

**Analysis of Consumer Research Evidence on the
Impact of Plain Packaging for Tobacco Products
(Updated to 2012)
Prepared by:
Professor Timothy M. Devinney**

29 June 2012

Table of Contents

1.	REPORT INTRODUCTION	2
2.	CRITERIA FOR EVALUATION	3
3.	SELECTION OF THE STUDIES	18
4.	REVIEW OF THE STUDIES	18
	CLASSIFICATION OF STUDIES	18
	DETAILED ANALYSIS OF THE STUDIES	22
	Gendall et al. (2011).....	22
	Moodie et al. (2011).....	27
	Hammond, Doxey et al. (2011).....	31
	Munafò et al. (2011).....	37
	Moodie, Hastings et al. (2011).....	40
	Thrasher et al. (2011).....	42
	Gallopel-Morvan et al. (2011)	44
	Borland et al. (2011)	49
	Moodie, Ford et al. (2011)	58
	Al-Hamdani (2011)	65
	Moodie & Ford (2011).....	70
	Hoek et al. (2012).....	73
	McCool, Webb et al. (2012).....	75
	Wakefield, Germain et al. (2012).....	79
	CRUK (2012).....	82
	Carter & Chapman (2006).....	85
	GfK Bluemoon (2011)	86
5.	CONCLUSION	96
6.	EXHIBIT ONE – QUALIFICATIONS OF PROFESSOR DEVINNEY	109
7.	EXHIBIT TWO – RESUME	111
	Examples of Research Support Received	115
	Professional Consulting Experience (Selection).....	117
8.	EXHIBIT THREE – SELECTED PUBLICATIONS	118
9.	EXHIBIT FOUR – OTHER MATERIALS CONSIDERED	123
	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	123
	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	125

1. REPORT INTRODUCTION

1.1 My full name is Professor Timothy M. Devinney. I am a Professor of Strategy at the University of Technology, Sydney (“UTS”) in Australia. 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. Most recently, I have been involved 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 résumé and a list of a sample of 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 author of the report “Analysis of Consumer Research Evidence on the Impact of Plain Packaging for Tobacco Products” dated 30 November 2010 (the “2010 Report”).

1.4 In my 2010 Report, I concluded that the publicly available consumer surveys and experiments do not provide reliable evidence that plain packaging would be effective in changing actual smoking behaviour.

1.5 Since the completion of my 2010 Report, several additional studies purporting to be in support of plain packaging have been published or have otherwise been 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 Report and to address the extent to which they provide 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 initiation) among minors;
- (b) reducing smoking consumption 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 and consistent with the finding in my 2010 Report, it is my expert opinion that none of the Studies provide 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.

1.7 For ease of exposition, I repeat the evaluation criteria employed in my 2010 Report and 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 Report I build on the criteria outlined by Dr. Keegan.¹ For ease of exposition, I repeat the outline I presented in my 2010 Report as way of a summary of the criteria discussed in detail by 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 Report 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),² the American Association for Public Opinion Research (AAPOR),³ the World Association for Public Opinion Research (WAPOR),⁴ the Council of American Research Organizations (CASRO),⁵ the International Statistical Institute (ISI),⁶ the Canadian Market

¹ As discussed in paragraph 1.3 of my 2010 Report, 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".

² Market Research Society (2005), *MRS Code of Conduct*. <http://www.mrs.org.uk/standards/downloads/code2005.pdf>.

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

⁴ World Association for Public Opinion Research, *WAPOR Code of Professional Ethics and Practices*. <http://www.unl.edu/WAPOR/ethics.html>.

⁵ Council of American Research Organizations (2008), *Code of Ethics and Standards for Survey Research*. <http://www.casro.org/pdfs/CodeVertical-FINAL.pdf>.

⁶ International Statistical Institute (1985), *Declaration on Professional Ethics*. <http://isi.cbs.nl/ethics.htm>.

Research and Intelligence Association (MRIA),⁷ and the European Society for Opinion and Marketing Research (ESOMAR).⁸

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 Report and are repeated here for ease of exposition. Specifically:

- (a) Because plain packaging does not currently exist and consumers are being asked to state an ‘intention’ relating to purchasing circumstances that are not currently available 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

⁷ Market Research and Intelligence Association (2007), *Code of Conduct for Members*. <http://www.mria-arim.ca/STANDARDS/CODE2007.asp>.

⁸ 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/ICCESOMAR_Code_English_.pdf.

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.⁹ Good research will attempt to reduce this problem by focusing on three factors, which I will add to my list of criteria:¹⁰

(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:

⁹ 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.

¹⁰ 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.

- 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 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.¹¹ Statistical efficiency is important because most experiments cannot possibly have all individuals look at every possible combination of products 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**.¹² In product choice

¹¹ 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.

¹² 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

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.¹³ This is achieved by the fact that 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.¹⁴ 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.

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.

¹³ 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 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.

¹⁴ 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.

		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**.¹⁵ 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 2 times each 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 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.

¹⁵ 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).

2.11 Point (a) in paragraph 2.3 relates also to intentionality. Morwitz, Steckel and Gupta¹⁶ 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 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.

¹⁶ Morowitz, V., Steckel, J. and A. Gupta (2007) “When Do Purchase Intentions Predict Sales?” *International Journal of Forecasting*, 23: 347-364.

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

Evaluative Criteria	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 opportunity to give a “don’t know” or “no opinion” answer.
Interviewer response bias	To the extent that it may bias the results, neither respondents nor persons

Evaluative Criteria	Short Description of Criteria and Its Relevance
	responsible for the data collection should be informed as to the sponsor or purpose of the study.
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>

Reliability by Data Collection Method

Reliability	Data Collection Method	Research Type
Most reliable	Direct observation	Observed Behavioural
↓	Recent recall of behaviour	Observed Behavioural

Evaluative Criteria	Short Description of Criteria and Its Relevance	
↓	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.¹⁷ 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 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

¹⁷Churchill, G., and D. Iacobucci (2005), *Marketing Research: Methodological Foundations, 9th Edition*. Orlando, FL: Dryden, 387-390.

Evaluative Criteria	Short Description of Criteria and Its Relevance
Focus groups	<p>reliable data.</p> <p>Focus group studies are exploratory. They generate hypotheses rather than findings that can be generalised to a wider population.¹⁸</p> <p>The reported findings of focus groups often have no statistical significance due to the small sample size and informal nature of the responses.</p>

5. Analysis

Statistical significance	<p>It is imperative that authors refrain from projecting results that are not statistically significant to general populations or markets.¹⁹</p> <p>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 and not caused by chance, error or other factors.²⁰</p>
--------------------------	--

¹⁸ Churchill, G., and D. Iacobucci (2005), *Marketing Research: Methodological Foundations, 9th Edition*. Orlando, FL: Dryden, 81-85.

¹⁹International Statistical Institute (1985), *Declaration on Professional Ethics*. <http://isi.cbs.nl/ethics.htm>, paragraphs 1.3 and 3.1.

²⁰Thompson, B. (1994), "The Concept of Statistical Significance Testing," *Practical Assessment, Research & Evaluation*, 4.5.

Evaluative Criteria	Short Description of Criteria and Its Relevance
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. ²¹
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.

²¹ 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. 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 above.

4. REVIEW OF THE STUDIES

CLASSIFICATION OF STUDIES

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

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

4.2 My review in this report has focused on key primary research studies published or otherwise obtained since my 2010 Report 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) 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. Hereinafter referred to as "Gendall et al. (2011)."
- (b) 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. Hereinafter referred to as “Moodie et al. (2011).”

- (c) 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. Hereinafter referred to as “Hammond, Doxey et al. (2011).”
- (d) 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. Hereinafter referred to as “Munafo et al. (2011).”
- (e) 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. Hereinafter referred to as “Moodie, Hastings et al. (2011).”
- (f) 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. Hereinafter referred to as “Thrasher et al. (2011).”
- (g) Gallopel-Morvan, K., Moodie, C., Hammond, D. et al. (2011). Consumer perceptions of cigarette pack design in France: a comparison of regular, limited edition and plain packaging. *Tobacco Control*, Online First, published on October 13, 2011. Hereinafter referred to as “Gallopel-Morvan et al. (2011).”

- (h) 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. Hereinafter referred to as "Borland et al. (2011)."
- (i) 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. Hereinafter referred to as "Moodie, Ford et al. (2011)."
- (j) 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. Hereinafter referred to as "Al-Hamdani (2011)."
- (k) 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. Hereinafter referred to as "Moodie & Ford (2011)."
- (l) Hoek, J., Gendall, P., et al. (2012). Tobacco branding, plain packaging, pictorial warnings, and symbolic consumption. *Qualitative Health Research*, 22 (5): 630-639. Hereinafter referred to as "Hoek et al. (2012)."
- (m) 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. Hereinafter referred to as "McCool, Webb et al. (2012)."

- (n) 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. Hereinafter referred to as “Wakefield, Germain et al. (2012).”
- (o) Cancer Research UK (2012). The packaging of tobacco products. The Centre for Tobacco Control Research. Hereinafter referred to as “CRUK (2012).”
- (p) 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).”²²
- (q) 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).”

DETAILED ANALYSIS OF THE STUDIES

Gendall et al. (2011)

4.3 This study is a mixed method interview and brand association study conducted with 66 adult smokers and non-smokers in New Zealand (aged 18 to 24 years old). It attempts to assess the degree to which brand representations and imagery influence a brand’s appeal. Its relevance to the plain packaging debate is seen in the conclusion that states: “[*These findings*] are strikingly at odds with tobacco companies’ claims that plain packaging

²² As I discuss further below, 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 has recently been 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).

would not reduce the appeal, attractiveness, or the uptake of smoking”.²³ However, the relevance of the brand image associations uncovered in the investigation to changes in behaviour is not established, as they are never investigated. Indeed the authors acknowledge this as a limitation: “*The fact that young adults associate brands with specific attributes does not demonstrate that tobacco branding increases the risk of smoking experimentation or uptake or that plain packaging would reduce these behaviors and prompt cessation*”.²⁴ Their conclusion – that the prior statement is qualified by the fact that “*it logically suggest[s] both*”²⁵ – is based on an assumption that the degree of “appeal” found or the related discussions of brand imagery are determinants of smoking initiation and cessation behaviour in realistic contexts. However, this link is not established.

Analysis

4.4 In my opinion, this study suffers from a number of methodological limitations. These relate to: (a) the sampling frame, (b) the ability to draw effective conclusions, and (c) the relationship to actual behavioural responses and market realities.

4.5 First, the sample size is limited to 66 young adult smokers and non-smokers. This significantly limits any conclusions from a statistical standpoint, which will be discussed below. Second, it makes any generalisations to a general population impossible. Although the authors indicate that they “*sought [participants] from all socioeconomic strata to maximize diversity and promote debate*”,²⁶ the limited sample size ensures

²³ See page 6, Gendall et al. (2011).

²⁴ Ibid.

²⁵ Ibid.

²⁶ Ibid.

that it is unlikely to be representative of the general population and that the conclusions can only be viewed as relevant to the sample itself and nothing beyond that sample with any degree of confidence. In addition, any specific inferences, such as differences between smokers and non-smokers, ethnic groups and educational attainment cannot be justified given the very limited number of potential points of comparison amongst the groups.

4.6 A more substantive issue is related to the brand association task. The researchers asked each individual to associate any number of 22 words with the specific brands. What is meant to be derived from this task is never made clear, nor is a link ever established between the conclusions drawn and actual behaviour. The conclusion: “*Participants ascribed very different images to different brands when exposed to the packaging alone, regardless of whether they had seen or heard of the brands before*”²⁷ is troubling in that it may imply that false associations will arise simply due to the exercise. This can be seen in two examples.

- (a) First, the data presented in Table 1 and Figure 2 appear to indicate that the Camel and Kool brands are competing on many of the same associations and that Camel is much more a “feminine” brand. Based upon the conclusions of the authors, it is suggested that companies are proactively using the brand and packaging imagery to create associations. Yet, there is no indication that the associations that are uncovered relate to what the manufacturers themselves are supposedly attempting to achieve. This is further compounded by the fact that we do not know (based on the information presented in the publication) whether the participants have any specific experience of smoking the brands and whether or not they are actual users of the brands. The fact that the sample includes non-smokers (the actual

²⁷ See page 1, Gendall et al. (2011).

numbers are not revealed in the publication) implies that the associations arising may be completely spurious as the assumption that some of the brands are “known” because they are sold in New Zealand would only apply if the individuals knew this to be the case and knew enough to have anything more than cue-attributed reactions to the stimuli being presented to them.

- (b) Second, brand association tasks can bring to the fore the problem of what are known as “pseudo opinions” – something discussed in detail in my 2010 Report. The act of asking for an association demands an answer for which the individual feels responsible for having an opinion and this is exacerbated when the issue is less important to them but is made salient by the sheer fact it is being researched. As noted in my 2010 Report, Bishop et al. (1980) utilised a series of polls that asked for people’s opinions on real and fictitious laws and found that 1/3rd of people gave responses concerning their opinions on fictitious laws. As they noted:

*“Of greater significance to many researchers is the question of whether respondents who offer opinions on the US Public Affairs Act [the fictitious law] will do the same on topics that are real but not particularly salient in their daily lives. Our results tell us that such people were indeed more likely to express an opinion on all other issues we investigated. This was particularly true ... for the more abstract matters of policy. ... Apparently, the more remote the topic becomes from day-to-day concerns the greater is the effect of this predisposition.”*²⁸

Hence, it is entirely conceivable that the subjects in these studies are expressing “pseudo opinions” – or in this case “pseudo associations”

²⁸ Bishop, G., Oldendick, R., Tuchfarber, A. and S. Bennet (1980), “Pseudo-Opinions on Public Affairs,” *Public Opinion Quarterly*, 44: 198-209, page 202.

– about brand options over which they have no knowledge but are picking up on vague or random cues. This is induced simply by the sheer fact that associations are being requested as part of the research and the participants’ logical tendencies will lead to what appear to be relevant patterns for no other reason than people are good at creating patterns even where none exist.

4.7 The issue of pseudo associations can be avoided by presenting the task in a manner that allows individuals not to respond when they do not have the knowledge to respond. It can also be addressed by attempting to extract the degree to which the individuals are simply engaging in associations that appear logical to them but may ultimately have no meaningful relationship. This can be achieved by having the individuals undertake another association task to determine the degree to which they can logically create associations based simply on the task at hand, or for such links to be controlled by having completely fictitious brands to determine what the natural base rate of quasi-random associations would be.

4.8 A second issue with the brand association task is that individuals will seek consistency and separation when presented with such a set of stimuli. When presented with different brands and a list, the natural assumption that the respondent makes is that there will be associations that matter. Why would the research be being done if this were not the case? In addition, there will be natural patterns in the associations – e.g., I will not label the same brand as “masculine” and “feminine” or “plain” and “attractive” or “traditional” and “trendy”. Hence, the task will cause individuals to naturally block logical and consistent attributes (which, for the most part is what is seen). Anyone presented with this task, even if they have no associations, will create associations because that is what the task is all about. Hence, all that may be arising is that the researchers are finding that individuals will make logical associations because they are cognitively wired to do this. So there is an aspect of this that can be seen. The brand Basic is viewed as very

basic (e.g., “budget”, “plain”, “ordinary”) and one would have expected this just by the brand name. Similarly, Holiday is “relaxing”. And the other brands show patterns that lead to the statistical conclusion that there is a difference. However, we do not know whether or not these associations are just being induced by the task or are functionally meaningful. The fact that individuals can discuss these differences in open-ended interviews can simply arise by attempting to avoid cognitive dissonance – the distress that arises when people are presented with their own inconsistencies.²⁹

Conclusion

4.9 Ultimately, this study simply reveals that when presented with an association task, participants in a study will engage in creating associations. The materiality of those associations and their relevance to actual behaviour is never established by the authors. In addition, the small sample size makes any generalisations impossible. Accordingly, I do not consider this study reliable evidence as to whether plain packaging would help to achieve the public policy goals listed at paragraph 1.5 above.

