality and participant involvement.⁴⁸

Conclusion

4.53 This study suffers from a number of significant methodological flaws that renders its validity suspect. As noted, it repeats many of the same study, design, measure construction, aggregation and analysis problems of related studies by this mix of authors. In addition, the authors make subtle but potentially meaningful adjustments to the measures they use in the study without noting the changes from their earlier work. Moreover, they argue

⁴⁸ See, e.g., Simmons, E. and A. Wilmot (nd). *Incentive Payments on Social Surveys: A Literature Review* (Available at: <http://www.ons.gov.uk/ons/guide-method/method-quality/general-methodology/data-collection-methodology/reports-and-publications/incentive-payments-on-social-surveys.pdf>); Deutskens, E., de Ruyter, K., Wetzels, M. and P. Oosterveld (2004). Response Rate and Response Quality of Internet-Based Surveys: An Experimental Study, *Marketing Letters*, 15 (1): 21–36.

that their study involves a ‘behavioural’ task that increases its validity in terms of market outcomes. However, their behavioural task is neither realistic (in that it does not include other relevant decision making factors – e.g., price) nor does it actually model or reveal any specific smoking related behaviour. The task also is out of line with the overall conclusions. While the authors argue that plain packs are overwhelmingly disliked, the supposed behavioural pack choice task reveals that almost half of participants choose the plain pack when logic would imply that given there is no price difference there is no reason at all to choose the plain pack at all.

4.54 Finally, the study draws conclusions that are beyond the purview of the data gathered. As noted on page one the study purports to find that “*marketing in the form pack branding remains a potent tool for increasing the appeal of tobacco products to young women.*” And that “*the findings provide empirical support for plain cigarette packaging regulations in Australia*”.⁴⁹ Overall, the study’s methodological flaws imply that its results are most likely reflective of the nature of the study design more than the realistic effect of any policy intervention.

Hammond, White et al. (2013)

4.55 This study is an online survey in which 762 11 – 17 year olds, who were recruited from a survey firm in the United Kingdom, rated a series of cigarette packs, where the pack design varied based on pack colour and type of health warning (i.e. graphic or text only). The study compared perceptions of brand appeal, attractiveness, taste, harm, tar, health warning impact and incentive to start smoking. The study’s structure and many of the questions are identical to a number of studies discussed in my prior reports from the same lead author (Professor Hammond), plus Hammond et al. (2013) discussed at paragraphs 4.47 to 4.54 above. Hence to avoid repetition, I will simply note areas in which the concerns discussed above are repeated and the

⁴⁹ See page 6, Hammond et al. (2013).

reader can also hence refer to my earlier reports and commentary on other articles by this team of researchers, as almost all of the major limitations of those studies are present here.⁵⁰ Overall, the authors of this study argue that: *“increasing the size of pictorial health warnings and standardizing the appearance and shape of packages may discourage smoking initiation among young people”*.⁵¹

Analysis

4.56 This study suffers from a number of critical flaws reflective of this stream of research.

4.57 First, as noted for all the studies from this research team there is no underlying behavioural model into which the analysis fits. As with many other studies I have evaluated, all that is done is that individuals are asked to make a series of inferences based upon a sequence of different packs presented to them. No model is tested, with the analysis simply being a search for effects that are then rationalised in an ex post facto manner.

4.58 The measures suffer from the same construct/index process problems outlined in my discussion of Hammond, et al. (2013) in paragraphs 4.47 and 4.48 above. The two most notable are that individuals are being asked to make decisions for which they lack knowledge (hence adding bias into the responses) and the aggregation process is arbitrary and not justified statistically or theoretically.

4.59 Second, the study makes it quite clear that the comparison is related to package design and health warnings and that what is being compared are changes against a fixed base as *“each pair included the same reference pack,*

⁵⁰ As above, examples include Hammond, Doxey et al. (2011), Gallopel-Morvan et al. (2011) and Wakefield, Germain et al. (2012) discussed in my 2012 Report and Hammond, et al. (2009), Hammond and Parkinson (2009), Doxey (2009) and Bansal-Travers et al. (ND) in my 2010 Report.

⁵¹ Page 6, Hammond, White et al. (2013).

a branded Benson and Hedges (B&H) pack”.⁵² This structure immediately creates a cognitive framing effect that biases the respondent toward inflating differences in a manner that would appear to be logically consistent.

4.60 This last point reflects a problem with the overall structure of the study and its failure to meet the criteria for experimental design outlined in paragraph 2.10 above (efficiency, orthogonality, and balance). The authors argue that they are using a factorial design but in reality all they do is use the six alternatives versus a base pack. What would be more correct (assuming that alternatives they are using are logically correct – which I will shortly argue is not the case) would be to use a full factorial design in which all variations in the pack designs were compared (with meaning that it would not be a 1 x 6 design but a variant of 7x7. The table below shows the design of the study (excluding the last pack comparison).

	A	B	C	D	E	F
PACK	W	W	W	B	B	B
WARNING	T40%	P40%	P80%	T40%	P40%	P80%
BASE CASE	B&H	B&H	B&H	B&H	B&H	B&H

4.61 This design allows only for simple comparisons against the base case. We can see immediately that (a) it is not orthogonal (individuals are not comparing all alternatives – for example we do not see a text descriptor on 80% of the pack), (b) it is not balanced (mixtures of the alternatives do not appear the same number of times (text alternatives appear twice but pictures four times), and (c) it is not efficient (it does not allow for a comparison of text versus pictures since there is never a comparison made). A complete design would have looked like that given below (assuming only the alternatives used by the authors).

⁵² See page 2, Hammond, White et al. (2013).

	W/T40%	B/T40%	W/P40%	B/P40%	W/P80%	B/P80%
W/T40%						
B/T40%	1					
W/P40%	2	7				
B/P40%	3	8	12			
W/P80%	4	9	13	16		
B/P80%	5	10	14	17	19	
B&H	6/A	11/D	15B	18/E	20/C	21/F

4.62 Comparing these two tables we can see immediately the inefficiency of the design used by the authors (where 15 of the possible 21 comparisons that should be made in a full factorial structure are not made). This may appear to be a minor technical issue but it is not since what it implies is that no statements can be made about the value of text representations versus pictorial representation since there are no direct comparisons of these that are independent of either the colour or the base case. The same is true of the impact of colour. For example, a comparison of the importance of white versus brown pack colour. This is a concern as this is exactly the sort of comparisons and conclusions drawn by the authors (e.g., *“The study also demonstrates that health warnings and standardized packaging have independent effects on perceptions of cigarette products. Compared with text warnings, pictorial warnings were perceived to have greater impact ...”*⁵³ and the text that follows on page 6). Such conclusions are not scientifically possible given the nature of the design that is being used.

4.63 Third, as noted in various places in my 2010 and 2012 Reports (and at paragraph 4.29 above) “one at a time” comparisons are an inappropriate means of analysis when individuals are using complex decision models and

⁵³ See page 6, Hammond, White et al. (2013).

when one wants to examine results that have natural inter-relationships. A more appropriate approach would have been to utilise a more appropriate design (as outlined in paragraph 4.56 above) and a multivariate statistical model in attempting to decompose the relationship between the instruments/stimuli and the respondents' evaluations.

4.64 Third, the use of a base case exacerbates the significance of the results. In this study, the 'base case' is that option against which the individual makes all of their comparisons (in this study, the Benson and Hedges pack). Independent of the nature of the instrument, the fact is that the base case will create an anchoring at an extreme. As noted by Friedman and Amoo (1999), "*By focusing mainly on the features of the subject [in this case the pack variations], respondents tend to overlook unique features of the referent [the B&H pack]. Thus, items of comparison that have unique positive features should receive more positive evaluations when they are subjects rather than referents [which is exactly the case in this study]; items that have unique negative features should receive more negative evaluations when they are subjects of the comparison rather than referents [which is the opposite of what we see in the case of this study]*".⁵⁴ What this implies is that by using the Benson and Hedges referent as the 'base case' and the variations as done in this study, we will see inflated results that will align with specifically what the authors expect.

4.65 Fourth, it is interesting that the entire survey took only 10 minutes on average to complete (see page 2 of the study). Given that each individual had to look at the stimuli and answer 42 questions (6 x 7) plus do a final task this implied that they spent a little more than a minute on each task and 14 seconds on each question.⁵⁵ This immediately raises the question whether the

⁵⁴ Friedman, H. and T. Amoo, (1999). Rating the Rating Scales, *Journal of Marketing Management*, 9 (3): 114–123.

⁵⁵ See, e.g., Schwartz, N. (1997). Questionnaire Design: The Rocky Road from Concepts to Answers, in Lyberg, L. et al., eds. *Survey Measurement and Process Quality*, Englewood

individual is involved cognitively in the task. 14 seconds may be sufficient in some cases but the fact that subjects are going through the survey quickly may also exacerbate bias when individuals are being asked to respond to questions for which they are making inferences rather than using evidence (as would be the case here in asking individuals questions about “taste, health risk, tar level, and health warning impact”) and where they are likely to want to respond in a manner that is consistent with what they perceived to be the intent of the survey (as would be true here in the case of “attractiveness and incentive to start smoking”). The importance of such “pseudo opinions” has been well documented and is a significant issue with respect to most all of the studies I have evaluated (see, e.g., paragraph 4.47 of this report).

4.66 Finally, like Hammond et al. (2013) above, the authors argue that they validate their results with a behavioural pack preference task. I discussed this in detail in paragraph 4.50 above and it is important to reiterate that this is not a realistic choice task as no real alternative is given. 81.5% of subjects never tried smoking (according to their self reports) yet 64.2% of them made a choice of a product that they previously indicated that they did not have any use for.

Conclusion

4.67 This study concludes by noting: “*the current study underscores the potential for packaging to communicate desirable product characteristics ...*”.⁵⁶ However, this is an invalid and erroneous conclusion. In my opinion, the structure of this study very likely led to a significant overstatement of the effects that it found and the conclusions drawn are not justified based upon the nature of the study design. The fundamental conclusion “*that health warnings and standardized packaging have independent effects on perceptions of cigarette products. Compared with text warnings, pictorial*

Cliffs NJ: Wiley, 29–46. The other papers in this volume also speak to the measurement issues discussed throughout this and my prior reports.

⁵⁶ See page 6, Hammond, White et al. (2013).

warnings were perceived to have greater impact ...” cannot actually be determined given the structure of the survey as the instrument used never allowed for an efficient and statistically valid comparison of text and pictorial warnings since every comparison was relative to the base case of a standard B&H pack.

Borland and Savvas (2013)

4.68 This study is an online survey of 320 adult “current” and “ex” smokers in Australia aimed at determining the impact of “variant descriptors” on plain packs on perceptions of ‘quality’, ‘strength’, tar and nicotine, and ‘harmfulness’. Overall, this study argues that: *“in summary, these results show that smokers impute differentiable characteristics to cigarettes based on variant labels. As such, variant descriptors on plain packaging can be used as a powerful branding tool. This role of variant descriptors will be even more important than it has been, as most other aspects of cigarette differentiation were removed from Australian packs from December 2012. While variant descriptors are allowed, it seems to us inevitable that smokers will try to make value judgements between them”*.⁵⁷

Analysis

4.69 This study suffers from a number of common methodological flaws that I have pointed out with respect to other studies from this team of researchers – in particular Borland et al. (2011) in paragraphs 4.57 to 4.77 of my 2012 Report. Of the issues discussed in that analysis, a number are repeated here and simply noted while specific limitations unique to the current study are outlined.

4.70 First, it is unclear exactly what each subject saw and did. Page 2 of the study notes that *“participants were asked to make judgements about sets of descriptors ...”* but we are never told exactly how many variants, whether

⁵⁷ See page 6, Borland and Savvas (2013).

or not the stimuli were truly generic (e.g., did it say “BRAND NAME” or was a brand name actually inserted). Inferring from what is given on page 2, individuals saw 17 packs in the first study and 19 in the second study. In addition, we are told that this study was effectively inserted into other studies but how those studies related to the one reported here is not clearly outlined in a manner that allows an evaluator to know whether those studies created other methodological problems for the results.

4.71 Second, let us assume that my inference is correct and individuals saw 17 packs in the first study and 19 in the second study. Accordingly the authors thus entailed (emphasis added):

*The sets of variant descriptor terms shown in the first survey and ratings made were Colour (Red, Blue, Gold, Silver, White)—taste and tar/nicotine levels; Flavour (Infinite, Ultimate, Full Flavoured, Smooth)—taste and tar/nicotine; Venting (Highly vented, Medium vented, Low venting, Unvented)—taste, tar/nicotine and harmfulness; Filter (Advanced filter technology, Charcoal filter, Filter tipped, Dual filter)—tar/nicotine and harmfulness. **The sets were presented in this order.***

*The second survey only asked about taste and quality for four sets of variants. It used variants differing in Colour (Red, Blue, Gold, Silver, White); Quality (Rich, Distinct, Premium, Premier, Refined); Style (Standard, Virginia, American blend, Oriental blend); Blend (Blend 001, Blend 004, Blend 012, Blend 333, Blend 879) **presented in that order.**⁵⁸*

4.72 This immediately creates bias issues, as individuals will clearly see patterns in what is being presented to them. Indeed, in nearly all of the results we see this pattern arising quite clearly as the authors acknowledge in

⁵⁸ See page 2, Borland and Savvas (2013).

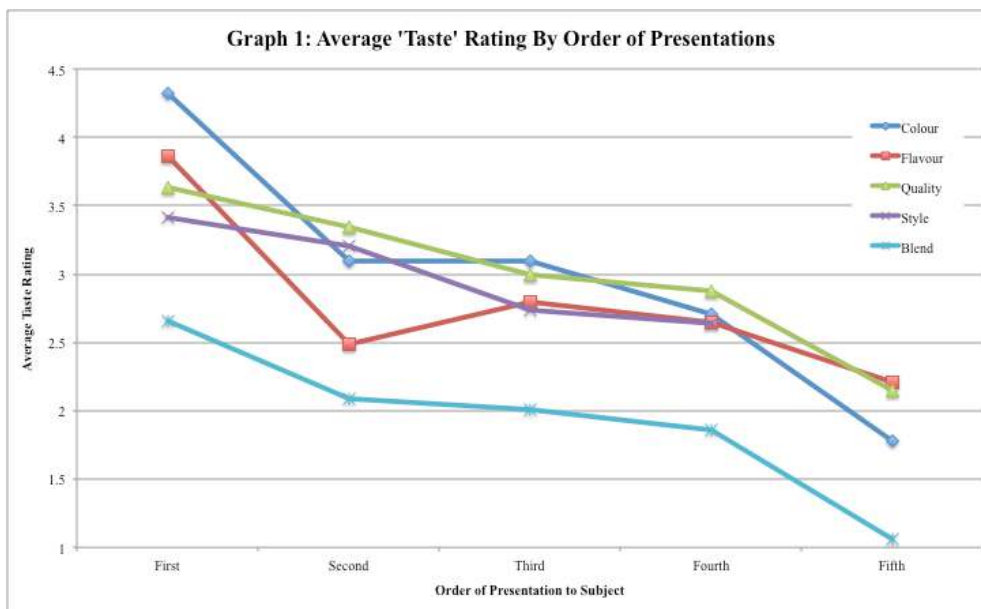
discussing their results: “Ratings of quality and taste were positively correlated in two of three comparisons (not for colour). Taste and tar/nicotine delivery were significantly correlated on all four comparisons, as were the two between taste and harm, and the only one tested between tar delivery and harm. In all cases, except for colour, the various ratings made by the same respondents on a set of descriptors were correlated. While a majority of respondents saw consistent differences, a minority (usually) saw none, and this varied by descriptor set (see tables 3–6)” (page 3).