Moodie et al. (2011)

4.10 This study sought to examine how using plain packs influenced pack perceptions, feelings about smoking, salience of health warnings and self-reported cigarette usage amongst a small sample of 48 adult smokers (aged 18 to 35) in Glasgow, Scotland. The authors conclude that their results support the contention that *“plain packaging could potentially help reduce tobacco consumption among young adult smokers, and women in*

²⁹ The link between cognitive dissonance and brand association is well established in the marketing literature and goes back almost 50 years. One of the earliest discussions is Stanley Kalish (“Cognitive Dissonance and the Classification of Consumer Goods”, *Journal of Marketing*, 31 (4) (1967): 28-31) who notes that “since goods offered in the marketplace are designed by the seller rather than by the buyer, they generally embody combinations of characteristics that are not exactly as the buyer would have them if he were customising the product to his own taste”. Consumers fill in the gaps by creating associations that ex post align the product association to what the consumer seeks even when no such association is intended by the producer.

particular".³⁰ The results, however, provide only limited information as to the true behavioural impact of plain packaging primarily because: (a) the final sample is small (attrition from the main study was greater than 50%); (b) the actual smoking behaviour was not monitored; and (c) it was unclear whether or not there is any validity to the interview and survey responses on which the main conclusions are drawn.

Analysis

4.11 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 to plain packages there is an attempt to get closer to the use environment in which smoking is done. However, this examination 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. This would have: (a) removed the artifact 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.

4.12 Although this study is field-based, it fails on two basic dimensions. First, there are no actual behavioural usage measures. All that is collected is perceptual self-reports. Second, this issue is made worse by the very small sample size, which makes interpreting the results problematic. Hence, without any actual usage measures, and given the very small sample size, the ability to draw any policy inferences from the results is lost.³¹

³⁰ See page 367, Moodie et al. (2011).

³¹ 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.

4.13 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 brown 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. I may like that my Starbucks coffee (that I am currently drinking) is in a red cup and find that more appealing than if it came in a brown cup, but that may have no impact on whether or not I will buy another coffee tomorrow. Also, the fact that the participants in the study are being asked to use a plain pack and then compare these results against what they would do if they thought about their usual branded pack is clearly a study that is aimed at finding differences. Hence, one would expect to see bias creep into the study responses.

4.14 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.³²

³² The approach discussed here is not dissimilar to that given in several studies reported in Belli, R. F., Stafford, F. P., and Alwin, D. F. (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

4.15 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.

4.16 Ultimately, in my opinion, this study’s most significant limitation is the small sample size. Hence conclusions about specific actions or specific groups – such as women – are going to be erroneous. For example, the statement “*about half reported some change in their smoking behaviour when using the Kerrod’s pack, again mostly women*”³³ may sound impressive until one realises that what is being discussed is eight people, six of whom are women and the behavioural measures are recall measures. One simply cannot make scientifically generalisable statements based on these sample sizes nor is it likely that they serve as a basis of good policy judgment. This fact is made even worse when one takes into account the very large drop out rate for the study.

Conclusion

4.17 Based upon this analysis there is no indication from this study that plain packaging actually impacts smoking behaviour. Although this study

Survey Method for Characterizing Daily Life Experience: The Day Reconstruction Method”, *Science*, 306 (5702): 1776-1780. Today, most such studies employ SMS messaging to randomly ‘buzz’ respondents to measure specific behaviour activities at random points in time.

³³ See page 370, Moodie et al. (2011).

represents a more realistic examination of smoking perceptions than other studies I have reviewed, it is still significantly removed from actual behaviour and is saddled by a small sample size and significant attrition from those initially recruited into the study. Hence, it is my conclusion that it does not provide reliable evidence in respect of achieving the public policy goals set out in paragraph 1.5 above.

Hammond, Doxey et al. (2011)

4.18 This study is similar to Doxey (2009) that I reviewed in my 2010 Report at paragraphs 4.38 to 4.47. The structure of the study is also similar to those of Hammond, et al. (2009) and Hammond and Parkinson (2009)³⁴ and, fundamentally, suffers from most of the major experimental design and statistical analysis limitations of those studies. The study's relevance to plain packaging is the conclusion that "*plain packaging was associated with fewer false beliefs about the health risks of brands compared with branded female packs, consistent with previous research ... The findings also indicate that the design of packs influences the likelihood that young women will accept the offer of cigarettes*".³⁵ The sample consisted of 826 females, smokers and non-smokers, between the ages of 18 and 19 in the United States.

Analysis

4.19 This study suffers from a number of serious methodological flaws, many of which follow directly from the fact that they repeat the same basic form of research method used by Hammond, et al. (2009) and Hammond and Parkinson (2009). Most obviously, the lack of actual behavioural outcomes or incentive compatible measures that represent how individuals would make choices in the broader contexts of purchasing are quite serious. They undermine the relevance of any conclusions drawn about perceptions.

³⁴ I have discussed Hammond et al. (2009) and Hammond and Parkinson (2009) at paragraphs 4.3 to 4.15 and 4.16 to 4.22 respectively in my 2010 Report.

³⁵ See page 8, Hammond, Doxey et al. (2011).

4.20 This issue of the perceptions is made worse by the artificiality of the question design and measure construction. Questions relating to “tar delivery” (“*How much tar do you think these cigarettes would have compared to other cigarette brands?*”³⁶) and “health risks” (“*How would the health risks of these cigarettes compare to other cigarette brands?*”³⁷) 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 other words, my definition of “health risk” is comparable to yours). 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.

4.21 The research uses 5-point scales to measure the authors’ perception constructs – e.g., where the scales range from “a lot better” to “a lot worse”. However, these 5-point scales are then arbitrarily aggregated so that they are dichotomous (i.e., “1” and “0”). Such arbitrary aggregation is completely unacceptable based on the norms and standards of market research as it imposes *a priori* restrictions on the data for no other reason than the researcher wants to use a specific empirical approach and to create an index by aggregating up the 0-1 measures. As noted by Rossiter:³⁸ “*the formation of such an index requires that constructs of interest be conceptually defined (described) in terms of (1) the object, including its constituents or components, (2) the attribute, including its components, and (3) the rater*

³⁶ See page 3, Hammond, Doxey et al. (2011).

³⁷ Ibid.

³⁸ Rossiter, J. (2002), “The C-OAR-SE Procedure for Sale Development in Marketing”, *International Journal of Research in Marketing*, 19: 305-335.

entity. Failing this, the conceptual definition of the construct will be inadequate for indicating how the construct should be (operationally) measured.” What this means specifically is that (a) it must be clear what is being rated – the ‘object’ must be clear to those being asked to do the rating, (b) what makes up the construct is well articulated – that the ‘attributes’ of the object are understood and valid, and (c) all of the raters are comparable – in other words, the raters are knowledgeable and relevant. The last statement implies that without this clear articulation the measure is conceptually invalid.

4.22 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?”*.³⁹ The ratings asked 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” and “don’t know” received a score of “0”. In essence, this measure treats someone who has no idea (i.e., “don’t know”) in the same category as “a lot less appealing” and “no difference”.
- (b) “Health risk” perception ratings were determined by a question: *“Compared to other cigarette brands on the market, would these cigarettes be...less/more harmful?”*.⁴⁰ The ratings asked were then aggregated so that “a lot less risk” and “a little less risk” received a score of “1” and “no difference”, “a lot more risk”, “a little more risk” and “don’t know” received a score of “0”. In essence, this measure treats someone who states that they have no idea of the “health risks”

³⁹ See page 3, Hammond, Doxey et al. (2011).

⁴⁰ Ibid.

of a brand (and is willing to say so) in the same category as someone who views the “health risks” as high!

4.23 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 1-8 scale that is meant to represent an overall “Brand Appeal” index, “Perceived Taste” index, “Tar Delivery” index and “Health Risk” index. This creates what amounts to a theoretic formative measure for Appeal, Taste, Tar and Health that supposedly measure the overall perception of the eight packs seen by the subject. However, what this measure means is totally unclear. How it relates conceptually to the choices being asked of the subject in the experiments is also unclear. To use Rossiter’s formulation as described in my 2010 Report, it is unclear what the components of this index are. For example, although it might be clearer what an “8” and “0” meant for such scales, intermediate measures are completely confusing. If someone rated the eight brands as a “4” on “tar” this could arise because four of the brands were perceived to have a “lot less tar” and four of the brands were rated “Don’t know” or because four of the brands were perceived to have “a little more tar” and four “a little less tar”. In other words, there is no way to determine what actual perceptions led to the intermediate scores. Mean statistics are provided for these indices in Table 2. For the most part, these indices are skewed toward the lower end meaning that the vast majority of subjects end up rating most of the packs in a manner that leads to a “0”. But a “0” is the most confused score as it confuses “a little” and “some” with “don’t know” and “no difference”.

4.24 The experimental conditions are, in my opinion, confused. They do not allow for effective and efficient comparison of the package attributes, as the design is not efficient, orthogonal or balanced. For example, the brands appear different numbers of times and the “Male brands” that are meant as controls have completely different brand names, pack dimensions and colours. A properly designed study would control for brand effects,

dimension effects, colour effects, price and other package and product attributes and do so in a way that would have proper experimental design characteristics. This would allow for the measurement of specific attribute and level effects (essentially the components of the packaging), along with the interaction between those factors and other components of the product.

4.25 In my opinion, this limitation in the experimental design means that there is no logical link between the design and the statistical model being estimated. The vast majority of the analyses are based upon pair-wise comparisons (such as “standard” versus “plain” and on a “brand by brand” basis) where the scores are completely dependent upon the alternative against which they are being compared. What this implies is that, at best, the analysis only allows us to make a statement about that specific package in that specific experiment against that specific alternative. This is why the lack of orthogonality and balance is important. An orthogonal and balanced design allows the researcher to make statements about the components of the product independent of the alternatives. In terms of this study, one is unable to say anything generalisable about any of the results of the analysis.

4.26 Econometrically, the study has a number of flaws. Statistical efficiency would require that analysis of the preferences for specific brands be estimated using a pooled regression where the independent variables are conditional on the alternatives examined. In other words, because each participant will see a different mixture of eight packages we would want to know whether or not the comparison set influences the choices. It is good practice to control for the choice set in which the evaluations are being made. Some bias may be mitigated by the “one-at-a-time” approach utilised in the study but it is also likely that the mixture of items seen in the set of eight will have an influence. Normally, good research practice is to control for the items that are appearing in the evaluative set as a means of seeing whether any undue bias occurs. Simply making recourse to the fact that the assignments were random may or may not resolve comparator effects but it

does not necessarily test whether or not they are present. This is particularly the case as the design is not orthogonal and different individuals will be potentially seeing different brands (particularly with respect to the “non-female packs”).

4.27 Additionally the one dependent variable at a time comparison approach may exacerbate effects that are themselves part of a general tendency. Hence, examination of “brand appeal”, “tar delivery”, etc. are undertaken via regression analysis one at a time as if there might not be a relationship between these measured factors represented in the dependent variables. In other words, there is no attempt to control for the effects that might exist across these variables. It also ignores the correlational structure that exists between the various dependent variables.

4.28 Resolution of this can be achieved in a number of ways. First, one can attempt to use econometric structures that account for the correlation amongst the dependent variables, e.g., via utilisation of a MANOVA structure or via restructuring the data and applying a seemingly unrelated regression modelling approach. Second, one could utilise a structural equations modelling approach where there is: (a) direct modelling of the structure of the constructs (hence removing the arbitrary creation of indices); and (b) the relationship between the dependent and independent variables is pulled together in a general structural model that indicates how the various influences flow to the ultimate dependent variable or variables of interest. Such modelling is also increasingly relevant to marketing and consumer related studies. However, creating such models requires a theoretical structure that is completely lacking in this study.

4.29 The authors argue that their “Pack Selection Task” is a behavioural measure. However, this is not the case and any conclusion drawn from this would be erroneous. For example, the authors could just as easily have offered participants candy bars that were branded “Snickers” or in a brown

bag labelled “Snickers” and found that the majority decided not to choose the one that was offered in the brown bag. In my opinion, that would in no way indicate that putting a Snickers bar in a bag rather than in its packaging would imply that brown bags reduce females’ “susceptibility” to becoming candy consumers.

Conclusion

4.30 This study is methodologically flawed in terms of its basic design, execution and statistical analysis. The structure of the experimental tasks and questions is flawed and fails to meet the standards of good experimental research. The use of arbitrary aggregation without theoretical justification renders the conclusions invalid as they are subject to a plethora of other explanations. The statistical modelling is theoretical and does not align with any behavioural model. The results are all perception based and provide no evidence of the veracity of the findings to actual choices.

Munafò et al. (2011)

4.31 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 young adults, comprising weekly smokers ($n= 14$), daily smokers ($n=14$) and non-smokers ($n=15$). Testing took place at the University of Bristol. The authors conclude that while non-smokers and weekly smokers pay more visual attention to health warnings on plain packs than branded packs, daily smokers do not.

Analysis

4.32 This is a modest pre-test study that provides a more advanced methodological and technological approach to perceptions than some of the other studies I have reviewed and which purport to be relevant to plain packaging. Also, the authors are quite open as to the limitations of what they

have done, as articulated on page 13 of their manuscript. First, the sample is small and based on convenience sampling. Second, even the daily smokers are non-representative of UK smokers in general (smoking less on average). Third, the majority of the sample was male. And, finally and most importantly, they “*do not know what the consequent behavioural impact ... is likely to be; individuals may habituate to plain packaging over time, and it is unclear whether increased visual attention to health warnings will translate to differences in actual cigarette smoking behaviour*”.⁴¹

4.33 In addition to the above, I will discuss two other issues that raise some questions as to the importance of the study. First, the authors noted that “*before the experiment began, participants were informed that they would have to perform a recall task ... This was in order to ensure that the participants viewed the images attentively*”.⁴² Normally, something like this is done when one is examining cognitive process, primarily because the researcher wants the participant to focus their attention so that the final measure is valid and subject to less variance. However, in the case of this study, in seeking to assess where attention is focused, the issue is more related to policy outcomes. Such prompting may lead to circumstances where individuals with less experience with a product, in this case non-smokers and infrequent smokers, would naturally attempt to pay more attention because their cognitive schema is less developed and they are preparing for the recall task. Hence, the results found may be an artifact of the threat of the recall task. This could have easily been resolved by simply having a control group where such prompting was not given and the recall task was done as a surprise.

4.34 A second issue is that while this recall task was clearly undertaken, it is never analysed or shown in the results. Hence, we do not know whether or

⁴¹ See page 13, Munafò et al. (2011).

⁴² See page 8, Munafò et al. (2011).

not the differences in attention noted in Table 1 of the study even translated into differences in recall capabilities, or whether aspects of that recall capability were related to some other aspect of cognitive capability. Therefore, while measuring the eye movements was the key dependent variable, understanding whether those movements mattered is material to understanding whether there is any likelihood that the eye activity translated into actual cognitive activity. In my opinion, the lack of this information is quite puzzling and would have helped in explaining some of the differences seen in Table 1 of the study.

4.35 Eye movement studies have had a mixed history in marketing and behavioural research. Early on, the instrumentation was bulky and hence quite intrusive. It also provided many inaccurate readings. Hence, the interest in their use waned. More recently, the advent of better and more accurate instruments has led to such studies (along with brain scanning methods seen in neuropsychology) becoming more common. They provide a better direct measure of where visual attention is being concentrated and are an alternative in some cases to subjective recall tests. However, without additional information, the results of such research are not able to determine: (a) how that visual attention is being utilised; and (b) whether that visual attention translates into anything practically meaningful. In addition, 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).⁴³

4.36 This last point is quite critical with this study. As noted in Table 1 of the study, the differences between the various groups are not statistically

⁴³ 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 42 subjects in Munafo et al. (2011).

meaningful as the mean of each group easily fits into the 95% confidence interval of the other groups (whether one is reading across the groups (columns) or across the packaging (rows)). However, even this may be an understatement as the differences between individual comparisons are assuming that the “true” individual level variance is the same for everyone. Generally, one would not expect this to be the case as noted in the prior paragraph. Hence, it is likely that much of the differences seen will be attributable to individual level differences that are not related to the stimuli being used.

Conclusion

4.37 Although this study is an improvement technically over survey-based perceptions, it suffers from sample size limitations, no theoretical behavioural model and a lack of more sophisticated econometric analysis that would account for individual level differences in terms of eye movements and cognitive capabilities. In addition, 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.

Moodie, Hastings et al. (2011)

4.38 This study is a focus group examination of consumers’ perceptions of the impact of plain packaging on illicit tobacco sales. Its conclusions are that “*packaging, whether branded or plain, has no impact on the decision to consume counterfeit tobacco*”.⁴⁴

Analysis

4.39 A total of 54 adult smokers in Glasgow, Scotland, aged between 18 and 35, participated in the study. The study is limited first and foremost by

⁴⁴ See page 2, Moodie, Hastings et al. (2011).

the use of focus groups. As noted in my 2010 Report, which examined a number of such studies, focus groups are non-representative and subject to significant research and social bias effects that render their applicability to significant policy and commercial decisions severely limited. Most of the normal checks that one attributes to good research practice as outlined in Section 2 are violated. One cannot control for truthfulness or incentive compatibility nor can the focus group questioning be subject to statistical analysis that would determine the degree to which any statements or effects matter materially to specific policy outcomes. At best, focus groups can be used to work through nascent ideas, but even in this case, they must be viewed sceptically.

4.40 However, a major issue with this study is not just that it is based on focus group research, but that broad policy conclusions are drawn based on a discussion with a small unrepresentative sample with no controls for bias. For example, the authors ask whether or not the participant has had experience with counterfeit or smuggled tobacco and concludes that men had more experience than women. Yet we do not know what this means. How many men? How many fewer women? Nor do we know the degree to which the main conclusion – *“packaging, whether branded or plain, has no impact on the decision to consume counterfeit tobacco”*⁴⁵ is based on a very small number of people. For example, if only 10 people of 54 had this experience, is this a relevant number, even assuming that the statements that are made are true? If these 10 people only had one experience each, is that sufficient to draw any conclusions?

4.41 More troubling is that the researchers queried *“whether the participants felt that there would be any change in behaviour if all cigarettes came in identically coloured, non-branded (plain) packs”*.⁴⁶ In essence, the

⁴⁵ Ibid.

⁴⁶ Ibid.

researchers are asking unknowledgeable participants whether they have a mental model about behavioural change and then using that response to make inferences about the validity of policy outcomes. Although this may be of interest, the scientific validity of such an exercise is questionable at best. One might similarly ask individuals if they believe that introduction of a carbon tax will change driving behaviour or whether a Tobin tax would influence currency and financial fluctuations. While individuals might have a feeling, that feeling is not going to be valid in any useful way, particularly in drawing large-scale policy conclusions.

Conclusion

4.42 Overall, it would be erroneous to draw policy conclusions based on this study for the reasons noted above and in Table 2 of Section 2 of this report: *“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”*.

Thrasher et al. (2011)

4.43 This study uses a bidding experiment to examine the price elasticity of cigarettes. They find that individuals will bid less for plain pack cigarettes. Adult smokers aged 18 years and over ($n = 404$) in four states in the United States (California, Florida, South Carolina and Pennsylvania) participated in the experimental auctions.

Analysis

4.44 As noted by the authors, the use of experimental economic approaches is demand revealing and consistent with creating an environment that is incentive compatible. In my opinion, the use of such an approach means that the authors of this study are getting closer to measuring a

behavioural response: in this case the willingness to bid on cigarettes possessing different package characteristics.

4.45 However, the design of the experiment is incomplete in two senses. First, there is incompleteness in the balance of the design. One way of interpreting what is being done is that as one moves from experimental condition 1 to 2 to 3 to 4, the severity of the warning gets higher relative to the branding and this is obvious to all participants. Hence, individuals are always faced with circumstances where they are bidding on what is clearly a “superior” and “inferior” product. At best, we can determine whether they value this inferiority in any real sense. I would not be surprised if survey-based perceptions had been taken that the respondents would favour 1 over 2, 2 over 3, and 3 over 4, simply because the differences were obvious and the study’s intent was clear. However, one cannot conclude much beyond this.

4.46 Second, the design lacks orthogonality. What would have been more appropriate would be to vary conditions orthogonally. For example, one would vary conditions based on brand x text x location x packaging. What would then be discernable would be the true effects of each component independent of the others. In the case of this study, all we can say is that condition 1 is worse than condition 2, which is worse than condition 3, which is worse than condition 4 when there are ONLY these four alternatives. As I noted in my 2010 Report, one can have non-orthogonal designs but in this case there is little need to do this, as all the important attributes can be included easily.

4.47 Additionally, while the bidding task is an improvement over perceptual measures, it is unrealistic to evaluate the importance of plain packaging. If plain packaging becomes the norm, the most appropriate case would be what happens when ALL brands are in plain packaging. Again, this could be incorporated into an experimental task by adding additional products into the bidding mixture. This solves two problems in the study.

There is the fact that the study is clearly about cigarettes and health warnings on pack (as people see their preferred brand and two different packs). However, this can be hidden by having a basket of items on which they bid and the experimental design is the mixture of the types of products in the basket. One can then estimate the hedonic price of the various packs. In specific situations, as laid out in the experimental design, there will be plain pack versus plain pack alternatives, allowing the researcher to examine the price elasticity in very different market conditions.

Conclusion

4.48 The experimental design employed in this study is an improvement on the survey work prevalent in the studies that purport to be relevant to plain packaging. However, this study is limited by the lack of a fully balanced orthogonal design and the fact that the alternatives reveal quite blatantly what the study is about. In addition, the validity of the study with respect to the demand for plain pack cigarettes is not established, as it never investigates demand realistically in circumstances where all the products are available in plain packs. Hence, the generalisability of the results that people bid less for the plain packs or packs with larger health warnings provides no information as to what the ultimate impact of a general plain packaging policy on behaviour would be in circumstances when all of the products would essentially look identical. I therefore do not consider this study to be reliable evidence that plain packaging would influence actual smoking behaviour.

Gallopel-Morvan et al. (2011)

4.49 This study is an examination of perceptions of package design in a sample of 836 French adult smokers and non-smokers. The study concludes that “*novel cigarette packaging, in the form of limited edition packs, had the highest ratings of consumer appeal, ahead of regular branded packs and also [plain packs]. ... [Plain packs] were perceived to be the packs most likely to*

promote cessation among those adults with quitting intentions. Plain packaging, therefore, may be a means of helping existing adult smokers motivated to quit to do so".⁴⁷

Analysis

4.50 This study is an improvement on studies that rely solely on ratings scales. In this case, the analysis is based on quasi choices among a set of alternatives. The quasi choices are which pack of X "*was (1) most effective in getting attention, (2) most attractive, (3) most effective in convincing non-smokers not to start, (4) most effective in motivating smokers to quit, (5) most effective in motivating smokers to reduce consumption and (6) most effective for motivating youth to purchase the pack*".⁴⁸ However, in my opinion, the study is subject to a number of limitations.

4.51 The first major limitation of the study is the lack of an effective experimental design in the structure of the pack alternatives. Each screen displayed three packs of the same brand. As has been discussed both in this report and my 2010 Report, the design of choice set alternatives is a well-developed area of research and needs to be linked directly to the decision model one is investigating.⁴⁹ In this situation, a more appropriate design would have mixed the brands and the packs and potentially also varied other factors as well (e.g., price). There are several reasons for this. First, by showing individuals three packs of the same brand the researchers immediately signal to the respondent that the variable of interest is package type (indeed the questions push this bias home). Hence, the respondent will know to concentrate their attention on packaging. Second, without the introduction of a price the respondent has no real cost of making any decision

⁴⁷ See page 1, Gallopel-Morvan et al. (2011).

⁴⁸ See page 2, Gallopel-Morvan et al. (2011).

⁴⁹ See, e.g., Street, D. and L. Burgess (2007). *The Construction of Optimal Stated Choice Experiments*, Hoboken, NJ: Wiley.

(in this case every variable of interest is very heavily correlated – as revealed by Table 2 – which is a signal that there may be a method bias issue). Third, the development of effective choice models requires orthogonality in terms of the design. Without this one cannot make general statements about the influence of specific independent characteristics of the choices being made and hence the model is necessarily limited.

4.52 The second major issue is in the econometric modelling. The use of a logistic regression model is appropriate but not a binary logistic model as formulated. What is done in this study is equivalent to a more technical “one issue at a time comparison” that I have discussed in paragraph 4.26. Here the problems are two-fold:

- (a) First, the correct logit model is what is known as a conditional logit, which is not what is estimated in this study. The appropriate methodology is not to estimate what is given in Tables 3 or 4. The correct model requires that the researcher conditions the choice of “pack type X of brand Y” on the alternatives that were available to choose amongst (this is where the “conditionality” arises – it is the choice of pack of type X given it was brand Y and what the alternatives in the specific set were).⁵⁰ If one were to estimate this as a binary logit – assuming an orthogonal design – the correct model statistically would look something like:

Choice (Yes, No) =

$f(\text{pack type, brand, price} \dots | \text{Covariates, Survey Questions}) + \text{error}$ ⁵¹

⁵⁰ The reasoning and importance of conditional logit models in choice tasks can be seen in K. Train (2003), *Discrete Choice Methods with Simulation*, Cambridge UK: Cambridge University Press. Strictly speaking even a conditional logit is not really correct here. The best alternative is a mixed-logit model that accounts for sources of heterogeneity across individuals.

⁵¹ Note that this is a gross oversimplification of a more complex representation.

This analysis would then allow the researcher to determine the effect of brand, pack type, price and so on, conditional on the characteristics of the individual including their responses to the various survey questions. Unlike Tables 3 and 4 this would be a single equation that would allow for simple interpretation.

- (b) Second, even if one were not to use this formulation, another alternative would be to run a multinomial logit where the dependent variables were the pack type choices. This would provide estimates of multiple equations for each dependent variable (e.g., pack type or brand). It would be more appropriate if the researcher theoretically believed that different, but related, decision models applied to different dependent variables. It would also make sense if the researcher wanted to examine multiple choices (e.g., allowing individuals to choose more than one item in a choice set).

4.53 A final methodological issue is related to the six dependent variable measures. As with many of the studies discussed in this report and my 2010 Report, the measures used here ask individuals for opinions that draw on knowledge they lack and hence create opportunities for pseudo opinions to be created. This is likely to occur as the participants attempt to please the researcher by generating a response.

4.54 In this case the items asked are: (1) most effective in getting attention; (2) most “attractive”; (3) most effective in convincing non-smokers not to start; (4) most effective in motivating smokers to quit; (5) most effective in motivating smokers to reduce consumption; and (6) most effective for motivating youth to purchase the pack. If items (1) and (2) query the individual as to “what is the most effective in getting YOUR attention” and “most “attractive” to YOU”, this is reasonable, as they would know what THEY may find “attractive”. However, items (3) to (6) are querying individuals about factors on which they may have an opinion, but that

opinion could very easily be erroneous, biased and/or entirely random. Indeed, the questions asked of respondents in this study assume, without any basis for doing so, that packaging is a singular dominant motivating factor in smoking behaviour. For example, suppose that I asked individuals a different question: “What do you think would motivate youth to purchase less, one of the three packs presented or a 50% change in the price of the product?” No doubt the answers would change and packaging would become less relevant. However, even this is a slightly meaningless question because the individual knows absolutely nothing about the price sensitivity of the demand for the product. This would be evidenced by my changing that question to say “What do you think would motivate youth to purchase less, one of the three packs presented or a 10% change in the price of the product?” This is now more difficult because maybe 10% is not enough. However, the individual has no way of really knowing since it is asking them to answer a question on price sensitivity for which they are unlikely to have knowledge.

4.55 Asking individuals questions for which they have no effective ability to answer leads to response biases. Not knowing an answer but being queried to answer by an investigator will create unwarranted inferences (e.g., higher price implies better quality, vibrant colours are better than black and white) and pseudo opinions. In this study, as in most of the studies I have reviewed, this issue arises continuously as a series of opinions are being queried at the same time, increasing the likelihood that individuals will subconsciously succumb to desire to reveal consistency in their responses.

Conclusion

4.56 This study concludes with the view that limited edition packs have a significant influence on choice. However, given the limitations of the experimental design and the empirical modelling it is difficult to assess the veracity of this statement. Without better design, analysis and the inclusion

of all of the factors relating to product choice, it is impossible to make any statement as to the impact of the effects examined by this study.

Borland et al. (2011)

4.57 This study used an online survey of comparisons amongst five different pack types. 160 adult smokers aged 18-29 years participated in the survey. Individuals were shown static images followed by a short video on use of the packs. Individuals rated the “attractiveness of the package”, the quality of the cigarette, the extent of distraction from the health warning, and a “most and least” preference overall. The study concludes that: *“Pack shape and pack opening affect ever smokers’ perceptions of the packs and the cigarettes they contain. This means that they have the potential to create appeal and differentiate products and thus should be regulated”*.⁵²

Analysis

4.58 The design employed in this study is not dissimilar to those that have been discussed elsewhere in this report and my 2010 Report. Hence, one can raise many of the same criticisms related to self-report perception measures that are not validated by any specific behavioural outcome or independent measurement (such as observation or usage). While the researchers draw very strong conclusions as to the impact of the packaging on behaviour, there is no evidence that the results of the study translate into behaviour in any specific way.

4.59 Although the authors argue that they are using an “experimental” design, it is difficult to understand specifically what that design entailed. A simple 5 x 2 x 2 design would imply that there are 20 alternatives that can be created but the paper does not outline how these alternatives are presented to the participants in the survey. It is indicated in the study that each alternative

⁵² See page 1, Borland et al. (2011).

is ranked from 1 (least) to 5 (most) implying that individuals saw five alternatives – however, exactly how the other factors were ‘interacted’ in this set of five alternatives is not clear. This has significant implications for their analysis.

4.60 First, if all that participants saw were five pack types in four alternative settings (two brands, two pack warning sizes) then what the authors have effectively done is limited the interaction effects and made the overall design less efficient. A better design would have required that individuals see mixtures of warnings x pack sizes x brand. This would have allowed for the true testing of direct effects without the nesting of the effects within brand and warning size.

4.61 Second, the analysis is stated to be “repeated measures” but there are two issues with this.

- (a) First, the measures are not repeated in the sense that the individual sees the same item more than once. If my first point is correct, then each individual sees each brand at most four times, each warning at most four times, and each pack type at most four times, all nested within blocks defined by the brand. However, none of the specific items appear to have been seen more than once. For example, if the individual sees Benson & Hedges in a flip box with a 70% warning, they never see this mixture again. This problem would have been resolved by using a balanced incomplete block design (“BIBD”) where blocks of pack were presented out of the set of 20 possible pack alternatives, each of which may be ‘repeated’. This would entail a more complex cross-over design that would allow for greater efficiency and more control for interaction effects.
- (b) Second, the logic of even using a repeated design approach is not clear. Normally one uses such an approach when there are sample

size issues, the design is overly complex and hence burdensome to subjects, and when one wants to examine longitudinal effects. However, none of this is the case here. The application of BIBDs easily allows for individuals to examine large numbers of alternatives (so having a more complex structure with a better design would have not added much to the strain on the subjects), having a larger sample than 80 would not be necessary if the design was better, and there is no attempt at measurement over time.

4.62 Additionally, there is a potential serious measurement response bias with their “repeated measures”. In essence by “repeated measures” the authors have four data points per individual. However, what the subjects are seeing is a pattern (i.e., they see four block that always contain the five package variants). This potentially creates an artificial effect where the presentation of five items identically motivates the respondent to infer that “this is an experiment about package design” and hence focus artificially on the pack design aspects of the study. This may be what the researchers intended. However, good practice, particularly in a situation where policy recommendations are being made, would imply that one should attempt to remove such effects as they will lead to artificial false positives due to instrument salience effects as subjects look for points of differentiation.

4.63 The approach to scale creation is incorrect and potentially masks variance. As noted on page 2 of the study:

Overall means for pack preferences were calculated where ratings were only of most and least by scoring 5 points for each most preferred, 1 point for least preferred and scoring all other cases 3 points.

4.64 Statistically this reduces individual subject variance and potentially hides key effects. There is no logic for making this transformation, either

statistically or theoretically. The impact will inflate significance and allow extreme points to take on more importance but to do 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).