4.73 To emphasise the importance of these order of presentation effects and the bias they potentially introduce, Graph 1 below simply shows the measure ‘taste’ against the order in which each “descriptor” is presented. The data is taken from Tables 2-4 in Borland and Savvas (2013) and the order in which each descriptor was presented to the subjects is given in the table below:

Order Seen by Subject	Colour (Table 2)	Flavour (Table 3)	Quality (Table 4A)	Style (Table 4B)	Blend (Table 4C)
First	Red	Full	Rich	Virginia	001
Second	Blue	Ultimate	Premium	American	002
Third	Gold	Infinite	Premier	Oriental	012
Fourth	Silver	Smooth	Distinct	Standard	333
Fifth	White	Fine	Refined		879

We see immediately the potential problem created by the structure of how the stimuli were presented. This point was also noted in paragraph 4.61 in my 2012 Report in discussing Borland et al. (2011) where I also discussed the statistical issues that arose from inferring conclusions using a repeated measures approach when all that was being revealed was that the subject was learning what the study was about and simply responding as they believed

they were expected to respond. What we see below in Graph 1⁵⁹ is that there is a direct relationship between the order in which the subject saw a descriptor and their rating of that descriptor (basically that the items seen first were rated “better” and those seen last were rated “worse”, with the intermediate descriptors being rated in a monotonically decreasing manner). Hence, it is impossible to determine whether it was the descriptors that mattered or the fact that the subjects simply used the order in which the descriptors appeared as an indicator about how they should “perceive” them.



4.74 Second, the study relies on impressions and opinions about items for which the individual has no knowledge – e.g., ‘quality’, ‘strength’, ‘tar’, ‘nicotine’, and ‘harmfulness’ – but on which the individual is queried (hence the potential for pseudo opinions generated by the nature of the study itself and something I note in the case of Hammond et al (2013) in paragraph 4.47 of this report). What is interesting in this case is that the authors

⁵⁹ Note that the data for ‘Blend Type’ (taken from row C1 of Table 4) was rescaled by taking (5 – the reported measure) to show the order effect on the same scale as the other descriptors more easily. In the case of ‘Blend’ the order effect was positive, meaning that individuals rated blends with higher numbers stronger (the effect, as is seen, is still an order effect). This actually reveals that an order effect may be the main driver as 4 of the effects are negative while one is positive, implying that individuals are cognitively attempting to respond in consistent manners based upon some heuristic.

acknowledge this fact but then go on to make their own inferences as to the cognitive model in use.

The most compelling conclusion, although hardly a surprising one from this study, is that given any set of attributes, people try to find order among them, and that some of this order is socially shared (i.e. they come up with similar judgments). For variant sets that are currently used, some of the commonality is likely due to some combination of industry communications and taste differences between the variants (page 6)

This would not be an issue except that the study is not designed to determine what that cognitive model is. Hence, any discussion as to what it is that the individuals are actually doing with the information they are presented with is simply pure speculation by the authors. One alternative is that the results are simply related to the order in which individuals saw the stimuli.

4.75 The approach to scale creation is incorrect and potentially masks variance (a point also noted as an issue for Borland et al. (2011) in my 2012 Report, and in paragraph 4.48 of my discussion of Hammond et al. (2013) in this report). As noted on page 3 of the study:

An overall mean for each descriptor in each rating was calculated by weighting as 5 points each ranking of most attractive, highest quality, strongest taste, or most harm; weighting as 1 point each ranking of the least attractive, lowest quality, weakest taste or least harm; and scoring all other cases 3 points (including cases where all were rated equivalent).

4.76 Statistically this reduces individual subject variance and potentially hides key effects. There is no logic for making this transformation, either statistically or theoretically. Its impact will be to inflate significance and allow extreme points to assume greater importance but doing so with the

need for fewer observations (i.e., a small number of individuals rating something at the extremes will inflate the overall apparent difference and significance).

4.77 Finally, the analysis is done in a “one at a time” manner that assumes complete independence of the results with broad conclusions being drawn from the application of a limited analytic perspective. The problem with this aggregation is discussed with respect to Borland et al. (2011) in my 2012 Report and with respect to Moodie, Ford et al. (2010) (see paragraph 4.85 in my 2012 Report). As noted there, the main issue is that the aggregation reduces variance and imposes artificial restrictions on the results. Hence, effects will appear to be more significant (since the standard error is reduced), leading to more statistical false positives, and will also be affected to a greater degree by outliers. In addition, analysis of variable importance in a one at a time manner can erroneously lead to what are known as “sunspots”. Sunspots arise when researchers assume that a relationship exists between A and B but the reality is that no such relationship exists.⁶⁰

Conclusion

4.78 This study concludes by noting that “*the results from this study show that all sets of pack-label variant descriptors systematically affected respondents’ judgements of the cigarettes associated with them, and this is so even when presented on plain, standardised packs. This is not surprising; however, it is of concern when it leads to erroneous inferences about relative harm*” and “*Indeed we suspect any variant set that induces inferences about relative strength is likely to also result in inappropriate inferences about harmfulness. We believe that variants should only be allowed where the differences are meaningful, and can be shown not to unduly create*

⁶⁰ Some explanations are seen in Jacobson, R. (1992), The ‘Austrian’ School of Strategy, *Academy of Management Review*, 17(4): 782–807; Plosser, C. and G. Schwert (1978), Money, Income, and Sunspots: Measuring Economic Relationships and the Effects of Differencing, *Journal of Monetary Economics*, 4: 637–660.

misleading perceptions of risk".⁶¹ However, one can draw no conclusions from this study as the methodological flaws are significant and the results are likely to be reflective of those flaws. In addition, the authors are drawing policy conclusions as to the cognitive and behavioural model being used by individuals without ever positing what that model (or alternatives) might be or testing that model directly.

Ramunno et al. (2012)

4.79 In this study, researchers examined visual attention to health warnings on existing branded packs and plain packs, as measured by eye movements with a small convenience sample of 28 adults, comprising weekly smokers ($n=7$), daily smokers ($n=9$), non-smokers ($n=8$) and ex-smokers ($n=4$). This study differed little from Munafo et al. (2011)⁶², which I evaluated in my 2012 Report in paragraphs 4.31 to 4.37. The authors conclude "*as expected, most time was spent looking at branding. Crucially, however, fixation times on warning labels increased in the unbranded compared with the branded presentations. This difference was not uniform across the experiment, but rather was significant only in the first pair of trials. The finding suggests that in contexts without recent habituation to warning labels (eg, first-time smokers) plain packaging helps to redirect the eyes towards text-only health warnings. However, the effect appears transient relative to that obtained with graphic warnings. Hence we suggest that a change to plain packaging might be most profitable if accompanied by a shift to front-of-pack graphic warning labels*". It should be noted that only a short abstract version of this study is publicly available.

⁶¹ See page 6, Borland and Savvas (2013).

⁶² Munafo, M., Roberts, N., Bauld, L. and U. Leonards (2011). Plain packaging increases visual attention to health warnings on cigarette packs in non-smokers and weekly smokers but not daily smokers. *Addiction*, Advance Publication. See also footnote 25 of this report.

Analysis

4.80 As noted with Munafo et al. (2011) the approach used in this study is more advanced methodologically and technologically than the other studies I have reviewed in the context of plain packaging. However, it suffers from all of the same limitations as Munafo et al. (2011), which I discussed in my 2012 Report:

- (a) First, the sample is small and based on convenience sampling. It should be noted that the sample of 28 subjects is even smaller than that of Munafo et al. (2011), which had 43 subjects. Eye movement studies require larger sample sizes and methodological approaches that account for a high rate of individual differences (or individual level variance driven by differences in how individual eye movements vary).⁶³ Unlike the authors of the current study, Munafo et al. (2011) clearly stated that their study was simply a pre-test. Hence, given this very small sample size the ability to make generaliseable policy statements is severely restricted.
- (b) Second, because of the small sample, the conclusions based on breakdowns of the various subgroups (e.g., the daily smokers vs. non-smokers) lack validity. For example, it is impossible to justify the conclusion that “*the finding suggests that in contexts without recent habituation to warning labels (eg, first-time smokers) plain packaging helps to redirect the eyes towards text-only health warnings*” as there is an insufficient number of first-time smokers to make statistically meaningful conclusions.

⁶³ A good example of a recent marketing/brand/advertising study using this method is Goodrich, K. (2011), “Anarchy of Effects? Exploring Attention to Online Advertising and Multiple Outcomes,” *Psychology & Marketing*, 28 (4): 417-440. This paper reveals the very small effect sizes found in such research but also the need to have both a clearly structured cognitive model and large sample sizes. That study used 100 subjects per cell and over 800 in total as compared to only 28 subjects here.

- (c) Third, the majority of the sample was female (only six were men), making the efficacy of any generalised statements about the results even more limited.
- (d) Fourth, and most importantly, there is no behavioural model that is either tested or informs us as to the meaning of the results. What this leads to is a series of ex post justifications and interpretations based on a very small and non-representative sample.

Conclusion

4.81 Although this study is an improvement technically over survey-based perceptions and attitudes, it suffers from sample size limitations and the lack of a theoretical behavioural model. Without additional information, we are not able to determine: (a) how that visual attention is being utilised and (b) whether that visual attention translates into anything practically meaningful. This leads me to conclude that the study does not provide reliable evidence as to whether plain packaging would help achieve the public policy goals outlined in paragraph 1.5 above.

Ford, et al. (2013)

4.82 This study examined survey based ratings and perceptions of five different cigarette pack designs by 1,025 “never smokers” who were aged between 11–16 years old in the United Kingdom. The study concluded: *“pack structure (shape and opening style) and colour are independently associated, not just with appreciation of and receptivity to the pack, but also with susceptibility to smoke. In other words, those who think most highly of novelty cigarette packaging are also the ones who indicate that they are most likely to go on to smoke. Plain packaging, in contrast, was found to directly reduce the appeal of smoking to adolescents”*.⁶⁴

⁶⁴ See page 1, Ford et al. (2013).

Analysis

4.83 Analytically, this study is econometrically more sophisticated than many other studies done on the subject of plain packaging. However, the sophistication of the econometrics is ultimately let down by the structure of the design of the task, the nature of the questions asked and their categorisations, and the distribution of the responses which ultimately invalidate the strong conclusions drawn by the authors.

4.84 The most significant structural issue with the study is the nature of the task itself. While one could argue with the validity of the specific pack designs chosen, that is not as relevant as the fact that all individuals saw all 5 pack designs all at once. This immediately creates a cognitive framing effect whereby the individuals attempt to structure consistent and logical responses to the stimuli they see. This is seen consistently in all studies of this type; where individuals immediately respond based upon very basic appearance factors and then infer what other responses are consistent with that. The most notable example of such effect is contained in my discussion of Borland and Savvas (2013) given in paragraphs 4.67 to 4.76 of this report.

4.85 The second major issue is related to the definition of “pack receptivity” and the meaning attributed to it by the authors. This is defined as an aggregation of four measures: “(6) *Not meant for someone like me/meant for someone like me*; (7) *Puts me off smoking/tempts me to smoke*; (8) *I dislike this pack/I like this pack*; (9) *I would not like to have this pack/I would like to have this pack*”. The authors conclude that this is (a) related to the willingness to take up smoking (susceptibility) and (b) that its relationship to pack design is causal.

4.86 Susceptibility in this study is “defined by the absence of a firm decision not to smoke” and was assessed with three items: “*never smokers were classified as non-susceptible if they answered ‘definitely not’ to the*

questions ‘If one of your friends offered you a cigarette, would you smoke it?’ and ‘Do you think you will smoke a cigarette at any time during the next year?’ and to the likelihood that ‘you will be smoking cigarettes at 18 years old’. Participants who answered anything other than ‘definitely not’ to any of the three items were classified as ‘susceptible’”. This amounted to categorisation of 28% of the sample as ‘susceptible’.

4.87 Thirdly, the model underlying this assumes that it is the relationship between the ‘attractiveness’ of the pack and the pack receptivity that matter and this, ultimately, is associated with susceptibility. However, the behavioural model for which this would be the case is not clear and the model that is estimated is behaviourally invalid. Susceptibility is ‘fixed’ and ‘exogenous’; that is, it is essentially a “trait” of the individual that will not vary with the experimental stimulus. In other words, because I am seeing 5 packs at once, no single pack is going to change my response to the questions related to susceptibility (they are ‘fixed’ in the sense that they have already been answered before the subject sees the stimuli and will not be able to be varied based on how the subject rates each pack on either receptivity or attractiveness). My susceptibility comes ‘before’ my rating of the packs and does not vary with them. Hence, it is ‘exogenous’ to the ratings.

4.88 However, the model as used in the study has this backwards, with susceptibility as a function of three factors – Block 1: Smoking Antecedents; Block 2: Demo/Sociographics; Block 3: Pack Receptivity – with Pack Receptivity being related to Pack Appraisal. Effectively this is:

$$\text{Susceptibility} = g(\text{Pack Receptivity, demo/sociographics, smoking antecedents}).$$

4.89 However, if ‘susceptibility’ is a ‘trait’ then its relationship to the measures such as ‘pack receptivity’ is unclear as the equation is implying that as pack receptivity goes up susceptibility goes up, but the discussion in the

study and the study design has causality going in the opposite direction. In other words, while the model makes logical sense for the demo/sociographics and smoking antecedents (which are essentially things known to cause susceptibility), it does not make sense for the relationship between pack receptivity (which is a random effect) which can in no way cause or precede susceptibility. Stated differently, susceptibility is viewed statistically as the ‘dependent variable’ that is influenced by pack design (an ‘independent’ variable) but the study design does not fit this structure. Hence, what we are seeing is a potential artifact in the results, which says that biases in measures of susceptibility are related to biases in measures of pack attractiveness and pack receptivity.

4.90 Econometrically, there are other issues related to the distribution in responses.

4.91 The number of individuals “receptive” to the packs is very small: *“Four per cent (n=35) indicated being receptive to the ‘regular’ Mayfair pack. For the ‘novelty’ packs, 5% (n=50) were receptive to Marlboro Bright Leaf, 6% (n=61) to Silk Cut Superslims and 7% (n=71) to Pall Mall. For the plain pack, 3% (n=27) indicated being receptive to this pack”*.⁶⁵ What is not clear is how many individuals are receptive to more than one pack – which is a critical piece of information, as we will see below.

4.92 The authors conclude: *“The GEE analysis showed that participants were significantly more likely to be receptive to the three ‘novelty’ packs compared to the ‘regular’ Mayfair pack. Participants were over 1.6 times as likely to be receptive to the Pall Mall pack (AOR=1.63, 95% CI 1.31 to 2.02, p<0.001, table 4), over 1.4 times as likely to be receptive to the Silk Cut*

⁶⁵ It should be noted that splitting the sample based on the median of the responses within the sample itself is generally not considered to be good practice and is uniformly frowned upon by statisticians. As noted by Maxwell, S. E. and H. D. Delaney (1993). Bivariate Median Splits and Spurious Statistical Significance, *Psychological Bulletin* 113 (1): 181–190, median splits generally will lead to a loss of power in statistical testing but when one relates one median split variable with another (as is being done in this study) there is a tendency to find spuriously significant effects.

Superslims pack (AOR=1.41, 95% CI 1.13 to 1.76, p=0.002) and over 1.2 times as likely to be receptive to the Marlboro Bright Leaf pack (AOR=1.27, 95% CI 1.03 to 1.58, p=0.027). There was no significant difference between the plain pack and regular Mayfair pack in terms of the likelihood of being receptive (AOR=0.85, 95% CI 0.68 to 1.07, p=0.172)".⁶⁶ However, what this does not emphasise is that most likely 80% of the respondents were not receptive to ANY pack; meaning that the comparisons that are being made are coming off a base in which the norm is to prefer none of the options.