4.65 In addition and most critically, the ultimate conclusions drawn by the authors are simply not justified by the data presented and reveal a biased view of their own results. Let us assume that the data presented is valid. The conclusions that one would draw from Table 2 are that:

- the “standard” flip-top pack is not the most preferred (it rates behind “bevelled” and “rounded” yet above “2 x 10” and 4 x 5”) both on “attractiveness” and mean preference.
- the “standard” flip-top pack does not lead to a perception of the highest quality cigarettes (it rates behind “bevelled” and “rounded” yet above “2 x 10” and 4 x 5”) both on “attractiveness” and mean preference.
- the “standard” flip-top pack leads to the least distraction from health warnings.

4.66 The conclusions that one would draw from Table 3 are that:

- the “standard” flip-top box is the most “preferred” but the least “attractive” (effectively tied with “case opening”).
- the “standard” flip-top box leads to a perception of the lowest quality cigarette.
- the “standard” flip-top box is the least distracting alternative packaging design.

4.67 Yet despite the obviousness of this information the authors conclude:

While on the surface one might conclude that even though the standard flip-top opening was the least likely to distract from the health warnings, it was also rated as the most desirable and thus should be considered for removal; we do not believe that we have made a case for this as it is the normative opening and that factor may have influenced the ratings, rather than any inherent quality of the opening. This suggests that although there could be an initial benefit of adopting a different opening to reduce product desirability, we think such effects would likely be temporary and thus hard to justify given the cost and complexity of forcing virtually all packs to be redesigned and the negative short-term, at least, effects on warning prominence. We think it is more important to standardise and do so with designs that are not overly designed to make the pack look better than it needs be. If tobacco companies were allowed to vary these elements of pack design, we would be surprised if they were not used, at least to differentiate some brands and thus form the basis of a marketing strategy.⁵³

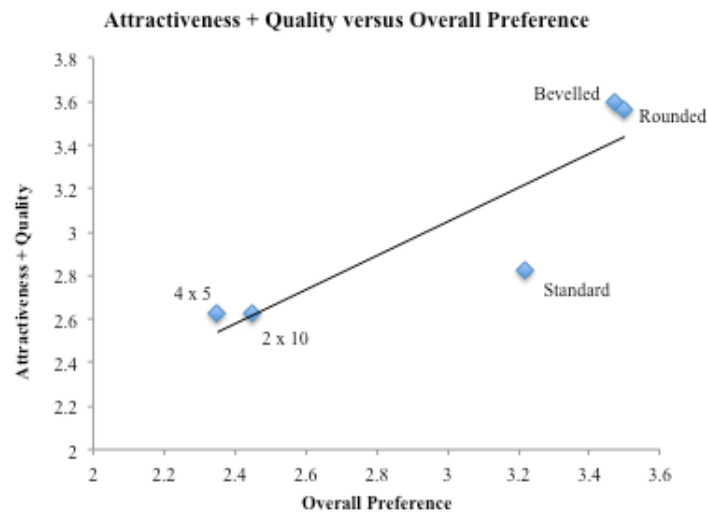
4.68 What this reveals is potential cherry-picking of results by the authors to support a conclusion that is not justified by a holistic examination of the results.

4.69 First, consider the statement “*it [the standard pack] was also rated as the most desirable*”.⁵⁴ This is indeed the case in Table 3, but not in Table 2. The standard pack is also not the most “attractive” as revealed by Tables 2 and 3. So what, specifically, are the authors arguing? Does “desirability” matter or is it “attractiveness” that really matters or the belief related to the quality of the cigarettes in the package (where again the standard pack does not do well)? Indeed, a conclusion of many studies I have reviewed in this

⁵³ See page 5, Borland et al. (2011).

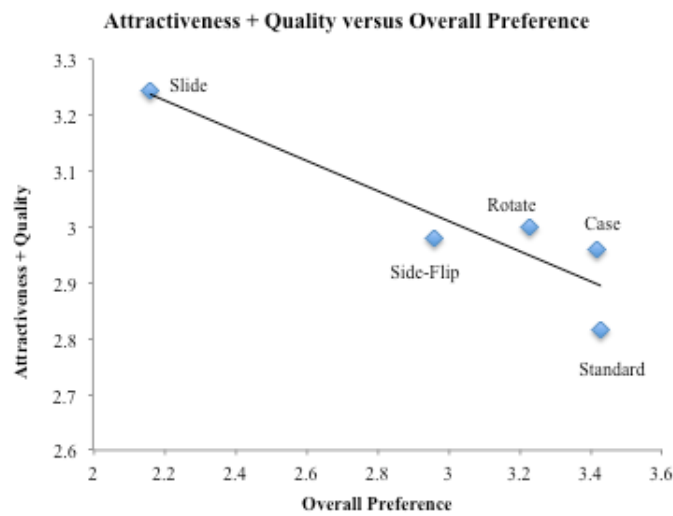
⁵⁴ Ibid.

report and my 2010 Report is that it is the perception of “cigarette quality” and “attractiveness” that drives preferences (we have seen this in the other studies examining what the authors call “false beliefs”). The Figure below shows just how unclear this statement is empirically. The x-axis is the measure of mean preference in Table 2. The y-axis is a simple average of the mean “attractiveness” and the perceived cigarette quality.



4.70 What one sees immediately is that the measures of pack preference do not align properly with respect to the standard pack. While the standard pack is low in “attractiveness” and “quality”, it is high in preference. This confused picture is made even worse when the information in Table 3 is examined in the same way as that from Table 2. The Figure below represents the data from Table 3 and shows that the relationship between “attractiveness”, “quality” and pack preference is actually negative. What this reveals is that the conclusions drawn will depend significantly on how one perceives the relationship between the specific measures. Indeed, if anything, the information in these two graphs indicates that there is potentially a significant cognitive familiarity effect. In other words, I prefer what I am familiar with. Hence, any conclusions relating to potential behaviour from constructs such as perceived “attractiveness” and “quality”

are going to be erroneous if one assumes that most behaviours are consistent across time.



4.71 Second, there are other instances where the authors go beyond the purview of their own study and draw conclusions that are not justified by the data. Consider the statement made by the authors that “*we think it is more important to standardise and do so with designs that are not overly designed to make the pack look better than it needs be*”.⁵⁵ It is unclear what leads to this conclusion, as the authors never attempted to examine the impact of “standardisation”. Indeed, one could speculate quite easily that the opposite of their conclusion is just as valid a hypothesis: i.e., that standardisation might enhance the salience of the brand as it is effectively the main piece of information that would be varying across the packs.

4.72 There is a hint in the results that there are also common method bias issues that may be arising in this study. It is noted on page 3 of the study that:

⁵⁵ Ibid.

Ratings of attractiveness and perceived quality of cigarettes for all pack shapes were moderately correlated at $p < 0.01$ (correlations ranged from $\rho = 0.32$ (2x10 pack) to $\rho = 0.54$ (4x5 pack)).

4.73 The reason this may signal a method bias issue is that there is no control for whether or not the individual has experience with the product – indeed, there are no controls for individual level variance and experience effects at all in the study. One explanation of these results is that the individuals are simply expressing opinions for which they have no basis. This is also seen in the fact that the most preferred pack is the pack that is in most use (as one would expect based on simple habitual bias). However, the study makes no attempt to control for method bias and experience related effects that may be significant sources of variance in the results.

4.74 A final concern that arises is whether one can draw any substantive conclusions based on the self-report information on health warnings (mean distracts). It is unclear why an individual would know what they are responding to when asked the extent to which a package would be distracting. Indeed, the best way to examine this would be through a recall task and not a simple Likert scale. The validity of their approach has to be called into question simply based on the fact that if one examines Tables 2 and 3 where there is a comparable measure for the standard flip-top pack we see that the two mean distracts measures (2.54 in Table 2 and 2.23 in Table 3) are significantly different ($t = 1.96$, $p = 0.025$). If this measure were valid, it should show no difference between whether it was collected as part of flip-top assessment or the assessment of the various other packs.

Conclusion

4.75 Overall, it is difficult to draw any conclusions from the findings of this study. It is also the case that the authors significantly overstate their conclusions. It is my opinion that, at best, all one can conclude from this

study is that based on some self-report measures of preference, individuals are skewed toward what they currently have experience with and may dislike case-opening boxes. Whether even this result matters materially to how consumers would react in a market environment cannot be concluded from the findings of this study.

4.76 Indeed, the conclusion that the researchers draw is seriously overstated.

“It is particularly notable that there were moderate to strong correlations between attractiveness and quality for each pack shape and opening, providing evidence that perception of the characteristics of the pack are generalised to characteristics of the cigarettes they contain. As we only showed respondents pictures of the packs and not of the cigarettes, we think it much more likely there is a causal link from pack to cigarettes than the converse or for some intermediary cause of the association. This finding also shows that the pattern of responding did not reflect random responding among the novel openings, rather variation in respondent preferences. Thus, while there is little preference beyond the current dominant opening, subgroups clearly have different preferences, and these could be used as the basis of targeted marketing.”⁵⁶

4.77 This is erroneous for three reasons. First, as noted by my analysis, the study suffers from serious methodological flaws that do not warrant such strong statements. Second, inferences of causation based on correlation are invalid in all precepts of science. The study was never designed to test a causal model and there is the potential that method bias and the use of self-reports may imply that the causal path is very different than from “pack to cigarettes”. The authors never posit a specific testable model but draw conclusions inductively from their data. Third, the conclusion that

⁵⁶ Ibid.

“*subgroups clearly have different preferences*”⁵⁷ is unsubstantiated as the researchers never sampled in a manner that would allow them to draw such conclusions nor did they apply any readily accepted methods that would be aimed at assessing individual or group level variation.⁵⁸ The authors make no attempt at all to control for individual level variation that may have influenced aspects of their results.

Moodie, Ford et al. (2011)

4.78 This study reports results from an on-line study of 658 young people (aged 10-17). In this study the respondents were queried about their beliefs and perceptions. The conclusion was that “*packaging appears to both attract young people and mislead them about product strength and relative harm. Innovative pack construction (novel pack shape and method of opening) and the use of color are instrumental in these effects*”.⁵⁹

Analysis

4.79 This study draws its conclusions based upon self-report survey responses of attitudes and perceptions in situations in which the individuals must draw inferences based on incomplete information.

4.80 As noted in my examination of other pieces of research in the area of plain packaging, it is impossible to determine specifically what the results of such research mean without some degree of external validation that links the perceptual data drawn from the surveys with more substantive behaviour response data (e.g., usage or purchasing), more structured experimental data (e.g., willingness-to-pay), or independently assessed pieces of information

⁵⁷ Ibid.

⁵⁸ For the most relevant and accepted methods to do this see, e.g., Rossi, P., Allenby, G. and R. McCulloch (2006), *Bayesian Statistics and Marketing*, Hoboken, NJ: Wiley and Wedel, M. and W. Kamakura (2001), *Market Segmentation: Conceptual and Methodological Foundations*, 2nd Edition, Dordrecht, NL: Kluwer.

⁵⁹ See page 1, Moodie, Ford et al. (2011).

(e.g., using test-retest methods). Without such validation, the respondents are likely to exhibit “halo effects” (where they assume relationships based on inference) or false beliefs (where they respond in a seemingly rational manner so as to complete the task). As I have discussed all of these issues earlier in this report and at length, I note only their relevance to the current study.

4.81 The dominant overarching issue with this study is the lack of any effective behaviour measures and the lack of a structure that takes into account the attributes critical to choice. Hence, the conclusions drawn from the study that pack design and colour matter are based on a study design which has failed to consider key drivers of choice or purchasing behaviour (such as price, availability and peer influences). As noted throughout this report and in my 2010 Report, to be able to effectively understand the role of packaging on choice it is important to have as complete a behavioural model as possible and to structure the empirical aspects of the study in a way that allow for the importance of all the key drivers to be considered in tandem. Focusing on isolated product attributes independently creates salience effects that lead to serious overstatement of importance as subjects respond. This is because the study focuses respondents on the issues that the researcher is studying, rather than the issues that are truly relevant to them. Failure to structure the studies correctly so that they are balanced and orthogonal across the key decision factors reduces researchers’ ability to estimate what the true impact of key decision drivers are.

4.82 A second methodological issue is related to the choice of dependent variables. First, the questions relating to colour are likely to engender pseudo opinions as they query the individual about experiences or factors that they may know nothing about – (a) the strongest tasting cigarettes; (b) the weakest tasting cigarettes; (c) the most harmful cigarettes; and (d) the least harmful cigarettes. Second, the questions on usage used to address plain packaging asked: (a) which pack they liked the most; and (b) which pack people their

age would be most likely to smoke. Question (b) is again a query that assumes that the individual possesses knowledge. In both of these sets of cases, inference bias will be an issue. Querying individuals on items for which they have no knowledge but a response is demanded by the investigator will lead to the tendency to use other cues in the study to rationalise their response.

4.83 For example, we can see this in the conclusion that: “*when shown four differently colored completely plain cigarette packs, displaying only health warnings, many young people associated pack color with product strength and also level of harm*”.⁶⁰ For this to be meaningful, one needs to ask, compared to what? The results reveal that about half the sample make an association. However, suppose that I changed this and gave them not colour but a picture that varied in complexity. Might this engender a similar response? If the effect is meaningful will it show up with alternative samples? Is there predictive validity from this? In other words, can I use this information to make superior predictions about what consumers will choose?⁶¹

4.84 The third methodological issue is the arbitrary aggregation of responses into binary (dichotomised) measures. As noted on page 3 of the study “*all responses were measured on 5-point scales ranging from 1 (very important) to 5 (not at all important). These were later recoded to binary variables to calculate the proportion who rated each as “very important” or “important” (Codes 1–2) and those who did not rate each as important*

⁶⁰ See page 6, Moodie, Ford et al. (2011).

⁶¹ One simple and well-established way to do this is to employ a holdout sample. With a holdout sample the researcher keeps a portion of the sample out of the analysis (e.g., 20 percent of the people studied). S/he then uses the results from their analysis to predict how that holdout sample would react. The difference between the predicted and actual results hints at how valid the model developed is from a predictive standpoint. For example, I might create a demand model for a product based on the product's price and colour plus the purchaser's gender, income, and age using 2,000 people from a total sample of 2,500. I would then predict how the remaining 500 people held out of the sample would purchase using the model created with the other 2,000.

(Codes 3–5).” There is no statistical logic behind this transformation without evidence of its impact.

4.85 This is then analysed 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) (see paragraphs 4.57 to 4.77 above). However, the main issue (and one repeated throughout) 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 also more affected by outliers. However, more importantly, this approach means that information which should be modelled is ignored.

4.86 Additionally, 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.⁶² Instead, there is a relationship between A and C and B and C and the research assumes that because A and B are correlated that some material relationship is there. Having a causal theoretical model and an empirical analogy that accounts for the inter-relationships reduces the sunspot problem. I noted in paragraph 4.28 above that one needs to apply structural modelling approaches to address this issue empirically. It is also the case that such a model cannot be designed without a theoretical structure that generates the hypothesised interactions.

4.87 A further limitation of a general nature to this study is its lack of correspondence with Borland, et al. (2011) (as discussed above). The table

⁶² 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.

below compares the results of the two studies in terms of package preferences:

	Moodie, Ford et al. (Percent Preferred)	Borland et al. (Mean Preference)
Flip Top	10%	3.43
Slide Pack	25%	3.53
None	50%	
Any	42%	
Others		2.74

4.88 This study concludes that slide packs matter very significantly while flip tops are disliked. However, Borland, et al. (2011) concludes that flip tops, being the standard in Australia, matter quite a lot and are very close or equally rated with respect to preferences. It is difficult to compare these two studies as the results above are subject to serious methodological issues in terms of how the data was collected and the biases associated with arbitrary aggregation. However, if one is to draw the sorts of generalised conclusions that are seen in this study, one has to question what that generalisation is based on. Do the packs matter? Or do they not? Or does it depend on who you ask and how you ask them? If it is the latter then the validity of the results for policy purposes has to be questioned.

4.89 Overall, one cannot take many conclusions from this study given its lack of interdependent behavioural validation. But even despite this there are conclusions one can point to that call into question its own conclusions. I will focus on two of these to illustrate the issues.