4.93 Overall, 28% (n=286) individuals were rated as 'susceptible', and 733 not susceptible, while significantly fewer individuals were found to be 'receptive' to any of the pack designs. The key comparison being made in the study is a supposed analysis between individuals who are 'susceptible' and those 'receptive' to specific packs. The essence of the relationship being tested is given in the table below, which is a simplification of the logit model estimated and reported in Tables 5 and 6 of the study and discussed in the section entitled "*Association between pack receptivity and susceptibility*".

4.94 What a logit model that uses an experimental design approach does statistically is attempt to predict which individuals fall into which cells of a cross table. What we have below is a representation using a single table that represents the logit model estimated in this study pared down to just the essence of the analysis in such a cross table.⁶⁷ This is easy in this situation as the focal comparison is simply whether or not an individual is 'susceptible' (YES or NO) and whether they are 'receptive' to the pack (Yes = Y or No). Hence, what the researchers want to know is if you answer 'Y' to being 'receptive' to a pack do you answer 'YES' to being 'susceptible'? In other words, they want to know how many of the responses fall into the cells designated '1'.

⁶⁶ See page 5, Ford et al. (2013).

⁶⁷ Note that this representation includes more pack/brands than the authors use. As I note in paragraph 4.91, why the authors only analyse two brands is unclear and most likely related to their ability to estimate more complete and correct models.

4.95 However, if there is a relationship between susceptibility and receptivity to various packs it can show up in two ways. First, Receptivity = Y can be related to Susceptibility = YES (this is given by the cells designated with '1'). Alternatively, the relationship can be driven by Receptivity = No being related to Susceptibility = NO (this is given by the cells designated with '2'). What this means is that the fit statistics can be wholly related not to the ability to predict the relationship between Y and YES but the relationship between No and NO. In other words, if I have a lot of individuals who are not receptive to a pack (and hence answer No) and they are also not susceptible to smoking (and hence answer NO) then my model has made a "correct" prediction, independent of whether or not I have predicted that those that answer YES and Y are matched well. This essentially means that I can have what appears to be reasonably well fitting model not because I am filling in the '1' cells correctly but because I am filling in the '2' cells correctly. The reason that this occurs is that 'incorrect' predictions are those represented by the cells designated as '3' ([YES, No] and [NO, Y]). Hence while the authors believe that what they are doing is comparing the cells designated as '1' to '2' and '3' the model as specified is actually comparing the cells designate '1' and '2' to the cells designated '3'.

		Receptive to:												
		Mayfair Regular		Novelty Slim		Silk Cut		Marlboro Bright Leaf		Pall Mall		Plain		Nr
		Y	No	Y	No	Y	No	Y	No	Y	No	Y	No	
Susceptibility	YES	1	3	1	3	1	3	1	3	1	3	1	3	286
	NO	3	2	3	2	3	2	3	2	3	2	3	2	733
Number		35	984	50	969	61	958	50	969	71	948	27	992	1019

This is made worse when the distribution of individuals appearing in the cells is heavily skewed toward specific responses. And this is exactly what we see

in this study. The distribution of responses that we can determine from the data in the study is that there are very few individuals likely to be in the cells represented as '1' and a much greater probability that there are individuals in the cells marked '2' (you can see this by just noting that there are more individuals answering "No" and NO (these are the numbers in the last row and last column)). This is why it is important not just to report the parameter estimates and general 'fit' statistics but also what the 'hit' rate (the number of correct cells predicted – or more simply, the number of times when the model predicts a Yes/Y relationship and the response by the subject is Yes/Y); the hit rate is not reported. In other words, what is important to know is the number of individuals "predicted" to be in each of the 24 cells and the number that actually appear in those cells.

4.96 Finally, the analysis in Tables 5 and 6 of the study (which relate to the "Logistic regression of association between susceptibility to smoke and pack appraisal of the 'novelty' Marlboro Bright Leaf pack" and "Logistic regression of association between susceptibility to smoke and packaging receptivity to Silk Cut Superslims") are extraordinarily curious as it is not clear why they chose to run two separate models instead of simply regressing all of the pack designs against susceptibility. A proper model would have included all of the packs as this would have provided information not just about the relationship the authors are discussing but would also have allowed them to align the models in Tables 5 and 6 with those given in Tables 3 and 4. Given that there is no logical reason to estimate these equations for the single pack type separately (note that there is no real difference between the structure of Tables 3 and 4 and that of Tables 5 and 6), the issue must be one related to the ability to estimate the equations (in other words, it is most likely that the full model that includes all the packs simply did not converge and was unestimable). Referring again to the table above one can guess that the reason that the models in Tables 5 and 6 exclude four of the other pack types is because the models cannot be estimated due to the lack of valid

responses in the cells marked ‘1’ above. Hence, what the authors are doing is attempting to generate models when the data is simply not available to estimate those models correctly and efficiently. It is also not clear why, on page 5 of the study, the authors indicate that 141 individuals were positively receptive toward Marlboro Bright Leaf but in Table 5 this number is 140 and that for Table 6 the number positively receptive toward Silk Cut was 58 while on page 5 the number is given as 61.

Conclusion

4.97 This study concludes by noting that it “*provides the first direct evidence that the attractiveness of cigarette packaging is associated with susceptibility to smoke. Differences among the packaging styles highlight the influence of innovative and unique branding elements on adolescents’ future smoking intentions*”.⁶⁸ However, this is an invalid and erroneous conclusion. The study does not relate ‘attractiveness’ to susceptibility either directly or causally. Indeed, it confuses the relationship by utilising susceptibility as a dependent variable when logically and statistically this is incorrect. In addition, because the study does not actually relate the different package style to changes in ‘smoking intentions’ directly, the second conclusion relating branding to minors’ future ‘smoking intentions’ is also invalid and potentially spurious.

Dewe et al. (2013)

4.98 This study analysed 204 tobacco advertisements in the United Kingdom spanning 1950 to 2009. The study authors conclude that their work provides “*some evidence in support of the call for compulsory plain packaging as the box cannot be considered a neutral object that has no impact on consumer choices. Instead the box can be viewed as a form of branded packaging that is oriented towards persuading the target audience*

⁶⁸ See page 9, Ford et al. (2013).

*to purchase the product. After the recent bans on advertising, the box continues to function as a mobile marketing tool that is presented to potential consumers by existing consumers as they smoke their cigarettes. The introduction of compulsory plain packaging of cigarettes would eliminate this advertising medium and its persuasive function”.*⁶⁹

Analysis

4.99 Overall, the dominant problem with this study is that its conclusions are unsupported by the study’s results. The authors make significant and overarching conclusions about topics, processes and activities for which there is no evidence presented and for which the study was never structured to investigate.

- (a) First, the authors of the study discuss that the role of advertising is to ‘persuade’ the individual to purchase and consume the product (e.g., *“theories of persuasion provide insights into how changes in tobacco companies’ advertising foci may have oriented towards influencing consumption behaviour”* and *“the focus over time on specific key positive features that were ascribed to cigarettes can be seen as a refinement of persuasion endeavours.”*). However the study does not measure or investigate ‘persuasion’ in a direct manner. All that it does is examine what was emphasised in the advertisements selected and then makes inferences as to what the marketing strategy of the firm was. In addition, while the study argues that persuasion and consumption are linked (e.g., *“Ultimately persuasive advertising leads to the consumer forming positive associations about specific products or brands which promote consumption.”*) individual level purchasing behaviour is not examined and no link between these key constructs is established. Hence the statement *“the box can be viewed as a form of branded packaging that is oriented towards persuading the target*

⁶⁹ See page 9, Dewe et al. (2013).

audience to purchase the product” has no support, as it is not what is investigated. Nor does the statement “the introduction of compulsory plain packaging of cigarettes would eliminate this advertising medium and its persuasive function.”

- (b) Second, concerning the conclusion that *“after the recent bans on advertising, the box continues to function as a mobile marketing tool that is presented to potential consumers by existing consumers as they smoke their cigarettes”* no evidence is presented. There is no data presented to support the conclusion and even the information given in Figure 2 shows that there is no information about the package beyond the year 2000. Hence the authors give no evidence that justifies the statement that *“the box continues to function as a mobile marketing tool”*.
- (c) Third, the conclusion that the study *“provides further evidence that the cigarette box is indeed an advertising vehicle in its own right”* is not supported by any data in the study. It did not investigate the box as a singular piece of advertising and its effectiveness as such.
- (d) Fourth, the conclusion that the study’s *“results also indicate that the meanings associated with cigarettes were transferred from the cigarette onto the box in the years prior to 2003 ban on print advertising, particularly those meanings associated with quality”* is not supported by any data in the study. The research did not examine the content of the box nor did it investigate how this process of transfer occurred, if it did indeed occur.
- (e) Finally, the relevance of the study to current policy is unclear. How important the content of cigarette advertising in the 1950s, 1960s, 1970s, 1980s, and 1990s, when various forms of tobacco advertising

were permitted in the United Kingdom, to current policy debates around plain packaging is never addressed by the authors.

4.100 There are two significant methodological flaws evident in this study.

- (a) First, while the authors discuss how advertising “might” work in generating ‘persuasion’ and consumption (the theory suggested by the authors at page 3 of the study), there is no behavioural model as to how they posit advertising “does” work in this case nor is there any relationship between the analysis they conduct and a behavioural model of advertising persuasion.
- (b) Second, the analysis aggregates the results arbitrarily into decades. While the authors argue that this is justified by a citation to Anderson et al. (2006), there is no logical, practical or scientific justification for such an aggregation. Good process would require that the advertisements were selected by year and only if there were consistencies across the years would aggregation be justified. In addition, by aggregating the results by decade there is no way to determine if the content of the adverts used were influenced where they appeared in the decade. A simple solution would be to do a robustness check on the aggregation by picking different 10 year periods to determine if the different groupings led to significantly different interpretations. If they did, then the aggregation is generating bias and should not be done.

Conclusion

4.101 This study represents a simple content analysis of tobacco advertising during a certain period by manufacturers, and therefore presents no primary research in respect of plain packaging. The study’s major limitation is that it attempts to draw policy conclusions on plain packaging without ever examining this process directly. In addition, it makes assumptions as to the

actions of the advertisers without any evidence as to what it was that the firms were attempting to achieve. However, what is most intriguing (and somewhat ironic) about the study is that while the authors argue that the cigarette box is a “powerful advertising vehicle”, they ultimately have to conclude, on the basis of their study, that “*changes in advertising and the meanings evoked were unrelated to changes in smoking behaviour*”.⁷⁰

Lund and Scheffels (2013)

4.102 This study is an online survey of 1,022 15 – 22 year old smokers and non-smokers in Norway. The stated aim of the study is to determine the degree to which plain versus branded packs are associated with positive images and characteristics. Significant, and in my opinion, unnecessary emphasis is put by the authors on the differences found in results between boys and girls (which I discuss further below).

Analysis

4.103 This study is a basic standard variant of many other studies examined in this report and my 2010 and 2012 Reports.⁷¹ Each survey participant evaluated a mixture of branded and plain packs and indicated whether or not, in their opinion, a positive characteristic applied to the ‘typical’ smoker of that pack.

4.104 There are three major design and method issues with this study:

- (a) First, the structure of the study makes it clear to the subjects that the intention of the study is to identify user images related to pack variation and that some of the variants are plain packs that currently do not exist in the market. Hence, the individuals assigned to the plain

⁷⁰ See page 1, Dewe et al. (2013).

⁷¹ See, e.g., Gendall et al. (2011) and GfK Bluemoon (2011), which did brand association tasks, and McCool, Webb et al. (2012), which was a focus group looking at what images individuals associated with smoking. These studies are all discussed in my 2012 Report.

pack conditions are most likely to notice that: (i) they are being asked to evaluate a radical change in the market presentation of a product; and (ii) the study is specifically about smoking and, therefore, related to a product category with specific regulatory connotations. As with every study I have evaluated in this report, and my 2010 and 2012 Reports, there is no attempt to blind the subject to the intent of the study. This potentially leads to a conscious desire on the part of the subject to look for differences in the pack variations and to make those differences fit with a logical cognitive model that they believe is compatible with the intent of the study.

- (b) Nearly 80% of the subjects are non-smokers. Smokers were defined as “*those who had smoked at all during the last 30 days*”⁷², a different definition to those used in many other studies, making the comparability with other studies using different definitions of smokers difficult. Also, given that non-smokers, as the vast majority of the sample, have no direct purchasing experience of the brands being investigated, the bias referred to in paragraph 4.104(a) above is exacerbated by the fact that the individual has no or little practical experience on which to base their judgements. In these circumstances, it is likely that the subjects will utilise a heuristic⁷³ to determine their evaluation of the packs and that what is being picked up statistically is that heuristic, and it is that heuristic that generates pseudo opinions as to the characteristics of the typical smoker of the

⁷² See page 7, Lund and Scheffels (2013).

⁷³ Heuristics are “simple, efficient rules, learned or hard-coded by evolutionary processes, that have been proposed to explain how people make decisions, come to judgments, and solve problems typically when facing complex problems or incomplete information” (<http://en.wikipedia.org/wiki/Heuristic>). They have been extensively studied both for their bias (e.g., Kahneman, D., Tversky, A., and P. Slovic, eds. (1982) *Judgment under Uncertainty: Heuristics & Biases*. Cambridge, UK, Cambridge University Press) and efficiency in making decision making quicker (e.g., Gigerenzer, G., et al. (1999). *Simple Heuristics That Make Us Smart*. Oxford, UK, Oxford University Press)

pack.⁷⁴ This may partially explain the difference between the responses of boys and girls in the study, as the heuristics used may be gender based.

- (c) The break used by the authors of the study to determine when an image generated by a cigarette pack is ‘positive’ is whether or not there is “at least” one positive user characteristic identified by the subject. This is arbitrary. It is also quite surprising, particularly given that there are potentially seven positive characteristics – ‘boy/girl’ (gender), ‘glamorous’, ‘stylish’, ‘popular’, ‘cool’, ‘sophisticated’ and ‘slim’. As noted by the authors, “*a relatively small proportion of the respondents assigned 2 or more positive characteristics to each pack*”.⁷⁵ What this reveals is a heavily skewed distribution where rather than the subjects having a wide degree of consensus as to how a pack user is perceived, the images are extremely random and generally are not positive at all. If the authors decided that 2 positive characteristics were necessary for a positive image, it is highly likely that none of the packs would come out as being rated ‘positive’. It is also the case that the regressions reported in the section of the study headed ‘*Pack comparisons*’ (preceding the ‘Discussion’ section) will essentially be meaningless as the differences that are being evaluated are very small. Hence, while the results might be ‘significant’, they will be effectively immaterial from the standpoint of the magnitude of the effects. They will also be highly sensitive to outliers (in other words, extreme responses by specific subjects) as the norm is to not evaluate any of the pack options given positively at all.

⁷⁴ See, e.g., Tourangeau, and N. Bradburn (2010). The Psychology of Survey Response, in Marsden, P. and J. Wright (eds.), *Handbook of Survey Research*, Bingley, UK: Emerald, 315–346.

⁷⁵ See page 7, Lund and Scheffels (2013).

4.105 The study focuses very heavily on the differences between the responses of boys and girls to the branded and plain packs. It is not clear why this should be the focus of the study. However, given that it is, the authors are remiss in not accounting for the fact that there are well known survey response differences between the genders. For example, generally, women are more likely to be willing to respond to surveys – meaning that it is necessary to contact fewer women to meet the sampling quota (i.e., a balance in the gender of the subjects in the sample).⁷⁶ Women are also known to differ from men in how they respond to survey questions and where their sources of bias arise.⁷⁷ For example, it is well recognised that women and men answer “don’t know” questions differently and this is generally a serious issue when knowledge is lacking (as outlined in the political arena by Mondak and Anderson).⁷⁸ The latter point is critical for this study as the sizeable number of non-smokers (79.5%)⁷⁹ may mean a significant proportion of responses were influenced by a lack of knowledge. The arbitrary use of the “no”/“don’t know” aggregation further inflates the error by conflating two different cognitive responses. In addition, men are more likely to be certain about their answers⁸⁰ and engage in more random guessing when they lack knowledge than women⁸¹ (a fact that might explain the lack of significant results for boys). Finally, there are gender differences in responses to “sensitive” questions and smoking behaviour is normally

⁷⁶See, e.g., Stoop, I. (2004). Surveying Nonrespondents, *Field Methods*, 16 (1): 23–54; Boulianne, S. (2012). Examining the Gender Effects of Different Incentive Amounts in a Web Survey, *Field Methods*, 25 (1): 91–104.

⁷⁷See, e.g., Rapoport, R. B. (1982). Sex Differences in Attitude Expression: A Generational Explanation, *Public Opinion Quarterly*, 46 (Spring): 86–96.

⁷⁸Mondak, J. and M. Anderson (2011). The Knowledge Gap: A Reexamination of Gender-Based Differences in Political Knowledge, *Journal of Politics*, 86 (2): 492–512.

⁷⁹ See page 7, Lund and Scheffels (2013).

⁸⁰See, e.g., Hirschfeld, M., Moore, R. and E. Brown (1995). Exploring the Gender Gap on the GRE Subject Test in Economics, *Journal of Economic Education*, 26 (1): 3–15.

⁸¹ See, e.g., Kenski, K. and K. Jamieson (2000). The Gender Gap in Political Knowledge: Are Women Less Knowledgeable Than Men about Politics? In K. Jamieson (ed.), *Everything You Think You Know About Politics*, New York: Basic Books, 83–89; Frazer, E. and K. Macdonald (2003). Sex Difference in Political Knowledge in Britain, *Political Studies*, 51 (1): 67–83.

considered a sensitive issue, particularly as it pertains to adolescents.⁸² All of this implies that by failing to account for the natural gender based differences and biases in survey response, the authors of the study may simply be attempting to rationalise results that are nothing more than reflective of gender bias.

Conclusion

4.106 While this study argues that branded cigarette packs are associated with positive images, the results are driven dominantly by how a ‘positive’ image is measured and defined in this study. In addition, the results themselves reveal a large degree of heterogeneity in the subjects’ responses. While the authors argue that “*the large variability in how different brands were evaluated indicates that cigarette packaging is important for young people’s differentiation between, and possibly also identification with, brands*”,⁸³ it is also possible that this variability simply reflects the fact that different individuals used different points of reference to come up with their responses to the task because they did not have strong evidence or experience on which to base them. The conclusion that, “*exposure to branded packs generally resulted in a more positive characterization of smokers*” would follow logically from this reasoning not in the sense that the brands really matter but that absent any other information, uninformed individuals will come up with what they consider to be a consistent rule (a heuristic) for generating their responses. And, as noted, this rule can be biased by the structure of the study, the stimuli presented to subjects, and the questions asked of them. In addition, the fact that responses by girls are very different from those given by boys may be hinting at the fact that girls are, in general, simply more consistent in the rule they use to respond to the survey.

⁸² See, e.g., Tourangeau, R. and Y. Ting (2007). Sensitive Questions in Surveys, *Psychological Bulletin*, 133 (5): 859–883.

⁸³ See page 10, Scheffels and Lund (2013).

4.107 In the end, the authors draw policy conclusions for which their study has no evidence. They conclude, *“It is not unlikely that introducing plain packaging might have some unintended effects that could undermine the effect of the policy. Smokers may decide to import more cigarettes legally or illegally, producers of cigarette cases could potentially see plain packaging as a new business opportunity, and the tobacco industry may counteract the similarity of the packs by changing the design of the cigarettes. However, most of these effects would in all likelihood be short term, while the positive effects of plain packaging could be expected to increase as both smokers and potential smokers start to put their outmoded ideas about user characteristics behind them in the medium to long term both within and across makes”*.⁸⁴ None of these points were examined in the study and hence none of the policy conclusions being made here can be considered as based on evidence from the authors’ investigation.

Scheffels and Lund (2013)

4.108 This study consists of two online surveys of 1,010 15–22 year old smokers and non-smokers in Norway. The stated intent of the study is to determine the degree to which plain versus branded packs generate more positive responses on ratings of perceived ‘appeal’, ‘taste’ and ‘harmfulness’ in the first survey, and ‘taste’, ‘harm’, ‘quality’, ‘would rather try’ and ‘easier to quit’ in the second survey.

Analysis

4.109 This study is a basic standard variant of many other studies examined in this report and my 2010 and 2012 Reports.⁸⁵ Each subject evaluated a mixture of plain or branded packs and responded to a series of survey

⁸⁴ See page 12, Lund and Scheffels (2013).

⁸⁵ These include, Hammond, White et al. (2013), Hammond et al. (2013), Borland and Savvas (2013), Ford et al. (2013), Hammond, Doxey et al. (2011), Gallopel-Morvan et al. (2011), Moodie, Ford et al. (2011), Hammond et al. (2009), Germain et al. (2009), Doxey (2009).

questions about each pack. In the first survey, they responded to 12 packs individually; in the second survey they responded to five pairs of packs.

4.110 The relationship between Lund and Scheffels (2013) and Scheffels and Lund (2013) is not clear. There are some apparent similarities between the studies: both use the same research company (TNS Gallup) to recruit the participants, both have nearly identical sample sizes (1022 vs. 1010) and the splits between smokers and non-smokers, and gender is again nearly identical. What is not stated, but would be important to know, is whether individuals from one study participated in the other study as this could potentially bias the results of one or the other (depending on which came first).

4.111 There are five major issues with the design and conclusions of the study. Most of these issues are identical to those that have arisen in other pack comparison studies I have evaluated.

- (a) First, as with all of the studies considered in this report, there is no behavioural model being investigated. The hypotheses given on page 6 of the study test no theory or model, and amount to little more than hypotheses of convenience. It is interesting to note that, once stated, these ‘hypotheses’ are never referred to again. The relationship between the tasks being asked of the subjects and the cognitive process that leads to the choice being investigated by the study is never articulated.
- (b) Second, there is once again an issue related to the fact that the purpose of the study is to investigate brand pack variations. As the vast majority of those taking part in the study (nearly 80%) are non-smokers, they have or are likely to have no or little direct purchasing experience on which to base their comparisons or their evaluations of the packs. As noted in my discussion of Lund and Scheffels (2013) at

paragraph 4.104(b) above, this leads the subject to utilise heuristics to generate their responses and the results are then likely to represent what those biased heuristics are rather than anything related to what amounts to realistic market based behaviour.

(c) The individual survey responses are arbitrarily aggregated. This aggregation is flawed in three ways.

(i) First, like many studies I have evaluated (e.g., Hammond, Doxey et al. (2011)), the authors attempt to turn nearly every response variable into a binary variable. For example, in this case the authors aggregate variables that have “no” and “don’t know” responses into a single “no/don’t know”. As I noted at many points in my 2010 and 2012 Reports (e.g., sections 4.22 – 4.23 of my evaluation of Hammond, Doxie et al. in my 2012 Report), such aggregation has no statistical justification and treats individuals who have no information (‘don’t know’ response) identically to those that know that “no” is the response they want to give.

(ii) Second, the authors attempt to create indexes by then aggregating aggregated variables. For example, in creating their ‘health risk awareness’ index the authors first asked subjects “*whether they knew that smoking could cause 12 different diseases ... [where the possible] response options were: yes, no, and do not know. [The no and don’t know responses were treated identically and] all positive answers were summed together to create a health risk awareness index. For the logistic regression analyses, the index was recoded into a variable with three values (0–4, 5–8, 9–12)*”.⁸⁶

⁸⁶ See page 6, Scheffels and Lund (2013).

Such aggregation upon aggregation makes understanding and interpreting the results nearly impossible.

- (iii) The authors then use these aggregated variables in comparison to other aggregated variables. As noted previously, the aggregation process removes information about individual variability and can lead to spurious results (see paragraph 4.78 and footnote 41, as well as paragraphs 4.22 and 4.33 of my 2012 Report). Unless there is a logical and theoretical reason for the aggregation (rather than simply attempting to get the data into a form amenable to an analysis that the authors find convenient) such aggregation is unjustified scientifically.

- (d) In the choice of the packs to evaluate and compare in the study, there is no meaningful experimental design. As noted many times in this Report and my 2010 and 2012 Reports, it is important when making comparisons based on the attributes of a stimuli that the range of the comparisons be such that they allow for meaningful interpretation – this being ensured when the criteria of efficiency, orthogonality and balance are met (see paragraph 2.10 for a discussion).

Looking at the brands in Table 2 on page 5 of the study, – which are related to the individual pack evaluation survey – Prince and Marlboro appear 3 times each, Lucky Strike twice, and all other brands once. We then have the word ‘Taste’ appearing 4 times, the words ‘Gold/Golden’ 3 times (twice for Marlboro) and the word ‘Original’ 3 times (twice for Lucky Strike) and so on. What we see in this is that the range of potential pack attributes being considered is vast, unbalanced, non-orthogonal and statistically inefficient. As structured, this means that the responses to the packs cannot be meaningfully evaluated (or must be done with extreme caution and in a manner that accounts for the fact that the stimuli is varying in a non-

systematic way) and drawing conclusions as the authors do is fraught with immense difficulties.

- (e) Looking at the second survey – the paired comparison exercise – we see that there is no real design at all. The pairs of packs used are “*Prince Rich Taste versus Rounded Taste, Marlboro Red versus Gold Original, Kent Original versus HD Taste System, Lucky Strike Original versus Blue and Petteroes Original versus Lys blå (Pale blue)*”.⁸⁷ How this fits into any experimental design is unclear. For example, how is one to know how to evaluate the difference between Prince Rich Taste versus Rounded Taste and Lucky Strike Original versus Blue when every descriptor is different? Again, this is a simple balance and orthogonality issue. To be able to understand the meaning of the responses, the options presented to the subjects need to be structured so that the components of the pack that drive the decision are identified clearly. This can only be done when the features on the pack are varied consistently using well-understood design principles as outlined in paragraph 2.10 of this report.

4.112 Interpreting the results of this study is fraught with difficulties given the poor design. If we examine Tables 1 and 2 on pages 4 and 5 of the study, which aim to assess the extent to which the branded pack is statistically significantly different from the plain pack on ‘appeal’, ‘taste’, and ‘harm’ (the numbers denoted with asterisks in the tables), we see that of the 102 possible results, only 21 are statistically significant for females (20.5% of cases, Table 1) and 14 for males (13.7% of cases, Table 2). In the case of the males, half of the packs show no statistically significant effects, while for females one quarter of the packs show no statistically significant effects. If we look at how many packs show more than one statistically significant effect across the evaluation criteria (e.g., ‘appeal’ and ‘taste’ or ‘taste’ and

⁸⁷ See page 3, Scheffels and Lund (2013).

‘harm’), only one quarter of the packs do so for males, while half of the packs do so for females. In essence, the effects seen in the first survey are much less clear cut and strong than the authors argue in the study.

4.113 Table 3 on page 7 of the study presents the results in a manner comparable to Lund and Scheffels (2013) (see Table 2 on page 9). Again, we see that rather than the branded packs having strong positive effects, all that is revealed is that all of the packs have relatively low positive responses and that ‘significance’ is driven not by the magnitude of the effect but by the sample size.

4.114 In addition, when examining the data in Table 4 on page 7, we see yet another confused set of results. According to the authors’ hypothesis 2: “*In a direct comparison of ‘regular’ and ‘lighter’ packs from the same brand family (eg, Marlboro vs Marlboro Gold), the lighter pack will more often be perceived as more appealing, better tasting and less harmful in the branded condition than in the plain (with descriptors only) condition.*”⁸⁸ However, the results show that lighter packs were, contrary to the hypothesis, no better on ‘taste’ ($\beta = -0.19$, not significant), more ‘harmful’ ($\beta = -0.77$, significant), less likely to want to ‘be tried’ ($\beta = -0.32$, significant), harder ‘to quit’ ($\beta = -0.58$, significant), and no higher in ‘quality’ ($\beta = 0.04$, not significant) than the regular packs.

4.115 Finally, it is simply noted that all of the gender response bias issues that were discussed with respect to Lund and Scheffels (2013) in paragraph 4.105 above apply equally to this study.

Conclusion

4.116 Overall, this study lacks validity due to its weak design and methodological flaws in the aggregation and analysis of the data. The study

⁸⁸ See page 6, Scheffels and Lund (2013).

concludes by arguing that “*the results ... point to how packages communicate messages that allow consumers to identify with and differentiate between cigarette brands, and thus are essential in the processes branding works through*”.⁸⁹ However the study does not evaluate communication, messages, consumer identification or differentiation nor the process of branding. The authors further conclude that “*respondents so clearly make distinctions regarding harmfulness and ease of quitting between the brand varieties based upon colours and descriptors confirms the findings from a previous qualitative research in Norway and points towards the conclusion that cigarette descriptors such as ‘rounded taste’ (in contrast to ‘rough taste’) and colour codes such as ‘gold’ or ‘pale blue’ are perceived in a similar way as the prohibited terms ‘light’ and ‘mild’*”.⁹⁰ However, as discussed above in paragraph 4.111, the structure of the study means that it is impossible to draw conclusions. In addition, as noted in paragraph 4.112, the results presented point to a conclusion that is exactly the opposite.

Focus Group Studies

4.117 In this section, I review a series of qualitative focus groups studies together. These studies are (a) Ford, Moodie, et al. (2013), (b) Van Hal et al. (2012), (c) Uppal et al. (2013) and (d) Edwards et al. (2012). This is being done simply because it is more efficient to outline the limitations of these studies together as they all apply the same basic methods and the limitations of focus groups (see paragraph 4.118) are common to all. In addition, my 2012 Report evaluated more than half a dozen focus group studies and the reader can see in those examinations the fundamental problem with drawing any conclusions based upon small scale, qualitative focus groups.⁹¹

⁸⁹ See page 8, Scheffels and Lund (2013).

⁹⁰ *Ibid.*

⁹¹ The studies examined in my 2012 Report were: Moodie, Hastings, et al. (2011) – paragraph 4.38-4.43; Moodie & Ford (2011) – paragraph 4.108-4.113; Hoek et al. (2012) – paragraph 4.114-4.118; McCool, Webb, et al. (2012) – paragraph 4.119-4.125; CRUK (2012) – paragraph 4.135-4.140; Carter and Chapman (2006) – paragraph 4.141-4.146; and GfK Bluemoon (2011) which included 6 focus group studies – paragraph 4.147-4.169.

4.118 As I noted in Table 2 at paragraph 2.12 (and in my 2010 and 2012 Reports): *“focus group studies are exploratory. They generate hypotheses rather than findings that can be generalised to a wider population. The reported findings of focus groups often have no statistical significance due to the small sample size and informal nature of the responses”*. Hence, in conducting focus groups it is important to recognise and identify clearly that such studies are meant to be part of a larger stream of research in which they are effectively inductive information exercises. One cannot draw conclusions based on focus groups. All that one can do is utilise the information gathered to better formulate more scientifically valid studies that follow on from the limited information gathered. Hence, overall, no policy conclusions can be drawn from focus group research, no matter how sophisticated they appear to be on the surface.