4.90 First, and as noted above, the authors conclude that: *“when shown four differently colored completely plain cigarette packs, displaying only health warnings, many young people associated pack color with product strength and also level of harm...This is consistent with the meanings that*

people, often automatically, attach to color...Although past research has typically explored pack color in the presence of branding (including brand name or pack descriptors), this study suggests that color alone is sufficient to mislead consumers”.⁶³ This conclusion is based on the results in Table 3 of the study.

4.91 The first issue with this conclusion is the statement “*this study suggests that color alone is sufficient to mislead consumers*”.⁶⁴ There is no evidence of this at all. Firstly, because the study examined each factor separately it cannot make any statements about whether something alone mattered. This can only be done by excluding all other effects, as would be the case in a structural equations model.

4.92 Second, the conclusions about the influence of colour are based on a limited part of the sample. Suppose that one reframed the data in Table 3 (as is done in the table below which shows the percentages making any or no colour association). I concentrate only on the total results and smokers since their definition of “susceptibility” is questionable.⁶⁵

	All	Ever-smokers	Never-Smokers
Any Association	48.25	60.25	43.75
No Association	51.75	39.75	56.25

4.93 What we see is that overall there is a 50:50 chance that a respondent makes any association at all with colour. Even in the case of ever-smokers, this number is not that materially different in that 40 percent make no

⁶³ See page 6, Moodie, Ford et al. (2011).

⁶⁴ Ibid.

⁶⁵ The authors’ definition is given as “*nonsusceptible never-smokers were those who indicated that they would “definitely not” smoke a cigarette during the next year, with susceptible never-smokers those whose response was anything other than definitely not.*” (see page 3 of the study). As noted with other recordings this is dubious statistically as it makes an arbitrary assumption that can be easily addressed by effect coding the responses.

association and the ever-smokers are 28 percent of the sample. What we do not know, of course, is whether that association is related in any way to anything material. For example, I can make an association because I am knowledgeable about a relationship (hence ever-smokers might use information about the products they have used). Alternatively, I may make an association because I am being asked to make an association and am more susceptible to assuming that such an association exists when I have no knowledge. It is even possible that both of these factors are at play in the different groups. Hence, the fact that an association exists does not de facto imply that it related to misperceptions of pack features.

4.94 Second, in the case of plain packs it is argued that “*one in three thought that young people would be most likely to smoke either the plain slide or the superslims pack. Of particular interest is the fact that susceptible never-smokers were almost three times more likely than nonsusceptible never-smokers to show a preference for the superslims pack and twice as likely to like the slide pack, which suggests that package design features such as shape and method of opening either appeals to, or perhaps reinforces, susceptibility to smoke*”.⁶⁶ There are two issues with this conclusion.

- (a) First, the statement that suggests pack design features such as shape and method of opening either appeals to, or perhaps reinforces, susceptibility to smoking is based on a specific definition of susceptibility that can be questioned, and makes an inference about appeal and reinforcement that is not tested in any way (as the conclusions are drawn by comparing across individuals). Without a behavioural model, no conclusion can be made on whether this is true or not.
- (b) Second, the preference for the slide-pack is conditional on individuals expressing an opinion at all and then the way in which the query is

⁶⁶ See page 7, Moodie, Ford et al. (2011).

framed. Based on the data in Table 5 of the study, 57 percent express no preference at all. When they do express an opinion the slide-pack is found to be more “attractive” (25 percent to 10 percent) but when the question is reframed to what they believe others would do, this result changes to 23 percent and 21 percent in favour of the flip-top box. This brings up the basic concern that the responses given are related entirely to the structure of the question design.

Conclusion

4.95 Overall, this study attempts to draw causal conclusions about behaviour without having a causal behavioural model. This creates a dilemma where policy conclusions are being inferred rather than proven in any meaningful way. As noted above, the data suffers from methodological and empirical limitations that make its conclusions difficult to justify. Put simply, this study does not provide reliable evidence to suggest that plain packaging will influence the public policy goals outlined in paragraph 1.5 above.

Al-Hamdani (2011)

4.96 This study is a quasi-replication of Wakefield et al. (2008).⁶⁷ Using a convenience sample of 220 adult smokers and non-smokers, subjects answered 10 perceptual questions about a signal pack of cigarettes, which was one of four designs. The goal stated by the author was to see whether or not plain packaging influences the recall of health warnings.

⁶⁷ Wakefield, M., Germain, D. and S. Durkin (2008), “How Does Increasingly Plain Cigarette Packaging Influence Adult Smokers Perceptions About Brand Image”, An Experimental Study, *Tobacco Control*, 17 (6): 416-421. I note this study has been reviewed by Dr Keegan at pages 21 to 24 of his 2009 Report.

Analysis

4.97 This study is a Masters thesis. Compared to certain studies published in journals, there is considerably more information about what was done, allowing for a more complete review of the underlying methodology and findings.

4.98 One of the first issues that arises in the study is researcher bias. The author has to be commended for revealing this. On page 16, the author notes that he is not employing focus groups “*because of his bias against the tobacco industry*”.⁶⁸ However, how and whether this influences other aspects of the study is not clear.

4.99 The study design also suffers from a number of flaws. First, the sample is essentially a convenience sample where individuals opted into the study and were paid to participate. Although paying subjects to participate is generally acceptable when the sampling is controlled, recruiting subjects using an advertisement and an opt-in with payment can lead to an unnecessarily large number of individuals simply doing the study for the reward. One easy check to control for this is to only use a specific percentage of individuals who meet certain criteria; such as how quickly they did the task (with online surveys this is easy since one piece of information given is the time taken to do the survey).

4.100 Second, the survey questions are structured to induce common method bias. Indeed, this is seen in the skewed responses and the author’s attempt to “transform” the data using various approaches (all of which are, in my opinion, unsuccessful). The problem is seen immediately on pages 67 and 68. Participants are asked about pack characteristics (“*how would you rate each of the following attributes of the displayed pack?*”). All of the questions would be good if rated high (10); i.e., more “attractive”, greater

⁶⁸ See page 27, Al-Hamdani (2011).

value, more exclusive and more tempting. Similarly for perceived individual characteristics (high is “trendy, young, masculine and confident”), taste (high is “rich, low in tar, satisfying and high quality”), and social standing (high is “outgoing and high class”). Hence, a participant wanting to complete the survey quickly would tick “high, high, high, etc.”; which is exactly what is revealed in the results. The author does not present the correlation matrix but relies on the VIF statistic⁶⁹ arguing that, “*none of the perception attributes were equal to or greater than 5 which meant that none of the variables were collinear with one another*”.⁷⁰ Unfortunately, this is not a correct conclusion to draw as it is possible that within each individual the responses are effectively identical and the skewed nature of the data hints that this is the case. In my opinion, this points to a clear common method issue. Hence, although the author sought to increase comparability by using the same materials as Wakefield et al. (2008) and Germain et al. (2009),⁷¹ he would have been better off structuring the task so that items could be reverse-coded or via some other scaling technique (such as a semantic differential).⁷²

⁶⁹ The VIF for a variable X_i compared to other variables – e.g., X_j , and X_k – is defined at $VIF_i = 1/[1-R_i^2]$, where the R_i^2 is the fit statistic from the ordinary least squares regression, $X_i = \alpha_i + \beta_{1i}X_j + \beta_{2i}X_k + \varepsilon_i$. VIF is useful to determine by how much the variance of the coefficient estimate is being inflated by multicollinearity. The choice of the cut off of $VIF > 5$ for hinting at multicollinearity is not definitive with different authors arguing for larger or smaller tolerances. Normally, one would use the tolerance from the VIF to adjust the model accordingly. However, in this case the author chose simply to assume that no multicollinearity existed and proceed accordingly.

⁷⁰ See page 48, Al-Hamdani (2011).

⁷¹ Wakefield, M., Germain, D. and S. Durkin (2008), “How Does Increasing Plain Cigarette Packaging Influence Adults Smoker’s Perceptions About Brand Image,” *Tobacco Control*, 17 (6): 416-421. Germain, D., Wakefield, M. and S. Durkin (2009), “Adolescents’ Perceptions of Cigarette Brand Image: Does Plain Packaging Make a Difference,” *Journal of Adolescent Health*, 46 (4): 385-392. I have previously discussed the latter study at paragraphs 4.32 to 4.37 of my 2010 Report.

⁷² A semantic differential differs from a standard rating scale in that the subject is presented with two defined extremes. For example, I could ask individuals their perception of a brand using extreme adjectives as the anchor points:

Trendy	–	–	–	–	–	Boring
Safe	–	–	–	–	–	Risky

4.101 A third issue is that the plain packs used were identical to those used in Wakefield et al. (2008). Again the author states that this was done to achieve comparability, but another potential issue arises in that all the packs were labelled with the same health warning: (a) “Smoking causes lung cancer” (given on page 27). This creates issues in the recall data. All the individuals who make an error in recall make that error by choosing one of three other warnings: (b) “Smoking causes impotence”; (c) “Smoking kills”; and (d) “Get help to stop smoking: Consult your doctor or pharmacist”. Health warning (b) is similar to health warning (a) in the use of the words “smoking causes”. Health warning (c) differs in that it is shorter by one word but starts with “smoking”. Health warning (d) is very different in that it is much longer at 10 words and starts with “get”. Good practice here would be to make all the health warnings the same length and to rotate the warning in case any one warning was more “recallable”. In this case, the author chose to use Wakefield et al.’s health warning and not address this issue. Unfortunately, one does not know which health warnings were mis-recalled as the author simply aggregated them into a correct and incorrect recall category.⁷³

4.102 A fourth issue is related to the coding of missing data. Page 34 of the study reveals that 23 to 32 values are missing for each perception question and the author replaced these missing values with the mean response. A more correct approach would also have been to create a variable for each measure that was coded [0,1] when a missing value was present for that variable. This would then allow the author to immediately determine whether those giving missing responses were different from others. There was no attempt to determine whether those not answering questions were different. By doing

Such scales make the extremes well defined but do not have known measurement properties. It is generally viewed as more cognitively complex and hence reduces common method issues when used appropriately.

⁷³ See page 32, Al-Hamdani (2011).

the recoding, the author is explicating that these people were “average” when there is no evidence that this is true at all.

4.103 Most concerning scientifically is that this analysis has absolutely no controls for individual level heterogeneity. The assumption is that other than the measures used in the analysis and given in the table everything else is unimportant. However this may not be the case. A more proper model would be to have covariates in the MANOVA that accounted for the individual level variability as fixed effects or to have a more complicated structure where they are sources of variability. Either way, the explicit assumption that the differences between the four package groups revealed in Table 4-1 are not important needs to be tested.

4.104 In my opinion, the most critical problem with the study is given in Table 4-4, which gives the health warning recall results. This analysis is simplistic and incorrect. A more appropriate model would have been to run conditional logit on whether the recall was correct or not. In essence, the dependent variable is whether the choice was correct or not [1,0] and then the independent variables are the experimental conditions (Original, PP1, PP2, PP3) conditional on the covariates (gender, smoking status, etc. as well as the perception measures). One could build a more complex structural model but the logic would be the same. In essence, the author is assuming that everything else is controlled for and that simply comparing A (correct) to B (correct) is sufficient. This is generally not a good assumption and good research practice requires that a researcher attempts to control for important as well as extraneous effects. There is no attempt to do so in this case and that is particularly critical in Table 4-4 because it is the heart of the entire thesis.

4.105 A final issue that makes one question the quality of the study is that the sample sizes make no sense. To illustrate this, in the Table below, I have

recreated the samples based on the information in the thesis given in Tables 4-1 through 4-3.

	Original	PP1	PP2	PP3
Table 4-1 (Gender)	47	51	43	51
Table 4-2	55	55	48	62
Table 4-3	57	58	50	66

4.106 What we see immediately is that the numbers do not match. The analysis in Table 4-2 is based on different pieces of information as that in Table 4-3 and the information on gender in Table 4-1. Good practice in consumer survey research is that you only use the data that is comparable across the entire analysis. In this case, the researcher is simply excluding or including cases where there is data for that specific analysis. If the key data is missing for any part of the analysis, it should be excluded throughout.

Conclusion

4.107 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.

Moodie & Ford (2011)

4.108 This study is a focus group examination of package design, including innovation, colour and plain packaging. Overall, the results of the study appear to be inconclusive, with some participants suggesting that plain packaging matters and others holding a neutral or indifferent opinion. If

anything, the results reveal indifference and a general opinion that one cannot draw much in terms of conclusions. As the authors note: “*When asked about the possible impact that plain packaging may have on young people, no consensus was reached in any group*”.⁷⁴

Analysis

4.109 A total of 54 adult smokers (aged 18-35) in Glasgow, Scotland, participated in 90-minute focus groups with 6 or 7 participants each.

4.110 This study is limited first and foremost by the use of focus groups. As noted in a number of places in this report and in my 2010 Report, focus groups are non-representative and subject to significant research and social bias effects that render their applicability to significant policy and commercial decisions severely limited. Most of the normal checks that one attributes to good research practice as outlined in Section 2 are violated. One cannot control for truthfulness or incentive compatibility nor can the focus group questioning be subject to statistical analysis that would determine the degree to which any statements or effects matter materially to specific policy outcomes. At best, focus groups can be used to work through nascent ideas that serve as a basis of more rigorous investigation. However, in this case, the value of the conclusions must be viewed sceptically as the conclusions of the authors outlined in section 1.1 have no scientific basis and amount to little more than a “considered opinion” of the researchers.

4.111 In this study, the focus group discussions were motivated by examinations of pack types – where the pack differed by opening, and also a ‘perfume’ pack style – colour (red, light blue, green and white), and plain pack colour. The only conclusion that appears to be drawn which is consistent in the study is that women reacted differently from men, with women being slightly more responsive to styling when discussing the

⁷⁴ See page 178, Moodie & Ford (2011).

alternative packs. Other than this conclusion, the remainder of the discussion was ad hoc inductive justifications of quotations and opinions of the various participants. Rather than go into more detailed criticism of the issues associated with focus groups research, I will simply note that the comments I will make shortly with respect to another study – McCool, Webb et al. (2012) (see paragraphs 4.119 to 4.125 below) – are applicable here. As the limitations are more demonstrable in that study, I leave it to that discussion to make specific points on potential alternatives that could have improved the work.

4.112 In the current study, what is most concerning is the fact that the results do not line up with other studies by the author. If we examine the conclusions here we find that: (a) pack type (other than the “perfume pack”) does not influence participants much; and (b) that most people do not believe that the plain packaging will alter their behaviour. Yet Moodie, Ford et al. (2011) and Borland et al. (2011) both conclude that packaging innovation matters significantly. Rather than attempting to rationalise why such differences arise, the simplest explanation is that the structure of the studies and the lack of any attempt at behavioural measurement render the results questionable from a policy perspective.

Conclusion

4.113 Overall, this study adds some confusing conclusions when compared to other work by some of the same authors. Independent of this, the relevance of a small-scale focus group study to major policy decisions must be questioned. The sample is non-representative and any conclusions drawn from it can only be considered as potential hypotheses for more rigorous scientific studies. Therefore, in my opinion, this study does not provide reliable evidence as to whether plain packaging would help achieve the public policy goals listed in paragraph 1.5 above.

Hoek et al. (2012)

4.114 This study is a focus group examination of plain packaging and young people's perceptions. 86 people participated in the focus groups with 42 being non-smokers and 44 smokers. The study concluded that: *“without brand symbols to confer these benefits, and with larger pictorial health warnings that actively challenge and usurp the cachet formerly associated with tobacco brands, plain cigarette packaging exposes tobacco products as unexciting sources of nicotine. Not only would the lack of branding fail to augment smokers' social personae, it would actively detract from the qualities they seek, identify them with undesired groups, and reduce the status they formerly obtained and still sought”*.⁷⁵

Analysis

4.115 As noted in my 2010 Report and again in this report, focus groups are non-representative and subject to significant research and social bias effects that render their applicability to significant policy and commercial decisions severely limited. Rather than repeat this criticism, I simply refer the reader to paragraph 4.110 above.