4.119 **Ford, Moodie et al. (2013)**. This study is a focus group study of 48 15 year olds split into eight groups of six individuals each of whom were shown and discussed eight different cigarette designs. The intent of the study was to examine reactions to different cigarette designs. The study argued, *“the exploratory study provides some support that standardising cigarette appearance could reduce the appeal of cigarettes in adolescents and reduce the opportunity for stick design to mislead young smokers in terms of harm”*.⁹² However, it is flawed in the following ways:

- (a) Overall, it is difficult to conclude what is the point of this study, as it is not clear what the next more scientifically valid stage is that follows on from what was done. As noted in paragraph 4.118 above, the point of focus groups is not to draw definitive conclusions but “to be part of a larger stream of research in which they are effectively inductive information exercises.” In doing this, it is important to identify the larger stream of work into which the study fits.

⁹² See page 1, Ford, Moodie et al. (2013).

- (b) What this study does reveal is that when asked to make comparisons amongst cigarettes that varied in terms of design, individuals responded by discussing differences in design and spent their time attempting to draw inferences that explained those differences. This is evident in the conclusion that “*we found that cigarette appearance can generate significant interest among adolescents*”.⁹³ Given that the point of the study was to generate discussion about the stimuli presented, this conclusion is hardly surprising as: (a) it is highly unlikely that a trained facilitator would not work hard to generate the expected discussion; and (b) the novelty of the stimuli prompts the subjects to want to discuss that novelty. The latter point is particularly germane since the subjects themselves have no real depth of experience with the product category (e.g., “*participants were surprised with the amount of variation in cigarette appearance*”⁹⁴). When this occurs the tendency for individuals to generate rationalisations, pseudo opinions and to try to formulate their discussion in a manner that conforms to the expectations of the research are exacerbated.
- (c) The conclusion that, “*intrinsic clues, such as colour, length, diameter, and decorative features easily communicated messages and images related to gender suitability, price, glamour and coolness*”⁹⁵ oversteps the validity of the study. The authors are making definitive conclusions and causal links where the design and structure of the study simply does not justify such conclusions. This tendency is seen in other statements where there is no way to justify them other than as speculation. For example, the authors state that, “*These differences may suggest that adults and adolescents prefer and place importance*

⁹³ See page 4, Ford, Moodie et al. (2013).

⁹⁴ See page 3, Ford, Moodie et al. (2013).

⁹⁵ See page 4, Ford, Moodie et al. (2013).

on different features of cigarette design”.⁹⁶ Yet, the study included no adults on which to base this comparison. In addition, the study goes on to make strong conclusive statements, inferences that the research “supports” scientific conclusions, and definitive policy recommendations that cannot be justified based on a sample of 48 Glasgow adolescents (emphasis added): “*The findings show that differences in cigarette appearance can generate interest, provide novelty, communicate positive imager and mislead adolescents in respect to product harm. While **providing support** for the ban on branding, colour and other decorative elements on cigarettes in Australia, **this study suggests** that standardising diameter could further reduce the opportunity for the tobacco industry to communicate with and influence young people. In this respect, the fact that the European Parliament has voted against the proposed ban on cigarettes less than 7.5mm in diameter would appear a **missed opportunity** for improving tobacco control policy in Europe*”.⁹⁷ None of this is justified in the context of small sample non-representative focus group.

4.120 **Van Hal et al. (2012).** This study is a focus group study with 55 15–18 year old Flemish adolescent smokers. The intent of the study was to examine opinions and perceptions of plain pack cigarettes. Individuals were shown and discussed two different cigarette packs; one a Marlboro pack and the other a plain pack variant. The study concluded that “*in this first scientific study in Flanders on this topic it emerged that plain packaging could be a strong policy tool to reduce the number of adolescents starting smoking*”.⁹⁸

⁹⁶ *Ibid.*

⁹⁷ *Ibid.*

⁹⁸ See page 1, Van Hal et al. (2012).

- (a) It is difficult to justify this study at all. The authors argue that “*this qualitative research [focus group discussion] method is very appropriate when exploring opinions, attitudes and experiences*”.⁹⁹ The key here is the word “exploring”. According to Stewart et al. (2007),¹⁰⁰ the foundation of focus group research is “focus” – the fact that the investigator is interested in ‘focusing’ in on individuals who have “*experienced some specific concrete experience*”. The researcher wants to get access to this experience and learn something inductively from it, as a means of building and enhancing theory and then to go on and test it in a more comprehensive and scientifically valid manner. Stewart et al. contrasts this with the more invalid form of ‘out-of-focus’ research that “*arise from pragmatic concerns about research time and money as well as [a] generally diminished interest in theoretical and methodological subtleties.*”¹⁰¹ Hence, the validity of the study needs to be justified by showing how the experiences being probed by the focus groups are so unique that it adds to what has already been investigated in the existing literature. In some sense, the authors of this study attempt to do this by arguing that “*although a lot of evidence in different continents is available concerning the potential of plain packaging to influence smoking behaviour in adolescents, cultural and local factors might still play a role.*”¹⁰² However, the study does not compare across cultures (it is a within Flemish study), hence those cultural and local factors are never identified and accounted for in the study design. In addition, the protocol and design are not structured to bring out what those cross-cultural differences might actually be. For example, one area where package design might differ is in inferences that individuals make

⁹⁹ See page 2, Van Hal et al. (2012).

¹⁰⁰ Stewart, D. W., Shamdasani, P. N. and D. Rook (2007). *Focus Groups: Theory and Practice, 2nd Edition*. Thousand Oaks, CA: Sage.

¹⁰¹ *Focus Groups: Theory and Practice, 2nd Edition*. page 9.

¹⁰² See page 2, Van Hal et al. (2012).

with respect to colour (e.g., white signifies ‘death’ in Chinese culture but ‘cleanliness’ in Western cultures). However, the stimulus in this study is simply a plain pack and a branded pack. As a result, the study ends up not being structured to actually answer the questions it sets out to answer pursuant to the purpose of the investigation.

- (b) The authors further justify the use of qualitative methods in this study on the basis that “*in Flanders, no study on plain packaging had been done before. For that reason, a qualitative study on plain packaging in Flemish adolescents was performed to rule out erratic results compared to the existing evidence*”.¹⁰³ However, the fact that no study has been done in a region of Europe is no justification in itself for qualitative research (such a logic would imply that every region was so unique as to justify a never ending stream of studies). Studies need to be justified by the value they bring to theoretical and practical problems, not by the fact that they simply have not been done by others. More problematic is the assumed logic that a sample of 55 individuals would be sufficient to “*rule out erratic results*”, as focus groups have no scientific ability to achieve this. If this was the authors’ aim, they should not have created a de novo question/interview protocol. Rather, they should have replicated any number of the previously published qualitative studies and performed a direct comparison (which, as an aside, would have been far better justified using quantitative approaches). Hence, this point and the prior discussion (in paragraph 4.120(a)) leads one to question what it is that this study achieves given the form it is being done.
- (c) However, the major issue with this study is that while the authors argue that it is ‘exploratory’, they make definitive conclusions that do

¹⁰³ *Ibid.*

not relate back to the supposed ‘cultural and local’ factors that were part of the justification for the study in the first place:

- (i) First, despite one of the justifications being that no study on plain packaging had been done in Flanders, there is not a single statement in the Discussion that relates to the Flemish context of the study or the significance of the cultural and local context.
- (ii) Second, the authors conclude, “*Our study shows that the most important motives for young people to choose a certain brand of cigarettes are the price of the cigarettes, the taste of the cigarettes, the influence of the social environment and the recognition of the packaging.*”¹⁰⁴ However, the relative importance of each of these factors is not explained nor is how they are incorporated into the individual’s decision calculus. Further, it is hardly surprising that price, brand and taste, as well as the social environment, matter to a consumer’s decision. But what about the pack? Why does it appear to be emphasised by the subjects? This is a second serious misinterpretation of the results.
- (iii) Given that the stimuli for the focus group discussion were the packs, it is possible that the mentioning of the pack and the health warnings are influenced by the study itself. This is then reinforced by the fact that in the question protocol ‘brand’ is mentioned 10 times, the ‘pack’ 15 times, and the ‘health warning’ three times, while price and motive/environment are each mentioned less than three times. Overall, given that the questions asked primed the individual to focus on the pack, the health warnings and the brand, it is not surprising that the

¹⁰⁴ See page 3, Van Hal et al. (2012).

results seem to confirm that this is what the participants mentioned in the focus group discussion.

- (iv) Finally, at the end, the authors conclude: “*We therefore believe that introducing plain packaging in Belgium can result in a substantial health gain in young people.*”¹⁰⁵ Not only is this statement not justified by a non-representative sample of 55 individuals, it is also rather misplaced as the conclusion is drawn about Belgium when the sample is Flemish, and no Walloon adolescents were part of the study.

4.121 **Uppal et al. (2013)**. This study is a focus group study with 22 smokers categorised by their motivation to quit smoking (10 “high motivation” and the remainder “low motivation”). The intent of the study is to examine decision making in relation to the use of NHS Stop Smoking Services (SSS) programs using the PRIME framework as a structure for the investigation. As part of the study, the participants discussed how various factors affected motivation to quit smoking and decisions to continue smoking, including attitudes towards existing and impending tobacco control policies (such as plain packaging).

- (a) This study is generally one of the better focus group studies evaluated (when just looking at the structure of the study). Unlike all of the other studies, it begins with a basic theoretical model (PRIME) and works out the protocol based upon that model. In essence, PRIME is a ‘process’ model – i.e., it focuses on the stages and individual goes through when making a decision to quit smoking – and process models are generally more amenable to investigation via qualitative approaches. Overall, the authors work diligently to keep the discussion well within the confines of the study and the sample, and do not go on to make policy conclusions from the limited sample.

¹⁰⁵ See page 8, Van Hal et al. (2012).

- (b) While we cannot draw any definitive conclusions from the study due not only to the small sample size but also to the limitations noted by the authors,¹⁰⁶ it is a potentially useful study in that it highlights two important facts that can guide future work. First, it seems to hint that survey based measures (in this case the motivation to quit instrument) may not be capturing motivation to quit intensity and outcome – *“In this study, attitudes toward smoking and quitting were similar for both low motivation and high motivation smokers, despite participants demonstrating explicit differences in motivational scores to quit smoking during the screening process. This suggests that more sensitive measures of current levels of motivation need to be developed in further research.”*¹⁰⁷ Second, it suggests the PRIME model, which is the standard by which quit smoking programs operate, may need to be re-evaluated in light of some of the discussions of the study.

4.122 **Edwards et al. (2012).** This study is a focus group study comprised of 24 New Zealanders (smokers and non-smokers) that examined public support for specific tobacco control policy alternatives.

- (a) As noted earlier in this report, the core of the idea of a focus group is ‘focus’ and the ability to tap into an individual’s experience of “some specific concrete experience” to understand a phenomenon that is as yet not well enough structured for stronger theoretical and empirical study. In this case, the focus group does not tap into any real experience but simply involves discussions of policy alternatives where the individuals have no experience or knowledge on which to base their judgments beyond that which they can rationalise as their opinion.

¹⁰⁶ See page 8, Uppal et al. (2013).

¹⁰⁷ See page 8, Uppal et al. (2013).

- (b) Ultimately, the study does not articulate what the costs of the various proposals are and does not generate any specific conclusions that are operational.

‘Public Support’ Studies

4.123 In this section, I review two of the studies aimed at gauging public support for tobacco control policies. While these studies are different from the studies I have examined earlier in this report, in that they do not attempt to measure any “effects” of specific aspects of potential policies, I comment on the method and approaches of these studies. These studies are (a) Rosenberg et al. (2012) and (b) Hoek et al. (2013). I would note that Edwards et al. (2013) is also a ‘public support’ study but I chose to evaluate that study with the other focus groups studies as its methodological limitations were such that it was easier to outline these by including it with that group of studies. Overall, one major issue with both studies is what their ultimate goal is. It is very well recognised that the relationship between public policy and public opinion is complex and not easily stylised. The public can be supportive of policies today that they dislike tomorrow (once they discover the true costs) and can dislike policies today that tomorrow they support (e.g., such as gay marriage).¹⁰⁸

4.124 **Rosenberg et al. (2012).** This study is a telephone interview survey of 2,005 adults in Western Australia in 2010. The intent of the survey was to examine reactions to different forms of tobacco control measures, including plain packaging. The dominant part of the study was related to restrictions on smoking at specific venues in Western Australia. For brevity, I will focus only the responses to the questions regarding plain packaging (Table 2 of the study).

¹⁰⁸ See, e.g., Soroka, S. and C. Wlezien (2010). *Degrees of Democracy: Politics, Public Opinion, and Policy*. Cambridge, UK: Cambridge University Press.

4.125 Ultimately, there is little that can be drawn from the results of this study, as it is little more than a simple opinion survey. Three issues are relevant for addressing the relevance of what was done.

- (a) First, as the survey was done in 2010, and the Australian Government's plain packaging policy had only recently announced, it is not clear whether the opinions being given: (a) reflect informed opinions and (b) to what extent they represent pseudo opinions generated by the fact that the whole opinion survey was about tobacco control. As noted before, it is well understood that having a "Don't Know" option in policy surveys is important.¹⁰⁹ In the case of this survey they use the expression "*No Feelings Either Way*", which is different from saying that the individual is simply uninformed about the policy. In either case, we simply do not know whether or not the individual had sufficient knowledge when expressing their opinions. If they do not then the spectre of pseudo opinions rises. The study did not attempt to control for this by adding in fictitious policy initiatives so as to tell whether or not the individuals were simply giving socially acceptable answers.
- (b) Like most public opinion surveys, the instrument is potentially subject to social desirability bias. In other words, as there is no cost for the individual to express what they consider to be the "acceptable" politically correct answer, there will be a tendency to do so. This is one possible explanation for why women are most 'positive' toward the regulatory interventions, as, again as noted earlier in this Report, women are generally more subject to social desirability bias.¹¹⁰ My point here is simply to note that the simple breakdowns and "one at a

¹⁰⁹ Luskin, R. C. and J. G. Bullock (2011). 'Don't Know' Means "Don't Know": DK Responses and the Public's Level of Political Knowledge, *Journal of Politics*, 73 (2): 547–557.

¹¹⁰ See, e.g., Rapoport, R. B. (1982). Sex Differences in Attitude Expression: A Generational Explanation, *Public Opinion Quarterly*, 46 (Spring): 86–96.

time” comparisons used in the study do not account for alternative explanations of the results, particularly those related to well known biases in survey and opinion polls.

- (c) This last point is related to the fact that opinion surveys like this one are not incentive compatible. The individual is, therefore, not making realistic decisions related to one policy alternative versus another or given consideration as to what the cost of the policy might be to them and the public.¹¹¹ This problem is identical to what we have seen in nearly every study on plain packaging evaluated in this report and my 2010 and 2012 Reports. Given that there is no cost to the opinions being expressed, nor are they taken in realistic settings where the costs of the behaviour is clear, it is impossible to determine how they translate into actual behaviour or to actions that would be undertaken when the costs and benefits are realized.

4.126 **Hoek et al. (2012).** This study is a online survey of 836 adults smokers and non smokers in New Zealand in 2012. The stated intent of the survey was to examine individual agreement with specific public health arguments plus support for plain packaging – *“Harmful products like cigarettes and tobacco should have very large warnings on them”*, *“Attractive tobacco packaging encourages young people to experiment with smoking”*, *“Tobacco packaging encourages smokers to try new brands”*, *“Plain packaging would discourage young people from trying smoking”*, *“Plain packaging would encourage smokers to try and quit”*, *“Plain packaging is unfair because it would stop tobacco companies using their brands and logos to promote their products to different forms of tobacco control measures, including plain packaging”*.