4.116 The major scientific issue with this study is the inductive nature of the method. The logic that individuals co-create value and personal image with product consumption is certainly well known and recognised in marketing literature. However, the means by which this would be studied in a specific context must go beyond focus groups if its specific importance is to be determined and its relationship to behaviour validated. In the case of this study, there are no alternative hypotheses to the suppositions of the authors and hence alternative explanations are never given any consideration. For example, the emphasis in this study is on the brand effects of “defining social status” and “identity demolition” when packaging is plain. But what of other

⁷⁵ See page 636, Hook et al. (2011).

factors such as price? For example, work in consumer and individual ethics shows quite clearly that “ethics” has a price.⁷⁶ Individuals may co-create an “image” by trying to be “eco” (i.e., environmentally conscious) but that image can be removed quite easily when it is more expensive to achieve. Indeed, it can be shown that ethical preferences can be completely removed when a specific reservation price level is hit. Hence, independent of what the authors can show by picking specific quotations from focus groups, the real issue is how important what they discuss is when applied to market circumstances where price, habit and competing demands on people’s time arise.

4.117 A second issue is one that arises with nearly all the Studies, namely the salience bias created when examining a singular topic. In this case, the emphasis was focused on identity associated with smoking. A simple way to remove this would be to concentrate on other social or consumption issues so that a comparison can be made that separates out a general set of issues relating to identity and those specific to the smoking context.⁷⁷ In addition, this approach can be structured to incorporate scenarios where price, availability, branding, peer influences and many other effects can be examined in a more quasi-experimental manner, allowing the researcher to obtain a fully and more scientifically valid picture that accounts for potential alternative explanations *ex ante*.

Conclusion

4.118 Overall, this study provides limited value to the debate on plain packaging. The general conclusion that co-created self image matters is not surprising. The failure by the authors to utilise the theoretical logic more directly and in a manner that is scientifically valid and empirically supported

⁷⁶ See, e.g., Devinney, T., Auger, P. and G. Eckhardt (2010), *The Myth of the Ethical Consumer*, Cambridge UK: Cambridge University Press; Zwick, R. and X.-P. Chen (1999), “What Price Fairness? A Bargaining Study,” *Management Science*, 45 (6): 804-823.

⁷⁷ I discuss this more with respect to McCool, Webb et al. (2012) in paragraph 4.123.

by direct behavioural measurement makes it no more than a considered opinion of the authors and potentially a basis for later research of a more positivistic and scientific bent. In addition, the lack of any representativeness in the sample, as well as the small sample size, renders it invalid for drawing policy conclusions. In my opinion, the authors' conclusions amount to considered speculation given that there is no evidence as to whether the effects they discuss: (a) do actually exist within the decision calculus of individuals in a way that would influence the policy goals set out in paragraph 1.5 of this report; and (b) materially matter to a degree that they would influence behavioural change.

McCool, Webb et al. (2012)

4.119 This study is a focus group examination of how adolescents respond to plain packaging. It concludes that *“the findings are highly suggestive that plain packaging may be an effective intervention to prevent smoking uptake. Given the dire consequences for health of smoking among young people, there is a strong justification for taking a precautionary approach and introducing plain packaging”*.⁷⁸

Analysis

4.120 A total of 80 Year 10 and 11 adolescents (14-16 year olds) in New Zealand participated in the study. This study is limited first and foremost by the use of focus groups. As noted in my 2010 Report and this report, focus groups are non-representative and subject to significant research and social bias effects that render their applicability to significant policy and commercial decisions severely limited. Most of the normal checks that one attributes to good research practice as outlined in Section 2 are violated. One cannot control for truthfulness or incentive compatibility nor can the focus group questioning be subject to statistical analysis that would determine the

⁷⁸ See page 5, McCool, Webb et al. (2012).

degree to which any statements or effects matter materially to specific policy outcomes. At best, focus groups can be used to work through nascent ideas that serve as a basis of more rigorous investigation. However, in this case, the value of the conclusions must be viewed sceptically as the conclusions of the authors outlined in section 1.1 have no scientific basis and amount to little more than a “considered opinion” of the researchers.

4.121 Secondly, the study is limited by its sampling. Non-representative small samples are a perennial problem with focus groups and this issue is exacerbated further in this case by the fact that there is no real attempt to ever determine whether the youth being examined here are in any way vulnerable or even sufficiently interested to take anything other than a cursory interest in the study. If the respondent has little interest in the topic and the intent of the study is obvious, the motivation to please the researcher is a strong source of bias.

4.122 Related to this point, it is my opinion that the most significant issue with this study, and one that occurs in other focus groups’ studies examining plain packaging, is the failure to hide the obvious intent of the research project from the respondents. In this case, all that the adolescents are shown are a series of cigarette packs and queried on the health warnings and the impact of plain packaging. This immediately signals to the respondent that the study is about such health warnings and there is every possibility that the respondents would react in a socially acceptable manner. This is demonstrated by the fact that nearly all of the discussion reported reveals negative perceptions of smokers, which is in many ways at odds with much other work on adolescents and smoking where the assumed viewpoint is that rather than being “sad” and “stupid”, smokers are perceived as “cool” and “trendy”.

4.123 This social response bias issue is further exacerbated by the plain packaging comparison. It is unlikely that the individuals seeing the plain

packs, independent of whether they were for cigarettes or for “Snickers” bars, would perceive that the lack of any colour or design would be seen as “unattractive”. The study purports to use a “grounded theory” method to “*produce a coherent explanation of the pervasive and interconnections between dominant themes*”.⁷⁹ However, the end result is mostly ex post facto justifications of quotations rather than any coherent theoretical development that explain the process by which the packaging aspects of the product lead to specific behavioural outcomes. Although the author(s) acknowledge many limitations to the focus group approach (for example, in discussing their use of focus group research, the authors state that “*the findings from our study should be carefully considered in the light of several limitations that need to be clarified*”),⁸⁰ this does not stop them from drawing extremely strong policy recommendations from this very type of research.

4.124 Independent of the serious researcher-induced bias issues inherent in this study, there are other methodological limitations that could have been addressed more directly.

- (a) First, rather than relying solely on tobacco, the study could have introduced other health or safety related product labelling and pack issues (e.g., related to obesity or alcohol usage) that were relevant to adolescents. This would have immediately reduced the salience bias associated with concentrating on tobacco alone and also helped to address the more general issue of how the respondents react to the natural directional bias of the focus group itself (e.g., less obesity is good, less alcohol usage is good, etc.).
- (b) Second, in addition to the plain packaging issue, other potential “interventions” could have been added that, again, served to remove

⁷⁹ See page 2, McCool, Webb et al. (2012).

⁸⁰ See page 4, McCool, Webb et al. (2012).

the potential salience of packaging alone (e.g., in Australia, fast food outlets must now present caloric information on the menu) and would have given the researcher a more nuanced picture of how plain packaging stands relative to other potential policy alternatives.

- (c) Finally, rather than relying on semi-structured interviews a more appropriate approach would have been to create alternative scenarios to which the respondents could react. This provides more control but also allows the researcher to more accurately examine the responses (e.g., with scenarios and time limitations it is easier to track word counts and hence compare the responses in a more scientific and systematic manner) while accounting for other motivating factors (e.g., peer group pressure).

Conclusion

4.125 Overall, it would be erroneous to draw policy conclusions based on this study for reasons noted in Table 2 of Section 2 of this report: *“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.”* This conclusion is all the more relevant here given that the authors argue that their study provides *“strong justification”* for a policy on the basis of little more than 80 unstructured interviews.⁸¹ I therefore do not consider this study to be reliable evidence to suggest that plain packaging would influence changes to actual smoking behaviour.

⁸¹ See page 5, McCool, Webb et al. (2012).

Wakefield, Germain et al. (2012)

4.126 This study is a between-subjects examination of the interaction between plain packaging and pictorial health warnings. 1,203 adult smokers (aged 18 years and over) in Australia participated in the study. The study concluded that plain packaging has a stronger effect on the assessment of the pack and the image of smokers than increases in the size of the health warnings. In addition, the study concludes that those in the “plain packaging condition” were less likely to express an intent to purchase than those in the “branded packaging condition”.

Analysis

4.127 Two main issues arise in this study. The first relates to the structure of the design of the study. The second relates to the nature of the authors’ analysis.

4.128 In examining the design of the study, there is a noted improvement from many other studies examined in this report and my 2010 Report in that participants evaluated only either branded or plain packaged alternatives in this study. This alleviates one potential limitation that has plagued related research in the area – the fact that individuals would not be facing a mixture of plain and branded packages in a plain packaging regulatory environment.

4.129 However, a number of points should be discussed. The most notable factor is that when one examines the results pertaining to “taste” and “harm” there appears to be no substantive effect of either health warning or taste. This is in stark contrast to a number of other studies by several of these authors purporting to show that plain packaging removes “false beliefs” that cigarettes are less harmful and that they remove associations that lead to beliefs about health and taste. Although the study suggests plain packaging has a slightly significant effect on “taste characteristics” reported in Table 2, this is mostly driven by sample size, as the actual differences between the

two conditions are not materially different (as seen in the Table below). A difference of 0.2 on a 10-point scale amounts to only 2% of the rating scale and given the variance of the responses it is less than 10% of a standard deviation.

	Packaging	
	Branded	Plain
Positive Taste	4.9	4.7
Negative Harm	7.7	7.6

4.130 Hence one must conclude that the results from this study appear to indicate that the packaging does not materially influence perceptions of taste or harm, even if the study shows an impact of perceived “attractiveness”.

4.131 The second issue is related to the nature of the analysis itself. Elsewhere in this report and my 2010 Report, I have discussed the limitations of the research for its “one at a time” approach to analysis. This applies here as well. Rather than the ANOVA approach utilised in this study, a more effective model would be to examine all of the responses simultaneously (most simply with a MANOVA, but more sophisticated approaches could also be applied). This is an issue here for two reasons.

- (a) First, all of the responses are taken from the subjects at the same time and there may be common variance effects due to questions asked about the related issues (e.g., pack characteristics and smoker characteristics). As the study does not report the correlations between these variables it is impossible to see the degree to which this is important.
- (b) Second, as with other studies by the same authors, there is a tendency to aggregate data when it is better to keep them separate. In this case, the authors report that they aggregate all the brand responses together.

This reduces variance and will inflate significance (hence leading to a tendency to report more significant results than justified). A more appropriate approach would have been to model the responses to each brand individually, which could be done quite easily in a MANOVA and would have allowed for a truer picture of individual brand effects. The importance of undertaking an approach can be seen in the fact that most studies reveal that consumers are habitual and hence will prefer and over-rate their traditional brand. However, in this study no such effect is seen and there is no attempt to control for the subject's favourite brand in the rating tasks.

4.132 The most limiting factor in this study is the lack of a behavioural model. Although this study used a brand choice task, it failed to evaluate that task in anything other than the most superficial of ways – reporting the mean differences in “no choice” between the different conditions. If the choice task and the ratings task had any meaning, the minimum that would be expected is that there is a relationship between the constructs reported in Table 2 and the ultimate dependent variable – i.e., whether or not to choose a specific brand conditional on package type. However, this is never done. As discussed elsewhere in this report, what ultimately matters is whether the measured responses are predictive of behaviour outcomes. While this study does not measure actual behaviour it gets closer by using a structured stated choice task – *“If you ran out of cigarettes and only the packs below were available in the store you went to, which pack would you be most tempted to buy?”*.⁸² This would have been very amenable to a structural model that used the ratings responses as the antecedents to the choice task. Without such a model, simply reporting that more subjects gave “no choice” as a response to the choice task does not imply anything about the impact of plain packaging as other antecedents may be driving the response. While the

⁸² One could criticise the use of the framing “most tempted to buy” as the word “tempted” is not neutral. More appropriate wording would have been “which, if any, would you be likely to purchase” as it is closer to the process of purchasing rather than giving into temptation.

design provides a degree of control for some effects, the study is still quasi-experimental and the final choices made have to be analysed conditionally.

4.133 A final issue which must be queried with this study is that a number of questions were asked but reported on or used in any analysis. For example, a series of attitudinal questions are discussed that never appear elsewhere in the paper – “*I regret having started smoking*”, “*The health effects of smoking are exaggerated*”, “*I get a lot of pleasure out of smoking*”, and “*Smoking is a disgusting habit*”.⁸³ Why these questions were never used and what their value was in the design of the study is unclear.

Conclusion

4.134 Overall, this study adds some confusing conclusions when compared to other work by some of the same authors. While plain packaging shows the same attitudinal impact given in the authors’ other work, there is no real impact on perceived “taste” or “harm”. Yet the authors’ other studies argue that branding and packaging create “false beliefs” about taste and harm that are critical to the plain packaging debate. Ultimately, however, the main limitation of this study is the authors conclusion that “*plain packaging offers unique advantages in reducing brand appeal and purchase intention*”⁸⁴ is not justified by any behavioural model that links the supposed antecedents of appeal to intention and behaviour. Therefore, I do not consider this study to be reliable evidence that plain packaging would change actual smoking behaviour.

CRUK (2012)

4.135 My review of this study focuses on Chapter 4, which states that “*we explore if, and how, young people respond to branded and plain tobacco*

⁸³ See page 3, Wakefield, Germain et al. (2012).

⁸⁴ See page 8, Wakefield, Germain et al. (2012).

*packaging, providing qualitative research*⁸⁵ and subsequently presents results from focus group findings.

Analysis

4.136 Overall 48 15-year olds participated in eight focus groups. Of the 48 participants, nine indicated that they were smokers. Hence, most of the participants (if their self-reported smoking behaviour is revealed truthfully) are non-smokers and generally reveal themselves to be unaware of tobacco packaging.⁸⁶

4.137 As noted with respect to many other studies examined, focus groups represent an unscientific approach to examining issues. This is related to: (i) the small sample nature; (ii) the inability to undertake any statistical analysis and to control for biases that are introduced by the process used in the research (i.e., verbal protocols); (iii) group think-effects that arise based on the open nature of the discussion; (iv) biases are introduced when individuals attempt to rationalise in real time about topics where they may have no information or knowledge; and (v) interpretation bias that arises in situations where the research is clearly on an emotive topic and no attempt is made to hide the intent of the research.

4.138 A second issue is that the topic's salience is revealed clearly to the participants. Hence, their responses are by definition influenced by the task. As noted elsewhere in this report, a simple solution to this is to use more than one product category and work to minimise the bias that would arise from concentration on a signal topic (e.g., alcohol, drugs and obesity).

4.139 Like the other Studies, the main issue of CRUK (2012) is its inability to determine specifically what its results are and how they relate to achieving

⁸⁵ See paragraph 4.1, CRUK (2012).

⁸⁶ See, for example, paragraph 4.3.2, CRUK (2012).

the policy outcomes listed in paragraph 1.5 of this report. For example, one of the conclusions of the study is that: “*there were mixed reactions to the perceived importance of tobacco packaging ... Price was highlighted as being more important*”.⁸⁷ Yet there is also considerable discussion about perceived impact of innovative packaging,⁸⁸ and other aspects of the pack. While these are areas where other studies discussed in this report and my 2010 Report suggest that there is relevance to youth smokers and youth uptake of smoking, nearly all the discussion in the focus group reveals either indifference or mixed opinions. In the end, whether these responses are more meaningful generally or whether they related to the specific sample examined is impossible to say.

Conclusion

4.140 In my opinion, the main limitation of this study is that, while the results show a confused and indifferent picture amongst a small sample of dominantly non-smoking youth, the authors draw stark conclusions as to the supposed ‘value’ of tobacco packaging. For example, the authors state that “*innovation and image appear most important to young people*”⁸⁹ but this is based entirely on the fact that some respondents say positive things about the design. For example, the fact that I may believe that a Vauxhall is less well designed than a Porsche can say nothing about whether I would purchase the Porsche when other factors come into play (particularly if I did not know how to drive and had no intention of driving). In addition, the vast majority of the conclusions are presented as definitive when the reality is that there is no statistical analysis determining whether the conclusions are valid. In simple terms, and in my opinion, the interpretation of the researcher is driven entirely by the context in which the focus group was examined.

⁸⁷ See paragraph 4.3.9.1, CRUK (2012).

⁸⁸ See paragraph 4.3.4, CRUK (2012).

⁸⁹ See paragraph 4.4, CRUK (2012).

Carter & Chapman (2006)

4.141 This paper presents a small sample (N=138) focus group discussion of various regulatory options relating to smoking. The purpose was to give individuals the opportunity to articulate their views. Although the study is not specifically related to plain packaging, plain packaging was one of the 11 options discussed.

Analysis

4.142 Like many of the studies examined, this is based on a focus group approach and hence all of the prior discussion in this report and my 2010 Report applies to this study. I will not repeat this discussion here simply for brevity.

4.143 This study is different in that it allows individuals to compare alternative regulatory options. This is a more realistic approach than employed in most of the Studies, but the study also creates considerable confusion in terms of how the results are presented. Most of the discussion of the results suggests the absence of any conclusions from participants. In my opinion, much of this is undoubtedly related to the fact that individuals are being asked to discuss complex policy options on which most would have no real idea as to the short-term and long-term costs and benefits or their ability to be implemented effectively.

4.144 What is most intriguing about this study is that the researchers utilise a “Discussion Stimulus” (given in Table 2 of the study), yet make no attempt to provide any information on this. A more salient set of results would have been to: (a) have individuals respond with a survey before the focus group; (b) engage in the focus group; and (c) redo the survey after the focus group. Independent of the quality of the survey, and what it asked, this approach may provide interesting information that triangulates the responses to the survey with the qualitative data being generated by the focus group.

4.145 In addition, the application of different scenarios would also allow the researchers to manipulate key aspects of the regulatory options. One thing that is clear about the results presented is the extent to which individuals react to the cost of the regulatory option and their belief about its effectiveness in achieving policy outcomes. In this sense, the respondents are quite “rational”. The use of scenarios rather than open discussions would make this clear and also allow the researchers to examine how “price/efficiency” sensitive people are to the policy options being discussed.

Conclusion

4.146 Ultimately one must conclude that there is little valid information in the results given in this study. To a degree, the study is not dissimilar to television roundtables where audiences discuss topical issues with a moderator. Although interesting, the ultimate value of the results in terms of informing policy is limited. Ultimately, the study’s non-representativeness and inability to structure alternatives that lead to realistic and justifiable predictions about behavioural outcomes, which could be validated, renders it of marginal value. I therefore do not consider this study to be reliable evidence that plain packaging would influence actual smoking behaviour.

GfK Bluemoon (2011)

4.147 This report consists of a series of seven studies on cigarettes and three studies on “other tobacco products” (Roll Your Own (“RYO”), Cigarillos and Cigars). The studies which formed part of the report were aimed at examining three specific goals, namely the “*objectives of plain packaging as announced by the Australian Government to (a) reduce the attractiveness and appeal of tobacco products to consumers, particularly young people; (b) reduce the ability of the tobacco product to mislead consumers about the harms of smoking; and (c) increase the noticeability and effectiveness of*

mandated health warnings".⁹⁰ As there are a large number of studies included in this report, I have chosen to focus on general issues with the studies and make reference (where necessary) to specific studies either as points of example or in cases where there are specific and pertinent concerns with that study alone.

Analysis

4.148 The first point to note is that that the studies concentrate on very specific operational goals as outlined in section 1.2 of the report, namely: *"the overall aim of the market research was to assess the potential plain packaging design elements to determine which plain packaging options were optimal to achieve the policy objectives. More specifically, the research sought to identify the optimal combination of: (i) candidate colours for plain packaging; (ii) font and font size for brand name; and (iii) graphic health warning design (size and layout). In summary, the research sought to identify one plain packaging design (colour, font type, font size) that would minimise appeal and attractiveness, whilst maximising perceived harm and the noticeability of graphic health warnings"*.⁹¹ As such, the studies do not directly address any of the behavioural policy issues outlined in paragraph 1.5 of this report.

4.149 Moreover, as the intent of the studies which formed part of the broader report was to focus on package colours, font size and health warning labels, the authors directly assume that no other factors mattered to consumers (or that these factors can be ignored). This shows up in two ways in the reports.

4.150 First, as with all of the studies discussed in my report here and my 2010 Report, there is no specified behavioural model of purchasing and usage

⁹⁰ See section 1.1, page 6, GfK Bluemoon (2011).

⁹¹ See section 1.2, page 6, GfK Bluemoon (2011).

behaviour. These studies, like all of the other Studies evaluated in this report and in my 2010 Report assume, rather than show, that pack “attractiveness” is meaningful to actual behaviours.

4.151 Second, while the researchers assume that “attractiveness” is meaningful, they do not have any structure for an articulated model that accounts for “how meaningful” it is relative to all of the other factors that would influence purchasing (e.g., pricing and availability, etc.). In other words, the researchers assume that the results they are achieving are valid, without considering the possibility that other factors may be important. However, this is not necessarily true and it is the researchers’ duty to show that it is a correct assumption. Furthermore, the failure to account for these other effects will artificially inflate the importance of investigated factors.

4.152 I will now discuss the general issues that feature across the studies in the report. The first issue which arises is that nearly all of the studies suffer from sampling difficulties that reduce their value to providing meaningful insight into the policy goals outlined in paragraph 1.5 above.

4.153 First, because the research agency chose not to include any adolescents in the sample (with the exception of Study 5 (Face-to-Face)), the ability to make inferences relating to youth uptake of smoking is lost. In this regard, the researchers state that:

“Whilst it was recognised that there would be value in including 14-15 year olds in the study, this age group was excluded from the research due to difficulties associated with conducting research with tobacco products among young adults. Previous experiences have shown that despite the objectives of the research, such exercises can be viewed by some market research companies who provide online research panels as ‘promoting’ or ‘marketing’ tobacco products to children. As a result, it was determined in conjunction with the Department and the Expert Advisory Group, that the

research not include the age group of 14-15 year olds. Similarly, 16-17 year olds were also excluded from all studies with the exception of Study 5 Face-to-Face due to likely difficulties in recruitment. In addition to this group being a hard-to-reach audience, under the AMSRS ethical guidelines, it is also a requirement that parental approval is obtained when surveying 16-17 year olds about sensitive topics such as those related to tobacco and smoking. This pre-approval process can result in refusals from parents to participate or issues with the honesty of responses (for example, respondents claiming 'not to smoke' even if they are smokers, if parents are present).”⁹²

4.154 I simply note that a number of prior studies examined in my report (Moodie, Ford et al. 2011; McCool, Webb et al. 2012; CRUK 2012) and in my 2010 Report (Hammond, et al. 2009; Germain, et al. 2009) used adolescents in their sampling without difficulties.

4.155 Second, none of the studies include non-smokers in their samples. If one of the stated public policy goals in introducing plain packaging is to mitigate the initiation of smoking, then it is important to understand the “non-user” and “potential-user” segments.

4.156 Further, none of the studies use effective probability sampling. In most cases, the researchers rely on soft quotas or online panel responses and in a number of cases report problems with the quotas; e.g., studies 2 and 6 and the RYO study. If the samples are going to be utilised as guides to population impacts, there are a number of issues that arise with respect to on-line panels, particularly in this case, as internet penetration and factors that are related to smoking behaviour may be correlated (e.g., an income effect or the susceptibility of aboriginal communities).⁹³

⁹² See section 4.2, page 23, GfK Bluemoon (2011).

⁹³ The issues related to the use of internet panels versus probability sampling can be seen in a study prepared for the AAPOR Executive Council by a Task Force operating under the

4.157 Third, nearly all the empirical studies in the GfK report use a “weighting scheme” that is inappropriate. This is of particular concern given the issue noted in paragraph 4.153 above.

4.158 In this regard, the researchers state that “*the quantitative data was weighted to population statistics on current smokers in Australia sourced from the ABS – National Health Survey: Summary of Results, 2007-2008 (Reissue)*”.⁹⁴ The attempt here was to somehow create a number that was meaningful for the population of Australia. However, this is invalid for two reasons. First, nearly all of the studies have relatively small samples (the largest sample was 455) and using a general population weighting on small samples (which implies that there are a small number of responses in each cell being weighted) can cause quite skewed results. Independent of this, the weighting needs to be relatively minor in magnitude, otherwise the weighting will be what drives the results. This is particularly an issue in this scenario as the researchers are applying weighting to samples as small as 41 and 193 subjects when the total population of smokers may be on the order of 1 million in Australia. Second, studies 1, 2, and 5 (Face-to-Face) limited themselves to capital cities (or their suburbs) while studies 4, 5 (online) and 6 utilised online panels. To illustrate this further, the table below outlines the sampling undertaken in the report.

Study	Sample Size	Sample	Study Type	Weighting?
1	122	Sydney/Melbourne/Adelaide	Focus Group + Survey	No
2	41	Capital Cities	Online Survey	Yes
3	10	Sydney	Focus Group	No

auspices of the AAPOR Standards Committee, with members including: Baker, R., Blumberg, S.J., Brick, J.M., Couper, M.P., Courtright, M. and Dennis, J.M., et al. (2010). Research Synthesis: AAPOR Report on Online Panels. *Public Opinion Quarterly*, 74(4), 711–781.

⁹⁴ See section 5.1, page 26, GfK Bluemoon (2011).

Study	Sample Size	Sample	Study Type	Weighting?
4	455	Australia	Online Survey	Yes
5 (FTF)	193	Capital Cities	Online Survey + Focus Group	Yes
5 (OL)	409	Australia	Online Survey	Yes
6	205	Australia	Online Survey	Yes
RYO	209	Urban NSW/Victoria/Queensland & Capital Cities for Quantitative survey	Online Survey + Focus Group	Yes
C'rillo	30	Urban NSW/Victoria/Queensland	Focus Group	No
Cigars	8	Urban NSW/Victoria/Queensland	Focus Group	No

4.159 Aside from the general issues noted above, there are also a number of methodological issues that arise in the specific studies that should be noted when assessing their relevance to the policy goals outlined in section 1.5.

4.160 Study 1 is a replication of another study (Gendall et al. (2011)) that I have discussed above in paragraphs 4.3 to 4.9. Although the sample size here is approximately twice that in Gendall et al. (2011), the major issues of pseudo opinions and a desire on the part of respondents to create rational associations imply that the results provide no information as to the value of plain packaging in meeting the policy goals outlined in section 1.5. The researchers conclude that: *“The research strongly suggests that Australian smokers do have differing perceptions of the common cigarette brands. The influence of brand design and packaging inform and affect associations strongly related to appeal, attractiveness and perceived harm”*.⁹⁵ However, there is no evidence as to how this influences behaviour.

4.161 A number of the studies (Study 2 and a slight variant of Study 5 (Face-to-Face), plus the RYO study) apply maximum difference (“Max-

⁹⁵ Ibid, page 83.

Diff’) scaling as an assessment tool. Although Max-Diff scaling has many important properties that allow a researcher to avoid aspects of bias (as is noted by the researchers), it is not clear what the value of the approach is here or what the results mean. The researchers note that:

*Max-Diff is appropriate when researching a larger number of test items. Consider a set in which a respondent evaluates four items, A, B, C and D. If the respondent says that A is best and D is worst, these two responses inform us of five of six possible implied paired comparisons: A>B, A>C, A>D, B>D, C>D where “>” means “is more important/preferred than”. The only paired comparison that we cannot infer is B vs. C. Maximum difference scaling questionnaires are relatively easy for most respondents to understand. Furthermore, humans are much better at judging items at extremes than in discriminating among items of middling importance or preference. And since the responses involve choices of items rather than expressing strength of preference, there is no opportunity for scale use bias. This is an extremely valuable property for cross-cultural research studies.*⁹⁶

4.162 However, there are a number of issues with this explanation. First, the authors note that “[t]his is an extremely valuable property for cross-cultural research studies” yet the study here is not cross-cultural.⁹⁷ Second, they note that “*Max-Diff is appropriate when researching a larger number of test items*”.⁹⁸ However, in this case, there are only 8 items to be evaluated (and in the RYO study only 6). What is curious is that Study 1 requires individuals to undertake a large number of adjective associations and it is here that something like Max-Diff could be applied very readily and have some real value (e.g., presenting individuals with mixtures of brand attributes

⁹⁶ See section 14.3, page 89, GfK Bluemoon (2011).

⁹⁷ Ibid.

⁹⁸ Ibid.

and asking whether the mixture “best” or “least” represents the brand). The problem with the application of Max-Diff scaling used here is that all that is being desired is a rank ordering (e.g., $A > B$, $B > C$, etc.).⁹⁹ Given what the exercise is meant to achieve, it is conceivable that either asking a 1-8 ranking is more relevant or three overlapping ranking tasks (e.g., rank items 1-4, items 2-6 and items 4-8). Either would give a full preference structure and help avoid the issue of the ordering of “middling items” in the choice set. It is also interesting to note that while Study 2 and the RYO study used Max-Diff scaling, the Cigarillo study used a ranking task to gather the same information.¹⁰⁰

4.163 Another issue that is not discussed by the researchers is that Max-Diff scaling is valuable in not only providing preference rankings but also information about the “solidity” of preferences. To illustrate this, if based on the task, it is determined that I am very clear in my preferences – for example, for me A is preferred to B, B to C and C to D ($A > B > C > D$) – this is very different if my preferences show examples of intransitivity – for example, I say that A is preferred to B and B to C but C is preferred to A ($A > B > C > A$). This can arise for three reasons. First, I am paying no attention to the task. Second, I have ill-defined preferences (meaning I do not know what I want). Third, items A, B and C are extremely close and I just made a mistake in ordering them. However, to get at this issue, one needs to have items appearing more often.

4.164 For example, in Study 2 subjects are presented with 6 sets of four different packs. Given that each item only appears 3 times and the individual is making only 6 decisions on “best” and “worst” it is possible that items will

⁹⁹ There is a direct mathematical relationship between the outcomes of a Max-Diff task and a ranking task such that one can be transformed into the other; Marley, A.A.J and J. Louviere (2005). Some Probabilistic Models of Best, Worst, and Best-Worst Choices, *Journal of Mathematical Psychology*, 49(6): 464–480.

¹⁰⁰ It should be noted, however, that the Max-Diff task used in the RYO study was actually identical to a ranking task since it had only three items being compared (once you say which is “best” and which is “worst” the third item had to be in between).

simply never be rated as best or worst in a set. To alleviate this, one should increase the number of times an item appears. However, this increases the number of comparisons. For example, a more efficient design than 6 blocks of 4 items is 14 blocks of 4 items. What this means is that each item now appears 7 times ($=14*4/8$). However, such a design, while providing vastly more information, may also overburden subjects.^{101 102}

4.165 One can see the value of this in a simple exercise. Every time an item is rated “best” it gets a +1. Every time it is rated “worst” it gets a -1. In the Study 2 design, this score can at best range from -3 to +3 for any item with the overall total score being -6 to +6. However, if the more efficient design described above is used, the scores can range from +7 to -7 and the overall range goes from -14 to +14. In the design employed in GfK Bluemoon (2011), there is a much higher probability that any item will be rated 0, meaning that it provides much less real information about the preferences.

4.166 In addition, the researchers report that they utilised Hierarchical Bayesian (“HB”) models.¹⁰³ The rationale for this was given as follow:

“HB makes it possible to estimate stable item scores from relatively sparse choice data. It does so by borrowing information across the entire sample to stabilise the scores for individuals. ... Individual respondents’ responses are analysed using HB techniques to derive attribute importance or preference scores at the individual respondent level. In this case, a single

¹⁰¹ It also appears that the design utilised in Study 2 is not correct. My own assessment is that a design for 8 items requires 14 blocks (achieving 86% efficiency). An extremely efficient design (98%) is achieved by having 8 blocks of 7 items (meaning one item is left out of each set).

¹⁰² In the RYO study the design and actual task is not made clear. My assessment is that an efficient design for 6 items would require 10 blocks of three items each with each item appearing 5 times. However, it is unlikely that this is what was done.

¹⁰³ See section 14.5, page 92, GfK Bluemoon (2011).

*score is calculated that indicated performance of a pack in terms of the key dimensions.”*¹⁰⁴

4.167 However, exactly which HB model is used is never specified and none of the HB results appear to be reported other than references to a composite score whose form is never specified (nor are the statistics behind it revealed). Although the authors state that “*individual respondents’ responses are analysed using HB techniques to derive attribute importance or preference scores*”,¹⁰⁵ it is hard to understand what those “attribute scores” are when the dependent variables are not choices but independent measures of “appeal” and “harm” and other product features (such as price) are excluded from the analysis.