¹¹¹ Alvarez, R. M. and J. Brehm (2002). *Hard Choices, Easy Answers: Values, Information, and American Public Opinion*, Princeton, NJ: Princeton University Press.

4.127 This study is subject to all of the issues outlined in my discussion of Rosenberg et al. (2012). It is simply noted that problems relating to (a) the knowledge of the respondents, (b) the tendency of the questions to generate pseudo opinions, (c) the clear intent of the study and hence the likelihood of social desirability bias, and (d) and the lack of incentive compatibility are all issues that make the interpretation of the results difficult.

4.128 However, the most concerning problem with this survey is that it confuses individuals' "opinions" – e.g., "*Harmful products like cigarettes and tobacco should have very large warnings on them*" and *Plain packaging is unfair because it would stop tobacco companies using their brands and logos to promote their products to different forms of tobacco control measures, including plain packaging*" – with facts – e.g., "*Attractive tobacco packaging encourages young people to experiment with smoking*", "*Tobacco packaging encourages smokers to try new brands*" – and scientific predictions – "*Plain packaging would discourage young people from trying smoking*", "*Plain packaging would encourage smokers to try and quit*". While it might be appropriate to ask people their opinion about what they think public policy **should** be (in their role as citizens they have this right) or what is **fair or unfair** (as their role as members of a society they have this right), asking them what they believe are provable facts and causal relationships is inappropriate. It is the role of scientific studies to determine facts and make predictions. Randomly surveyed individuals in a society will be informed enough to know what the impact of policies are and are not going to be able to understand the complexities that lead to specific behavioural outcomes (such as encouragement to consume a specific product in specific contexts). As noted earlier, when individuals lacking knowledge are asked questions that appear to be related in a context in which the intent of the research is clear the respondents are likely to create heuristics to generate meaning both for the specific questions and the questions grouped together (see the discussion in paragraph 4.99 (a) and the citations given).

What is seen in the data will reflect those heuristics and hence confuse realistic results with results generated by the study's design and how specific items are measured.

Conclusion

4.129 Overall, neither of these studies adds anything of scientific value to the debate over the impact of plain packaging regulation. Both suffer from traditional serious limitations associated with public opinion polls. In addition, Hoek et al. (2012) confuses opinion with facts and predictions, making an interpretation of their results difficult and problematic.

5. CONCLUSION

5.1 In Section 4 of this report, I have provided a number of detailed comments about specific aspects of each of the Studies that relate to plain packaging. In this section of the report, I will summarise and expand upon that analysis by looking specifically at the criteria I discussed in paragraphs 2.4, 2.10, 2.11 and 2.12. Unlike Section 4, which focused on each study individually, in this section I will focus on the Studies as a group and relate them back to the specific criteria.

5.2 Consistent with my 2010 and 2012 Reports, I will first address the validity of the studies using the same reporting format. I will then summarise my findings using an overall format. For simplicity of exposition, I am excluding the two 'Public Support' Studies, Rosenberg et al (2012) and Hoek et al. (2012), and Pechey et al. (2013) from this discussion, as they are quite different from the standard consumer studies. I will come back to them separately.¹¹²

¹¹² I exclude Dewe et al. (2013) from this discussion as it is irrelevant to the discussion of the potential impact of plain packaging and its specific flaws are articulated in my more specific analysis.

5.3 The criteria outlined in paragraph 2.4 concentrate on the question of the extent to which a study provides valid and generalisable conclusions that align with what an individual would do across a range of contexts, such as might exist in purchasing situations. The importance of salience and incentive compatibility to this is that the Studies should be generating results that represent realistic behaviour (incentive compatibility) without bias induced by making aspects of the product or the situation more salient than it would in reality.

5.4 It is my conclusion that all of the Studies examined fail to meet reasonable **incentive compatibility** requirements. There is no indication that the studied individuals' attitudes and intentions, as measured, align with their actual or future behaviours. Nor does any study posit an *a priori* testable decision model that is underlying their investigation. Because the Studies focus on packaging absent other salient attributes of the products (such as price) and other factors, we do not know the degree to which the Studies provide realistic information about the degree to which packaging matters.

5.5 It is my conclusion that the Studies also fail to provide **experimental or situational contexts that created realistic scenarios** in which the individual would be applying the decision model that the individual used when making purchasing decisions or decisions that related to the policy goals outlined in paragraph 1.5. The fact that all of the Studies focused exclusively on package design, created a situation where packaging features dominated the evaluation because no other attributes were presented to counter this salience (which would have been the case if, for example, various prices were applied to the different packages). The artificiality of the studies and the fact that they invariably compared plain versus current packaging (an unrealistic scenario) most likely increased the salience of the packaging to a degree that it impacted artificially on the attitudes and perceptions seen.

5.6 Finally, it is my conclusion that, as all but one of the Studies (Ramunno et al. (2012)) **concentrated entirely on stated preferences, attitudinal measures and focus group opinions**, one cannot assume any predictive accuracy with respect to actual purchasing behaviour or the intended policy goals outlined in paragraph 1.5. For example, the context in which all the experimental studies were done did not provide a realistic set of circumstances in which the individual was considering plain packaged alternatives at different prices or alternatives that they could actually purchase. Overall, none of the Studies were able to provide a “line of sight” between their attitudinal and perceptual measures and actual purchasing behaviour, nor any of the measures outlined in paragraph 1.5 above, namely reducing smoking initiation among minors, reducing smoking consumption among minors and/or adults, or increasing smoking cessation among minors and/or adults. One would take an unjustified “leap of faith” about the importance of the various measures used in the Studies to relate them to actual behaviours.

5.7 The criteria outlined in paragraph 2.10 relate specifically to the experimental studies examined. In the case of Moodie & Mackintosh (2013), Hammond et al. (2013), Hammond, White et al. (2013), Borland & Savvas (2013), Ford et al. (2013), Ramunno et al. (2012), Lund & Scheffels (2013) and Scheffels & Lund (2013), it is my conclusion that these studies fail to meet even the most basic standards of good experimental design, efficiency, orthogonality and balance, implying that one can make no conclusions at all about the relevance of their findings with respect to the importance of the components of cigarette packaging that they study. Indeed, the criticisms made in my 2010 and 2012 Reports with respect to earlier studies by many of these authors remain valid as the Studies reviewed here are little more than repetitions of these prior studies using the same measures and biased methodological approaches. Even when the researchers do not appear to have previously published studies in this area (e.g., Ramunno, et al. and Lund

and Scheffels) there is a tendency to repeat the same methodological mistakes from earlier studies.

5.8 Ultimately, the goal of experiments is to determine the decision model underlying choice, hence it is important not just to have a well-designed experimental structure but also to include in the experimental conditions all the factors relevant to the determination of the different decisions individuals would reasonably be expected to make. Unfortunately, none of the Studies has either an effective decision model or a structure to estimate that decision model which would allow one to make a valid estimation of the impact of plain packaging. All of these studies **assume**, rather than **prove** or even **justify**, that what they are measuring is related to the actual behaviours that they discuss in their conclusions.

5.9 The criteria outlined in paragraph 2.11 help us understand the degree to which stated intentions align better with actual behaviour and it is worth noting that criteria are relevant to the experimental studies discussed in paragraphs 5.7 and 5.8 above. As noted above at paragraph 2.11, stated intentions align better to actual behaviour when: (a) they are for existing products, (b) they are for durable rather than non-durable goods,¹¹³ (c) they are for short-term horizon decisions rather for long-term time horizon decisions, (c) they relate to purchase intentions for specific brands rather than for the product category in general, (d) they are related to the here and now rather than the future, (e) individuals understand the purchasing context, (f) intention is related to trial rather than usage, and (g) trial is measured comparatively (e.g., one brand versus another). Overall, it is my conclusion

¹¹³ A non-durable good is a good that is consumed in its use. A durable good is one that is not consumed by its use. The distinction between durable and non-durable goods is generally considered to be a continuum. For example, at one extreme are pure non-durables and pure durables. A candy bar is a pure non-durable good because it is gone when we eat it. Whereas a washing machine would be a pure durable as it is expected to last years (or decades) and one use does not reduce the efficiency of future uses. A non-rechargeable battery would be considered to be a non-durable even though it did last for a limited amount of time it is ultimately consumed by its use. A rechargeable battery would be considered a durable as it can be renewed virtually indefinitely.

that, as with the studies evaluated in my 2010 and 2012 Reports, none of the Studies met enough of these criteria whereby one could argue that they would provide potentially valid predictions of actual purchasing behaviour. I discuss each criteria below:¹¹⁴

- (a) *Existing products.* For the most part the experimental studies use an existing product category for which there is some understanding of the product on the part of the consumers, particularly smokers. A number of the studies, however, use samples of that are either dominantly, or exclusively, non-smokers – Hammond et al. (2013), Hammond, White et al. (2013), Ford, et al. (2013), Lund and Scheffels (2013), Scheffels and Lund (2013). This creates a potential problem in that the products are not “existing products” for the subjects, as they do not use, or have never used, them. Hence, the logic by which they evaluate the products cannot be determined to be realistic as they have no active experience with the category.
- (b) *Durable vs. non-durable goods.* Cigarettes are a non-durable good, subject to frequent purchasing. Hence, this might imply that stated intentions are less accurate for cigarette consumption that is based a more complex heuristic that cannot be picked up in simply static surveys or experiments.
- (c) *Short-term horizon vs. long-term time horizon decisions.* The issue is that while the purchase of a specific packet of cigarettes is a short-term horizon decision, the policy questions given in paragraph 1.5 – smoking uptake, smoking reduction and smoking cessation – are long-term horizon decisions. Hence, from a policy perspective the stated intentions are unlikely to be accurate predictors of longer-term behaviour, independent of whether it is uptake, reduction or cessation.

¹¹⁴ More detail is given in paragraph 5.7 of my 2012 Report.

- (d) *Specific brands vs. the product category in general.* The results of the various studies are mixed on this, as the choices asked do not focus on the choice of specific brands but on specific packages. This is important in this situation, as cigarettes are a product for which there is often significant brand loyalty, but none of the studies related the choices made to the brands that individuals smoked. Hence, asking individuals to choose amongst packages (when their real decision model is to choose amongst brands) is likely to exacerbate the salience of the package relative to the brand, even when the brand is the most relevant factor.
- (e) *Purchase intentions measured as ‘trial’ rates.* Individuals are better at indicating if they might ‘try’ a product, which can be influenced by non-use factors. However, whether they continue to purchase the product will be heavily influenced by their views of the product. None of the studies asked about ‘trying’ the product as opposed to ‘purchasing’ it. By this I mean, the question should be asked “would you try the product?” rather than “would you purchase the product?”. In the case of the studies examined here, trial is only weakly measured, e.g. Hammond et al. (2013) and Hammond, White et al. (2013), where there is what the authors call a “behavioural task”. As noted, that task is not a behavioural task nor does it relate to a brand decision task (as it is simply about different packs within the same brands).
- (f) *Comparative purchase intentions.* Purchasing occurs via product comparison and what a product was compared to provides more information as to why one product was chosen. Most of the Studies had subjects evaluate packets one at a time except for Hammond, White et al. (2013), who had subjects compare each pack to a base B&H Pack and Ford et al. (2013), where subjects saw five packs at once. However, as noted, the comparative cases even here are not

realistic choices that relate to what individuals would be doing in the marketplace.

5.10 The criteria outlined in Table 2 above represent the final mixture of items to consider when evaluating the Studies. As noted throughout this Report, all of the studies fail on a number of these dimensions. For simplicity, I will discuss this based on the general categories given in Table 2: Field Administration Protocol, Appropriateness of the Sample Frame, and Analysis.

5.11 **Field Administration Protocol** asks whether the questions were appropriately constructed and relevant, there was no researcher bias, the researcher was objective, and the responses received were relevant. In the case of all of the Studies, there were significant limitations in the structuring of questions. For example, in many cases, questions were asked which assumed knowledge or led the respondent. It is my opinion that Wakefield, et al. (2013), Hammond et al. (2013), Hammond, White et al. (2013), Borland and Savvas (2013), Ford et al. (2013), Lund and Scheffels (2013), and Scheffels and Lund (2013) all suffer from significant methodological flaws in terms of poor question design and, most dominantly, inappropriate aggregation of data that inflate and hide important sources of variability. In addition, all of the focus group studies (Ford, Moodie et al. (2013), Van Hal et al. (2012), Uppal et al. (2013) and Edwards et al. (2012)) plus Ramunno et al. (2012), suffer from bias that is induced by the inability of the researcher to hide the intent of the study – leading to potential respondent bias.

5.12 **The sampling frame** in the Studies varied. Although some of the samples are broad, none are nationally representative, implying that none of the Studies can be used to make predictions about the general population. Some, such as Hammond et al (2013), Hammond, White et al (2013), Ford, et al. (2013), Lund and Scheffels (2013) and Scheffels and Lund (2013), concentrate on specific samples that are generally going to be representative

of those narrow groups. None of the focus group studies can be considered to have samples that can be generalised in any way and the eye tracking study of Ramunno et al. (2012) has an even smaller sample than the similar Munafo et al. (2011) study that I discussed in my 2012 Report.

5.13 **The analysis** in the Studies suffers from significant statistical validity issues. Most particularly, the experimental design limitations outlined in paragraphs 5.4 to 5.8 above, and discussed throughout Section 4, imply that the statistical analyses seen must be considered to be invalid. Their failure to structure their experiments correctly implies that their ability to make generalisable conclusions is lost. These issues are exacerbated by the use of invalid econometric structures and a tendency to conduct their analyses one measure at a time rather than building holistic and comprehensive decision models that account for the totality of the effects they are examining (a fact that is all the more important given the experimental design limitations). As noted throughout Section 4, a range of other statistical and methodological issues arise with respect to the Studies that collectively imply that their validity with respect to the policy questions outlined in paragraph 1.5 above is questionable.

5.14 Of the Studies I have evaluated in this Report, Pechey et al. (2013) is the most unusual and requires discussion in its own right. In my opinion, this is a very seriously flawed study, most particularly in that the sample of individuals being used to elicit “expert opinions” is small and biased (in that it is restricted to “tobacco control experts”). The use of the Systematic Review further reinforces the bias by presenting the body of research in a manner that overstates the validity of the research to date (e.g., the experts were not faced with an independent evaluation of that same research)..

5.15 Finally, the two ‘Public Support’ studies stand alone when examined in this report. Rosenberg et al. (2012) and Hoek et al. (2012) add nothing to the scientific evidence on the validity of the scientific evidence as they relate

to the public policy goals outlined in paragraph 1.5 and are rather naïve in failing to account for the many issues relating to the validity of opinion polling, in general, and, more importantly, polling when addressing emotive political issues more specifically.

5.16 To summarise, Table 3 below provides an overview of each of the studies examined in Section 4 and provides both a summary of their limitations on the criteria just discussed. It highlights a number of key points, in particular the lack of direct behavioural measures, methodological limitations and the lack of representativeness of the sampling frames used.

Table 3 - Overview of the Studies Evaluated in This Report

Study	Method	Sample	Issues Noted in Section 4	Behavioural Measures
Pechey et al. (2013)	Expert Elicitation	N=33, Tobacco Control Professionals	Failure of 'Expert' Judgment, Inadequate Justification of the Approach, Utilization of a Systematic Review Rather than a Planned Meta Analysis, Bias Sample – Experts Used were Many of Those Publishing Studies Examined, Sample Size Small, No Underlying Model Driving the Results	None
Wakefield et al. (2013)	Survey	N=,536 Smokers, Non-Representative, Australian Sample (Victoria Only)	Method Bias (Intent of Study Clear to Subjects) Poor Question Design, Recall Bias, Social Response Bias “One-at-a-Time” Statistical Analysis	None
Moodie & Mackintosh (2013)	Survey & Experimental Manipulation	N=301, Smokers, Non-Representative, Female Only, UK Sample (Scotland Only)	Poor Experimental Design, Lack of Balance, No Controls, Significant Sample Attrition	None (Usage was Self Reported)
Hammond et al. (2013)	Survey & Experimental Manipulation	N=947, 16–19 Year Olds, Female Only, Non-Representative, UK Sample	Poor Experimental Design, Lack of Orthogonality, Lack of Balance, Inappropriate Data Manipulation, Econometric Analysis Incorrect, “One-at-a-Time” Statistical Analysis	None

Study	Method	Sample	Issues Noted in Section 4	Behavioural Measures
Hammond, White et al. (2013)	Survey & Experimental Manipulation	N=762, 11–17 Year Olds	Poor Experimental Design Base Case Bias Lack of Orthogonality, Lack of Balance, Inappropriate Data Manipulation, Econometric Analysis Incorrect, “One-at-a-Time” Statistical Analysis	None
Borland & Savvas (2013)	Survey & Experimental Manipulation	N=160, 18–17 Year Old ‘Ever Smokers’, Australian Sample (Victoria Only), Non- Representative	Poor Experimental Design, Lack of Orthogonality, Lack of Balance, Order Bias Effects, Inappropriate Data Manipulation, Econometric Analysis Incorrect, “One-at-a-Time” Statistical Analysis	Yes
Ramunno et al. (2012)	Eye Tracking	N=28, Smokers/Non- Smokers, Non- Representative	Small Sample, Lack of Accounting for Individual Heterogeneity	None
Ford et al. (2013)	Survey & Experimental Manipulation	N=1,025, 11–16 Year Old ‘Never Smokers’, UK Sample, Non- Representative	Poor Experimental Design, Lack of Orthogonality, Lack of Balance, Bias Due to Presentation Effects, Poorly Constructed Measures, Inappropriate Data Manipulation, Erroneous and Flawed Econometric Modelling	None
Dewe et al. (2013)	Content Analysis of Advertisements	N=204, Advertisements from 1950 through 2003	Unjustified Aggregation of Advertisements, Lack of Measures of Variables of Interest, Unjustified Conclusions Based on Domain of the Sample of the Study	None that relate to the variables examined

Study	Method	Sample	Issues Noted in Section 4	Behavioural Measures
Lund & Scheffels (2013)	Survey & Experimental Manipulation	N=1,022, 15–22 Year Old Smokers and Non-Smokers, Norway Sample, Non-Representative	Poor Experimental Design, Lack of Orthogonality, Lack of Balance, Intent of the Study Clear to Participants, Poorly Constructed Measures, Inappropriate Data Manipulation, Simplistic Empirical Modelling, Lack of Control for Gender Bias Effects	None
Scheffels & Lund (2013)	Survey & Experimental Manipulation	N=1,010, 15–22 Year Old Smokers and Non-Smokers, Norway Sample, Non-Representative	Link to Lund & Scheffels Unclear, Poor Experimental Design, Lack of Orthogonality, Lack of Balance, Intent of the Study Clear to Participants, Poorly Constructed Measures, Inappropriate Data Manipulation, Simplistic Empirical Modelling, Lack of Control for Gender Bias Effects	None
Ford, Moodie et al. (2013)	Focus Group	N=48, 15 Years Old, UK Sample (Location Unknown), Non-Representative	Focus Groups Do Not Allow for the Drawing of Valid Scientific Conclusions	None
Van Hal et al. (2012)	Focus Group	N=55, 15-18 Years Old, Flanders Sample, Non-Representative	Focus Groups Do Not Allow for the Drawing of Valid Scientific Conclusions	None
Uppal et al. (2013)	Focus Group	N=22, Smokers, UK Sample (Nottingham), Non-Representative	Focus Groups Do Not Allow for the Drawing of Valid Scientific Conclusions	None

Study	Method	Sample	Issues Noted in Section 4	Behavioural Measures
Edwards et al. (2012)	Focus Group	N=24, Smokers and Non-Smokers, New Zealand Sample (Whanganui), Non-Representative	Focus Groups Do Not Allow for the Drawing of Valid Scientific Conclusions	None
Rosenberg et al. (2012)	Opinion Poll	N=2,005, Adult Smokers and Non-Smokers in Western Australia	Generation of Pseudo Opinions, Lack of Incentive Compatibility, Intent of the Study Clear to Respondents Simplistic Empirical Modelling,	None
Hoek et al. (2012)	Opinion Poll	N=836, Adult Smokers and Non-Smokers, New Zealand Sample	Generation of Pseudo Opinions, Lack of Incentive Compatibility, Intent of the Study Clear to Respondents Simplistic Empirical Modelling,	None

5.17 In conclusion, it is my expert opinion based on the publicly available consumer surveys and experiments that I have evaluated in this report that they do not provide reliable evidence that plain packaging would be effective in achieving the public policy goals of changing actual smoking behaviour, namely in:

- (a) reducing smoking uptake (also known as initiation) among minors;
- (b) reducing smoking consumption among minors and/or adults; or
- (c) increasing smoking cessation among minors and/or adults.

5.18 I confirm that insofar as the facts stated in my report are within my own knowledge I have made clear which they are and I believe them to be true, and that the opinions I have expressed represent my true and complete professional opinion.

Signature

A handwritten signature in black ink, appearing to read 'Timothy M. Devinney', with a stylized flourish extending to the right.

Name: Professor Timothy M. Devinney

Date: 3 January 2014

6. EXHIBIT ONE – QUALIFICATIONS OF PROFESSOR DEVINNEY

Ex.1.1 I am a University Leadership Chair at the University of Leeds in the United Kingdom. In addition, I am a Conjoint Professor in the Faculty of Medicine at the University of New South Wales, Australia and a Visiting Professor at the Institute of Management at Humboldt, Berlin. My educational background includes a B.Sc. (1977) in Psychology from Carnegie Mellon University, and M.A. (Public Policy, 1979), MBA (Economics and Statistics, 1981) and PhD (Economics, 1984) from the University of Chicago.

Ex.1.2 I have held academic positions at the University of Chicago (Lecturer), Vanderbilt University (Asst Professor), University of California – Los Angeles (Asst Professor), The Australian Graduate School of Management (AGSM) (Associate, Chaired Professor, Professorial Research Fellow), and the University of Technology – Sydney (Professor).

Ex.1.3 I have held visiting Professorships at the Universities of Trier, Frankfurt, Ulm, Hamburg and Humboldt University in Germany, London Business School in the UK, Copenhagen Business School in Denmark, Hong Kong University of Science and Technology and City University in Hong Kong.

Ex.1.4 I have taught MBA and doctoral courses at University level for over 25 years. I teach in the marketing, international business, strategic management, innovation and statistics/research methods areas. I was the Founding Director of the Executive MBA Program at the AGSM and have taught extensively on executive development programs around the world.

Ex.1.5 I am one of the leading researchers in the social sciences in Australia, having secured extensive research funding through the Australian Research Council and other external funding bodies.

Ex.1.6 I have published in the leading business journals in the field of social sciences including the *Journal of Marketing*, *Journal of International Business Studies*, *Management Science*, *Organization Science*, *Strategic Management Journal*, and many others. I am on the editorial board of 10 of the leading academic journals and serve as Co-Editor of the *Academy of Management Perspectives* (the #14 ISI listed journal in Business), the *Advances in International Management* Series published by Emerald, Editor of *Foundations and Trends in International Business and Management*. I am also the author or editor of more than 10 books.

Ex.1.7 I am a Fellow of the Academy of International Business, an International Fellow of the Advanced Institute of Management (UK), a Fellow (Distinguished Member) of the Australia New Zealand Academy of Management, a Research Awardee of the Alexander von Humboldt Foundation (Germany), and a Bellagio Residence Fellow of the Rockefeller Foundation. My work has been recognised by numerous organisations including being awarded the Researcher of the Year award by the Australia New Zealand Academy of Marketing.

7. EXHIBIT TWO – RESUME

Education

B.Sc. (Psychology), with highest honors, Carnegie-Mellon University, 1977

M.A. (Public Policy Studies), University of Chicago, 1979

M.B.A. (Economics and Statistics), University of Chicago, 1981

Ph.D. (Business Economics), University of Chicago, 1984

Academic Experience (excluding visiting positions):

University Leadership Chair and Professor, Leeds University Business School, University of Leeds, September 2013–Present

University Professor of Strategy, School of Business, University of Technology Sydney, July 2009–August 2013.

Professor (Conjoint), Faculty of Medicine, University of New South Wales, July 2009–present.

Professor of Management, Australian Graduate School of Management (now Australian School of Business), University of New South Wales, 1993–2009 (June). Director Centre for Corporate Change, 1999–2006. AGSM Professorial Research Fellow, 2006–2009.

Assistant Professor of Management, Anderson Graduate School of Management, University of California, Los Angeles, 1990–1992.

Assistant Professor of Management, Owen Graduate School of Management, Vanderbilt University, 1982–1990.

Lecturer in Mathematics, Graduate School of Business, University of Chicago, 1981–1982.

Academic Experience (examples of courses taught):

AT LUBS: International Business Strategy (MBA/EMBA).

AT UTS: Philosophy of Science (PhD).

At AGSM (Recent MBA/EMBA): International Business in Asia (MBA on site project course in China), Corporate Strategy (MBA/EMBA), International Business Strategy (MBA/EMBA), Strategic Management of Intellectual Property (MBA shortcourse), Philosophy of Social Science (PhD), Ph.D. Seminars (one on Corporate Strategy and one on Innovation), Globalization of the Knowledge Based Organization (MBA).

At AGSM (Executive Programs): Managing Competitive Strategy,* Technology Management,* The International Manager's Program,* Managing Intellectual Property,* the Accelerated Development Program, the General Manager Program, and the Development Program for Managers (* indicates program directorship and development).

At Vanderbilt: Pricing, Managerial Decision-making and EC92, Economics of the Firm, International Business, Industrial Strategy and Organization, Regulation and Antitrust Economics, Financial Institutions, Information Economics (Ph.D. Level)

At UCLA: Marketing Strategy, Product Management, MBA Projects.

Professional Associations/University Affiliations (including awards/recognitions):

Professional Associations and Activities Therein (Including Awards): American Economic Association, Econometric Society, INFORMS, The Product Development Management Association, Academy of International Business, Academy of Management, Australia New

Zealand Academy of Management, Australian New Zealand Marketing Academy

INFORMS: Organizing Committee, Marketing Science Conference, Nashville 1987; Program Coordinator, Euro XII/TIMS International XXXI, Helsinki 1992 and TIMS XXXIII, Singapore 1995; Organizing Committee, Marketing Science Conference, Sydney 1995

Academy of International Business (AIB): Chair, Annual Meeting, Sydney, 2001. Journal of International Business Studies (Editorial Board & Consulting Editor). 2006 JIBS Decade Award Committee Chair. Faculty 2006 AIB Doctoral Consortium. 2007 Program Organizing Committee (Track Chair). 2009 Program Organizing Committee (Track Chair). 2010 Program Organizing Committee (Track Chair). AIB Fox Best Paper Award Committee, 2009–2012. 2013 Program Organizing Committee (Track Chair). 2014 Program Organizing Committee (Track Chair). Best Paper Award Finalist, 2005 AIB Conference. Best Reviewer, 2012 AIB Conference. AIB Fellow, Elected 2008

ANZAM (ANZ Academy of Management): Executive Committee, ANZAM (2005–2006), Best Paper Award (2001, 2006 & 2007), 2007 ANZAM Conference Organizing Committee, Distinguished Member (Fellow), 2008

ANZMAC (ANZ Marketing Academy): Researcher of the Year Award, 2007

Academy of Management (AOM): Executive Committee and Chair, International Management Division of the Academy of Management (2004–2009). Raised funding for the Booz & Co./strategy+business eminent scholar award in International

Business (2004–2009). Organized IMD PDW program (2005). Organized IMD Conference Program (2006). Served as Division Chair (2007–2008). Received one of three AOM Enterprise Awards (2005) to develop multimedia delivery of IMD Program activities. Chair IMD Communications Committee (2008–2009). IMD Service Committee (2010–2012). BPS Junior Faculty Consortium Faculty (2006). New Doctoral Consortium Faculty (2008). *Carolyn Dexter Award Finalist (Social Issues in Mgt Division)—Best Intl Paper (2007)*. *Best Paper Finalist (Social Issues in Mgt Division), 2011, Social Issues in Management (SIM) Book Award (2012)*. Evidence-Based Management Collaborative (2007–2008), Founding Member. Academy of Management Perspectives, Co-Editor (2011-2015), Associate Editor (2006-2010)

Strategic Management Society (SMS): *Strategic Management Journal (editorial board)*, *Global Strategy Journal (editorial board)*, *International Management Interest Group: Representative-at-large (2008–2009)*, *Associate Program Chair (2011)*, *Program Chair (2012)*, *IG Chair (2013)*. *Conference Organizing Committee, Rio de Janiero (2010)*. *Conference Organizing Committee and Consortium Chair, Sydney (2014)*.

Editorial and Refereeing Duties (Formal):

Director, Social Science Research Network (SSRN), International Business Research Network and Editor, *International Business Strategy & Structure*, 2009–present, *Sustainability Research & Policy Network*, 2012–present

Co-Editor, *Advances in International Management*, Emerald (with T. Pederson and L. Tihanyi), 2009–present. Received the “Leading Editor” Award in the 2011 in the Literati Awards.

Founding Editor, *Foundations & Trends in International Business and Management*, 2014–present

Co-Editor, *Academy of Management Perspectives*, (with D. Siegel) 2011–present, Associate Editor, 2006–2011

Associate Editor, *Australian Journal of Management*, 1995–2005

Associate Editor, *Management Science*, 1988–1990

Consulting Editor, *Journal of International Business Studies*, 2011–present. Co-Editor of special issue on Measuring Culture (2014)

Editorial Board, *Strategic Management Journal*, 2007–present

Editorial Board, *Journal of International Business Studies*, 2003–present

Editorial Board, *Global Strategy Journal*, 2010–present. Co-Editor of Special Issue on Managing Global Stakeholders (2013)

Editorial Board, *Strategic Organization*, 2006–present

Editorial Board, *Long Range Planning*, 2013–present. Co-Editor of special issue on Innovating for Sustainability (2015)

Editorial Board, *Corporate Governance: An International Review*, 2007–present. Co-Editor of Special Issue on CSR, Institutions and Governance (2012)

Editorial Board, *European Journal of International Management*, 2010–present

Editorial Board, *Management Organization Review*, 2013–present

Editorial Board, *Journal of Product Innovation Management*, 1991–present. Editor of special issue on the Internationalization of Innovation (2000)

Editorial Board, *Journal of Strategy & Management*, 2008–present

Editorial Board, *European Management Review*, 2005–present

Editorial Board, *Asia Pacific Journal of Management*, 2003–present

Editorial Board, *BuR–Business Research*, 2007–present

Editorial Board, *Journal of Small Business Economics*, 1993–2002

(a) Examples of Research Support Received

1993–1994 Andrus Foundation, Washington D.C., Age as a Rating Factor in Automobile Insurance Pricing (with B. Cooil, Vanderbilt University).

1999–2002 SRG and UCG, Hong Kong, Measuring the Utility Value of Ethical Consumerism (with Patrice Auger, City University HK, and Jordan Louviere, Sydney) (SRG) (ARC) (granted 1/6/99)

2002–2003 CRC for Smart Internet Technology (Linkage Grant), Measuring Customer Response to Radical Future Technologies (with J. Louviere, UTS and industry partners Westpac) in 2004.

2003–2005 Australian Research Council (Discovery Grant), Cross-Cultural Differences in Perceptions Of Consumption Ethics (with G. Eckhardt, AGSM, and R. Belk, Utah).

2003 Nokia/Telstra, Using Lead User Research to Determine the Demand for 3G Service Delivery .

2003–2005 Transurban, Discrete Choice Modeling of Infrequent Road Users (with J. Louviere, UTS).

- 2004 AIM Fellowship (ESRC UK), Performance of UK Firms (with G. Yip and G. Johnson).
- 2005–2007 Australian Research Council (Linkage Program), Patterns of Rural Segmentation (with J. Louviere and S. Gudergan, UTS) (Described at: <http://www.ruralchoice.com.au>).
- 2005–2007 Australian Research Council (Discovery Grant), Information Provision and the Valuation of Social Attributes (with P. Auger, MBS, A. Gunnthorsdottir, AGSM, J. Louviere and M. King, UTS).
- 2006–2009 Australian Research Council (Linkage Program; Bluescope Steel), A Simulation Based Approach to Understanding Alternative Supply Chain Configurations (with T. Coltman, J. Gattorna and T. Spedding).
- 2007–2010 Australian Research Council (Linkage Program; ACT Health), An Action Research Project to Strengthen Inter-Professional Learning and Practice Across the ACT Health System (with J. Braithwaite, R. Iedema, J. Westbrook, R. Foxwell, R. Boyce, K. Murphy, M.-A. Ryall, J. Beutel, M. Budge, W. Ramsey).
- 2009–2012 Australian Research Council (Discovery Program), The Value of CSR to Close Stakeholders: A Discrete Choice Modelling Approach (with P. Auger, MBS, and G. Dowling).
- 2010–2014 Australian Research Council (Discovery Program), Extreme Values: The Anatomy of Civil Society Supporters and Protest Groups (with R. Belk, York U., J.

Schwalbach, HU-Berlin, P. Auger, MBS and A. Gunnthorsdottir).

2010-2012 Alexander von Humboldt Foundation. Research Reward (renewed).

2012–2016 Australian Research Council (Linkage Program), Technology and Innovation Management in High Risk Situations (with T. Coltman, R. Sharma, S. Gudergan, B. Brooks and N. Lin).

(b) Professional Consulting Experience (Selection)

Management consultant for various organisations including large corporations—United Press International, IMS/Dun & Bradstreet (London), Apple Computer, Martin-Marietta (Department of Energy), NationsBank, Dominion Bank, Nuturn Corporation, The Tennessee Valley Authority, LG (Seoul), Permanent General Insurance, Boral, AT Kearney (London), GEC-Alsthom, AMP, Morgan & Banks/TMP, GM/Holden, CSR, Mobil, Koppers Industries, SAP, Rolls Royce (UK), SAS Institute, Telecom Austria, Hanimex/Rabbit Photo, Thomson Publishing, Transurban, Nokia, Telstra, Sabanci Holdings (Turkey), Borusan (Turkey), Anadolu (Turkey) and Westfield Holdings.

Government and non-profits – the State Council of the PRC (China), the Government of PNG, Amnesty International, Greenpeace, WWF, The Property Council of Australia, Invest Australia, Australian Manufacturing Council and the City of Sydney.

8. EXHIBIT THREE – SELECTED PUBLICATIONS

Ex.3.1 *Philosophy of Science and Meta Knowledge in International Business and Management*, *Advances in International Management* (Volume 26), Emerald, 2013 (co-editor) (with T. Pedersen and L. Tihanyi).

Ex.3.2 “Modeling the Operational Capabilities for Customized and Commoditized Services,” *Journal of Operations Management*, 31, 7-8, November 2013 (with T. Coltman).

Ex.3.3 “A Research Agenda for Global Stakeholder Strategy,” *Global Strategy Journal*, 3, 4, November 2013 (with A. McGahan & M. Zollo).

Ex.3.4 “Corporate Social Responsibility and Corporate Governance: Comparative Perspectives,” *Corporate Governance: An International Review*, 21, 5, September 2013 (with J. Schwalbach & C.A. Williams).

Ex.3.5 “Is Microfoundational Thinking Critical to Management Thought and Practice?” *Academy of Management Perspectives*, 27, 2, May 2013.

Ex.3.6 “How Much Does a Company’s Reputation Matter in Recruiting?” *MIT Sloan Management Review*, Spring, 2013 (with P. Auger, G. Dowling, C. Eckert & N. Lin).

Ex.3.7 “Taking the “Non-“ out of “Non-Market” Strategy,” *Global Strategy Journal*, 3, 2, May 2013.

Ex.3.8 “Continuing Differences Between Health Professions’ Attitudes: The Saga of Accomplishing Systems-Wide Interprofessionalism,” *International Journal for Quality in Health Care*, 25, 1, February 2013 (with J. Braithwaite & 6 others).

Ex.3.9 *Institutional Theory in International Business and Management*, *Advances in International Management* (Volume 25), Emerald, 2012 (co-editor) (with T. Pedersen and L. Tihanyi).

Ex.3.10 “Perspectives on the Art and Science of Management Scholarship,” *Academy of Management Perspectives*, 26, 1, February 2012 (with D. Seigel).

Ex.3.11 “Social Responsibility, Global Strategy and the Multinational Enterprise: Global Monitory Democracy and the Meaning of Place and Space,” *Global Strategy Journal*, 1, 3-42 (launch issue), 2011.

Ex.3.12 “Customer Relationship Management and Firm Performance,” *Journal of Information Technology*, 26, 3, 2011 (with T. Coltman & D. Midgley).

Ex.3.13 “Dynamics of Globalization: Location-Specific Advantages or Liability of Foreignness?” *Advances in International Management* (Volume 24), Emerald, 2011 (co-editor) (with C. G. Asmussen, T. Pedersen and L. Tihanyi).

Ex.3.14 “Using Frontier Analysis to Evaluate Company Performance,” *British Journal of Management*, 21, 4, December 2010 (with G. Yip & G. Johnson).

Ex.3.15 “The Myth of the Ethical Consumer,” Cambridge: Cambridge University Press, 2010 (with P. Auger and G. Eckhardt).

Ex.3.16 “The Past, Present and Future of International Business and Management,” *Advances in International Management* (Volume 23), Emerald, 2010 (co-editor) (with T. Pedersen and L. Tihanyi).

Ex.3.17 “The Importance of Intangible Social Attributes in Consumer Purchasing Decisions: A Multi Country Comparative Study,” *International Business Review*, 19, 2 (with P. Auger, J. Louviere & P. Burke). A variant of this paper was nominated for the Carolyn Dexter Award for the best international paper at the 2007 AOM Conference.

Ex.3.18 “Measuring Organizational Performance as a Dependent Variable: Towards Methodological Best Practice,” *Journal of Management*, 35, 3, June 2009 (with P. Richard, G. Yip & G. Johnson).

Ex.3.19 “Measuring Long Term Superior Performance: The UK’s Long-Term Financial Performers, 1983-2004,” *Long Range Planning*, 42, 3, June 2009 (with G. Yip & G. Johnson).

Ex.3.20 “Is The Socially Responsible Corporation a Myth? The Good, Bad and Ugly of Corporate Social Responsibility,” *Academy of Management Perspectives*, 23, 2, May 2009 (shortlisted for best paper of 2009).

Ex.3.21 “The Financial Times Business Schools Ranking: What Quality is This Signal of Quality?” *European Management Review*, 5, 4, Winter 2008 (with G. Dowling & N. Perm-Ajchariyawong). There are three additional commentaries on this paper in the same issue.

Ex.3.22 “Formative versus Reflective Measurement Models: Two Applications of Formative Measurement,” *J. Business Research*, 61, 12, December 2008 (with T. Coltman, D. Midgley & S. Venaik).

Ex.3.23 “Do Social Product Features Have Value to Consumers?” *International Journal of Research in Marketing*, 25, 3, September 2008 (with P. Auger, J. Louviere & P. Burke).

Ex.3.24 “Do Managers Behave as Theory Suggests? A Choice Theoretic Examination of Foreign Direct Investment Location Decision Making,” *Journal of International Business Studies*, 38, 7, December 2007 (with P. Buckley & J. Louviere) best paper finalist, AIB Conference, Quebec City, 2005.

Ex.3.25 “Do What Consumers Say Matter? The Misalignment of Preferences with Unconstrained Ethical Intentions,” *Journal of Business Ethics*, 76, 4, December 2007 (with P. Auger).

Ex.3.26 “Using Best-Worst Scaling Methodology to Investigate Consumer Ethical Beliefs Across Countries,” *Journal of Business Ethics*, 70, 3, February 2007 (with P. Auger & J. Louviere).

Ex.3.27 “The Other CSR,” *Stanford Social Innovation Review*, Fall 2006 (with P. Auger, G. Eckhardt & T. Birtchnell).

Ex.3.28 “Dual Paths to Performance: The Impact of Global Pressures on MNC Subsidiary Conduct and Performance,” *Journal of International Business Studies*, 36, 6, 2005 (with D. Midgley & S. Venaik).

Ex.3.29 “Modular Strategies: B2B Technology and Architectural Knowledge,” *California Management Review*, 47, 4, Summer 2005 (with P. Richard).

Ex.3.30 “Client and Agency Mental Models in Evaluating Advertising,” *International Journal of Advertising*, 24, 1, 2005 (with M. Collins & G. Dowling).

Ex.3.31 “A New Perspective on the Integration-Responsiveness Pressures Confronting Multinational Firms,” *Management International Review*, 44, S11, 2004 (with D. Midgley & S. Venaik). Reprinted in: J.A. Krug and J.D. Daniels, *Multinational Enterprise Theory*, Thousand Oaks, CA, 2007.

Ex.3.32 “What Will Consumers Pay for Social Product Features?” (with P. Auger, J. Louviere & P. Burke), *Journal of Business Ethics*, 42, 3, 2003.

Ex.3.33 “Knowledge Management: Philosophy, Process, and Pitfalls,” *California Management Review*, 44, 4, Summer 2002 (with A. Deering, D. Midgley, & C. Soo). Best Paper, 2001 ANZAM Conference.

Ex.3.34 “Managing the Global Corporation: Case Studies in Strategy and Management,” 2nd Edition, New York: McGraw-Hill, 2001 (with José de la Torre & Yves Doz).

Ex.3.35 “E-Business: Revolution, Evolution or Hype?” California Management Review, 44, 1, Fall 2001 (with D. Midgley, T. Coltman & A. Latekefu).

Ex.3.36 “The Organisational Imperative and the Optimal Performance of the Global Firm: Formalising and Extending the Integration-Responsiveness Framework,” Organization Science, 11, 6, 2000 (with D. Midgley & S. Venaik).

Ex.3.37 “Understanding Institutional Designs Within Marketing Value Systems,” Journal of Marketing, 63, Special Issue, 1999 (with S. Carson, G. John & G. Dowling).

Ex.3.38 “Paying the Piper an Incentive to Play a Better Tune: Understanding and Resolving Advertiser-Agency Conflicts,” Journal of Business-to-Business Marketing, 6, 1, Spring 1999 (with G. Dowling).

Ex.3.39 “A Formal Model of Trust Based on Outcomes,” Academy of Management Review, 23, 3, July 1998 (with R. Bhattacharya & M. Pillutla).

Ex.3.40 “The Essence of Corporate Strategy: Theory for Modern Decision Making, Sydney,” Allen & Unwin, 1997 (with Jeremy Davis).

Ex.3.41 “How Well Do Patents Measure New Product Activity?” Economic Letters, 41, April 1993.

Ex.3.42 “New Products and Financial Risk Changes,” Journal of Product Innovation Management, 9, September 1992.

Ex.3.43 “New Product Innovations and Stock Price Performance,” Journal of Business Finance & Accounting, 19, September 1992 (with P. Chaney).

Ex.3.44 “The Return to Advertising Expenditure,” Marketing Letters, 3, May 1992, (with B. Cooil).

Ex.3.45 “European Markets After 1992: Implications for Business Strategy,” Lexington, MA: Lexington Books, 1991, (with William C. Hightower).

Ex.3.46 “The Impact of New Product Introductions on the Market Value of Firms”, *Journal of Business*, 64, October 1991, (with P. Chaney and R. Winer).

Ex.3.47 “New Products Over The Business Cycle,” *Journal of Product Innovation Management*, 7, December 1990.

Ex.3.48 “Diversification Strategy and Performance in Canadian Manufacturing Firms”, *Strategic Management Journal*, 11, September 1990 (with T. Nguyen and A. Seror).

Ex.3.49 “Rationally Determined Irrationality: An Extension of the Thesis of Rationality as Anti-Entropic,” *Journal of Economic Psychology*, 10, November 1989.

Ex.3.50 Numerous other articles in other journals, books, and magazines. Note that none of the above includes book chapters or other publications, patents or magazine/newspaper publications that have been excluded for space reasons.

9. EXHIBIT FOUR – OTHER MATERIALS CONSIDERED

Ex.4.1 In the process of preparing this report, I have identified a number of other studies and materials which appear to be related to the issue of plain packaging and which have not previously been considered by Dr Keegan in the Reports. These additional materials can be categorised as follows:

- (a) Studies and other materials which are related to plain packaging as a regulatory initiative (either by their conclusions or their content), but which do not generate any original consumer survey evidence in this regard; and/or
- (b) Studies and other materials which are potentially relevant to plain packaging as a regulatory initiative, but for which the survey analysis does not appear to be publicly available, therefore preventing any meaningful analysis.

Ex.4.2 Given that the scope of my report addresses the extent to which publicly available consumer surveys provide reliable evidence that plain packaging will achieve the public policy goals set out in my report at paragraph 1.5, I have not considered studies or other materials which fall into the above criteria in formulating my conclusions in this Report. For completeness, however I set out below a list of the studies that I have identified and which fall into the above criteria.

Studies which are related to plain packaging as a regulatory initiative but which do not generate any consumer survey evidence in respect of plain packaging

- (a) Stead M, Moodie C, Angus K, Bauld L, McNeill A, Thomas J, Hastings G, Hinds K, O'Mara-Eves A, Kwan I, Purves RI and Bryce SL (2013). Is consumer response to plain/standardised tobacco packaging consistent with Framework Convention on Tobacco Control guidelines? A systematic review of quantitative studies. *PLoS One*, 8(10): e75919.
- (b) Greenland, SJ (2013). Cigarette brand variant portfolio strategy and the use of colour in a darkening market. *Tobacco Control*, published online 29th August 2013.
- (c) Roland Berger (2013). The Potential Economic Impact of Plain Packaging for Cigarettes and Fine-Cut Tobacco in Ireland.

- (d) Kotnowski, K and D Hammond (2013). The impact of cigarette pack shape, size and opening: evidence from tobacco company documents. *Addiction* doi:10.1111/add.12183.b
- (e) McCool, J, Hoek, J, Edwards, R, Thomson, G and H Gifford (2012). Crossing the Smoking Divide for Young Adults: Expressions of Stigma and Identity Among Smokers and Nonsmokers. *Nicotine and Tobacco Research Advance Access Published 4 September 2012*

Studies which appear to be related to the issue of plain packaging as a regulatory initiative, but for which survey analysis does not appear to be publicly available

- (a) An opinion poll undertaken by Ignite Research in Ireland in September 2013.¹¹⁵

¹¹⁵ See: <http://www.businessandleadership.com/marketing/item/43846-minister-reilly-to-press-on/>.