4.168 While HB does indeed allow a researcher to work with sparse data, it is also important to recognise that using it as a crutch when a more complete design of the experiment exists is inappropriate. A more complete design would allow for more accurate and informative individual estimates. Put simply, with sparse data from each individual, HB uses the population data. With more information from each individual, that individual’s unique model parameters are more likely to be revealed.

Conclusion

4.169 Overall, the studies which form part of GfK Bluemoon (2011) are an attempt to gain a comprehensive picture of a number of aspects of the importance of plain packaging. However, the information provided by these studies is muddled and confused by the lack of behavioural measures, no clear behavioural model and an inconsistent, but eclectic, mixture of methodologies. The studies do not seem to be overly concerned with scientific validity and appear to be more driven by a desire to implement a

¹⁰⁴ Ibid.

¹⁰⁵ Ibid.

policy using the levers at hand; i.e., pack colour, health warning design and font size. Ultimately, GfK Bluemoon (2011) provides no clear evidence to suggest that plain packaging would have any impact on actual smoking behaviour.

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, I will focus on the Studies as a group and relate them back to the specific criteria.

5.2 Consistent with my 2010 Report, I will first address the validity of the studies using the same reporting format. I will then summarise my findings using an overall format.

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. In this regard:

- (a) It is my conclusion that all of the Studies examined – with the possible exception of Thrasher et al. (2011) – 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, with the exception of Thrasher et al. (2011), 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.

- (b) 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 – except for Thrasher et al. (2011) – 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). Even in the case of Thrasher et al. (2011), the artificiality of the experiment and the fact that it 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 bid prices seen.
- (c) Finally, it is my conclusion that, as all but one of the Studies 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 (even in the Thrasher et al. (2011) experiment the conditions

were artificial). 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 to an unjustified “leap of faith” about the importance of the various measures used in the Studies to relate them to actual behaviours.

5.4 The criteria outlined in paragraph 2.10 relate specifically to the experimental studies examined. In the case of Hammond, Doxey et al (2011), Borland et al. (2011), Wakefield, Germain et al. (2012) and Gallopel-Morvan, et al. (2011), it is my conclusion that they fail to meet even the most basic standards of good experimental design, 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 Report with respect to earlier studies by many of these authors remain valid.

5.5 In addition, Thrasher et al. (2011), while using a bidding game that added some behavioural validity, fails to meet the requirements of balance and orthogonality expected of such experiments, while excluding many other factors that might realistically influence choice and willingness to pay. Hence, even in this case, it is impossible to draw valid conclusions that would help determine the degree to which the policy goals outlined in paragraph 1.5 above would be influenced by plain packaging.

5.6 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.

5.7 The criteria outlined in paragraph 2.11 help us understand the degree to which stated intentions align better with actual behaviour.¹⁰⁶ Again, these criteria are relevant to the experimental studies discussed in paragraphs 5.4 to 5.6 above. As noted above at paragraph 2.11, stated intentions align better to actual behaviour when:

- (a) They are for existing products. This is because individuals generally understand the product category and products better. 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.
- (b) They are for durable rather than non-durable goods.¹⁰⁷ Individuals are more variable in how they purchase non-durables and use more deliberative decision models for durable products. Cigarettes are a non-durable good, subject to frequent purchasing. Hence, this might imply that stated intentions are less accurate for cigarette consumption.

¹⁰⁶ Morowitz, V., Steckel, J. and A. Gupta (2007) "When Do Purchase Intentions Predict Sales?" *International Journal of Forecasting*, 23: 347-364.

¹⁰⁷ 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.

- (c) They are for short-term horizon decisions rather than for long-term time horizon decisions. Individuals are generally poor at estimating what they will do at some vague future date as opposed to what they will do now or in the near future. 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.
- (d) Subjects are asked about purchase intentions for specific brands rather than for the product category in general. Like in paragraph 5.7(a), individuals understand the product and the purchasing context. The results of the various studies are mixed on this, as the choices asked do not focus on specific brands but on specific packages. This is important in this situation as cigarettes are a product for which, I understand, 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 are measured as ‘trial’ rates amongst existing purchasers in the relevant product segment, rather than being measured in terms of total market share. 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?”.

- (f) Purchase intentions are collected in a comparative mode, rather than monadically. 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 Borland et al. (2011), who had subjects rank the alternatives. Thrasher et al. (2011) had the bidding done on single packs only.

5.8 Overall, it is my conclusion that, as with the studies evaluated in my 2010 Report, **none of the Studies met enough of these criteria whereby one could argue that they would provide potentially valid predictions of actual purchasing behaviour.**

5.9 The criteria outlined in Table 2 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:

- (a) **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 Gendall et al. (2011), Hammond, Doxey et al. (2011), Gallopell Morvan et al. (2011), Borland et al. (2011), Moodie, Ford et al. (2011), Al-Hamdani (2011), Wakefield,

Germain et al. (2012) and GfK (2011) 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** (Moodie Hastings et al. (2011), Moodie & Ford (2011), Hoek et al. (2012), McCool, Webb et al. (2012), CRUK (2012) and GfK (2011)) suffer from bias that is induced by the inability of the researcher to hide the intent of the study – leading to potential respondent bias. Finally, in one case – Al-Hamdani (2011) – the researcher quite clearly outlines **researcher bias** by expressly indicating that they chose their method because of that bias. As discussed in Section 4, a number of the other studies reveal aspects of research bias in drawing strong or erroneous conclusions from mixed or weak results – e.g., Borland et al. (2011), Moodie, Ford et al. (2011) and McCool, Webb et al. (2012).

- (b) The **sampling frame** in the Studies varied. The sampling is representative in only one study (Gallopel-Morvan et al. (2011)) implying that none of the Studies can be used to make predictions about either the general population or the population of minors (both smokers and non-smokers). GfK Bluemoon (2011) attempted to engage in some aspect of representativeness by weighting their data but this was more a methodological flaw than a positive aspect of their sampling. This, in my expert opinion, renders their findings irrelevant to the policy questions outlined in paragraph 1.5 above.
- (c) The analysis in the Studies suffers from significant **statistical validity** issues. Most particularly, the experimental design limitations outlined in paragraphs 5.4 to 5.6 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.

5.10 To summarise, Table 3 below provides an overview of each of the studies examined in Section 4 and provides 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
Gendall et al. (2011)	Interview, Association Task	N=66, Smokers/Non-Smokers, Non-Representative	Generation of Pseudo Opinions, Response Bias, Small Sample	None
Moodie et al. (2011)	Survey & Experimental Manipulation	N=48, Smokers, Non-Representative	Poor Experimental Design, Lack of Balance, No Controls, Small Sample	None
Hammond, Doxey et al. (2011)	Survey, Experimental Manipulation	N=826, Females Only, Non-Representative	Poor Experimental Design, Lack of Orthogonality, Lack of Balance, Inappropriate Data Manipulation, Econometric Analysis Incorrect, "One-at-a-Time" Statistical Analysis	None
Munafo et al. (2011)	Eye Tracking	N=43, Smokers/Non-Smokers, Non-Representative	Pre-Test Study Only, Small Sample, Bias Induced by Recall Task, Missing Results	None
Moodie, Hastings et al. (2011)	Focus Group	N=34, Non-Representative	Focus Groups Do Not Allow for the Drawing of Valid Scientific Conclusions	None
Thrasher et al. (2011)	Bidding Experiment	N=404, Smokers, 4 States, Non-Representative	Inappropriate Design, Lack of Balance, Lack of Orthogonality, Lack of Realism in Bidding Task (Comparison between existing branded packs and plain packs)	Yes

Study	Method	Sample	Issues Noted in Section 4	Behavioural Measures
Gallopel-Morvan et al. (2011)	Survey, Quasi-Experiment	N=836, Smokers/Non-Smokers, Over 18 Years, Representative	Poor Experimental Design, Lack of Orthogonality, Lack of Balance, Inappropriate Data Manipulation, Econometric Analysis Incorrect, “One-at-a-Time” Statistical Analysis, Generation of Pseudo Opinions	None
Borland et al. (2011)	Survey	N=160, Smokers, Non-Representative	Poor Experimental Design, Lack of Orthogonality, Lack of Balance, Inappropriate Data Manipulation, “One-at-a-Time” Statistical Analysis	None
Moodie, Ford et al. (2011)	Survey	N=658, 10-17 Year Olds, Non-Representative	Inappropriate Data Manipulation, “One-at-a-Time” Statistical Analysis	None
Al-Hamdani (2011)	Survey	N=220, Smokers, Convenience Sample, Non-Representative	Researcher Bias, Sampling Bias, Common Method Bias, Poorly Designed Instruments, Inappropriate Data Manipulation	None
Moodie & Ford (2011)	Focus Group	N=54, Smokers, 18-35 Years Old, Non-Representative	Focus Groups Do Not Allow for the Drawing of Valid Scientific Conclusions	None
Hoek et al. (2012)	Focus Group	N=86, 18-25 Years Old, Non-Representative	Focus Groups Do Not Allow for the Drawing of Valid Scientific Conclusions	None
McCool, Webb et al. (2012)	Focus Group	N=80, 14-16 Years Old, Non-Representative	Focus Groups Do Not Allow for the Drawing of Valid Scientific Conclusions	None

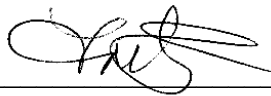
Study	Method	Sample	Issues Noted in Section 4	Behavioural Measures
Wakefield, Germain et al. (2012)	Survey	N=1,203, 18 Years and Older, Online Panel, Representativeness Unknown	Common Method Bias, Inappropriate Data Manipulation, “One-at-a-Time” Statistical Analysis	None
CRUK (2012)	Focus Group	N=48, 15 Year Olds, Non-Representative	Focus Groups Do Not Allow for the Drawing of Valid Scientific Conclusions	None
Carter & Chapman (2006)	Focus Group	N=138, 63 Smokers, 75 Non-Smokers (Three Age Groups; 18-24, 35-44, 55-64), Non-Representative	Focus Groups Do Not Allow for the Drawing of Valid Scientific Conclusions	None
GfK (2011) Study 1	Interview, Association Task	N=122, Smokers, Non-Representative	Generation of Pseudo Opinions, Response Bias, Small Sample	None
Gfk (2011) Study 2	Survey	N=41, Smokers, Non-Representative	Small Sample, Data Weighting, Experimental Design, Empirical Modelling	None
GfK (2011) Study 3	Focus Group	N=10, Smokers, Non-Representative	Small Sample, Focus Groups Do Not Allow for the Drawing of Valid Scientific Conclusions	None
GfK (2011) Study 4	Survey	N=455, Smokers, Non-Representative	Sampling, Data Weighting, Experimental Design	None
GfK (2011) Study 5 (Face-to-Face)	Focus Group + Survey	N=193, Smokers, Non-Representative	Sampling, Data Weighting, Experimental Design	None

Study	Method	Sample	Issues Noted in Section 4	Behavioural Measures
Gfk (2011) Study 5 (Survey)	Survey	N=409, Smokers, Non-Representative	Sampling, Data Weighting, Experimental Design, Empirical Modelling	None
Gfk (2011) Study 6	Survey	N=205, Smokers, Non-Representative	Sampling, Data Weighting, Experimental Design, Empirical Modelling	None
GfK (2011) RYO Study	Survey + Focus Group	N=209, Smokers, Non-Representative	Sampling, Data Weighting, Experimental Design, Empirical Modelling Focus Groups Do Not Allow for the Drawing of Valid Scientific Conclusions	None
GfK (2011) Cigarillo Study	Survey + Focus Group	N=30, Smokers, Non-Representative	Small Sample, Focus Groups Do Not Allow for the Drawing of Valid Scientific Conclusions	None
GfK (2011) Cigar Study	Focus Group	N=8	Small Sample, Focus Groups Do Not Allow for the Drawing of Valid Scientific Conclusions	None

5.11 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.

5.12 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 

Name: Professor Timothy M. Devinney

Date: 29 June 2012

6. EXHIBIT ONE – QUALIFICATIONS OF PROFESSOR DEVINNEY

Ex.1.1 I am a Professor of Strategy at the University of Technology, Sydney. In addition I am a Conjoint Professor in the Faculty of Medicine at the University of New South Wales and a Visiting Professor at the Institute of Management at Humboldt University – Berlin.

Ex.1.2 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.1.3 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.1.4 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.1.5 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.1.6 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.1.7 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* and the *Advances in International Management Series* published by Emerald. I am also the author or editor of more than 6 books.

Ex.1.8 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 Professor of Strategy, School of Business, University of Technology Sydney, July 2009–present.
- 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 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, Academy of International Business, 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. Best Paper Award Finalist (2005), 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 Nominee (Social Issues in Mgt Division)—Best Intl Paper (2007). 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). Conference Organizing Committee, Rio de Janeiro (2010).

Editorial and Refereeing Duties (Formal):

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

Co-Editor, *Advances in International Management*, Emerald (with T. Pederson and L. Tihanyi), 2009–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

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

Editorial Board, *Journal of Product Innovation Management*, 1991–present

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

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).
- 2013-2015 Australian Research Council (Discovery Program), Culture and the Individual: Advancing the Measurement of Culture (with D. Caprar, E. Maitland, P. Auger, C. Eckert, B. Kirkman and T. Peterson).

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

Ex.3.2 “Perspectives on the Art and Science of Management Scholarship,” *Academy of Management Perspectives*, 26, 1, February 2012 (with D. Seigel).

Ex.3.3 “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.3.4 “Customer Relationship Management and Firm Performance,” *Journal of Information Technology*, 26, 3, 2011 (with T. Coltman & D. Midgley).

Ex.3.5 “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.3.6 “Using Frontier Analysis to Evaluate Company Performance,” *British Journal of Management*, 21, 4, December 2010 (with G. Yip & G. Johnson).

Ex.3.7 “The Myth of the Ethical Consumer,” Cambridge: Cambridge University Press, 2010 (with P. Auger and G. Eckhardt).

Ex.3.8 “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.3.9 “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.3.10 “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.3.11 “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.3.12 “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.3.13 “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.3.14 “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.3.15 “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.3.16 “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.3.17 “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.3.18 “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.3.19 “The Other CSR,” *Stanford Social Innovation Review*, Fall 2006 (with P. Auger, G. Eckhardt & T. Birtchnell).

Ex.3.20 “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.3.21 “Modular Strategies: B2B Technology and Architectural Knowledge,” *California Management Review*, 47, 4, Summer 2005 (with P. Richard).

Ex.3.22 “Client and Agency Mental Models in Evaluating Advertising,” *International Journal of Advertising*, 24, 1, 2005 (with M. Collins & G. Dowling).

Ex.3.23 “A New Perspective on the Integration-Responsiveness Pressures Confronting Multinational Firms,” *Management International Review*, 44, SII, 2004 (with D. Midgley & S. Venaik). Reprinted in: J.A. Krug and J.D. Daniels, *Multinational Enterprise Theory*, Thousand Oaks, CA, 2007.

Ex.3.24 “What Will Consumers Pay for Social Product Features?” (with P. Auger, J. Louviere & P. Burke), *Journal of Business Ethics*, 42, 3, 2003.

Ex.3.25 “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.3.26 “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.3.27 “E-Business: Revolution, Evolution or Hype?” California Management Review, 44, 1, Fall 2001 (with D. Midgley, T. Coltman & A. Latekefu).

Ex.3.28 “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.3.29 “Understanding Institutional Designs Within Marketing Value Systems,” Journal of Marketing, 63, Special Issue, 1999 (with S. Carson, G. John & G. Dowling).

Ex.3.30 “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.3.31 “A Formal Model of Trust Based on Outcomes,” Academy of Management Review, 23, 3, July 1998 (with R. Bhattacharya & M. Pillutla).

Ex.3.32 “The Essence of Corporate Strategy: Theory for Modern Decision Making, Sydney,” Allen & Unwin, 1997 (with Jeremy Davis).

Ex.3.33 “How Well Do Patents Measure New Product Activity?” Economic Letters, 41, April 1993.

Ex.3.34 “New Products and Financial Risk Changes,” *Journal of Product Innovation Management*, 9, September 1992.

Ex.3.35 “New Product Innovations and Stock Price Performance,” *Journal of Business Finance & Accounting*, 19, September 1992 (with P. Chaney).

Ex.3.36 “The Return to Advertising Expenditure,” *Marketing Letters*, 3, May 1992, (with B. Cooil).

Ex.3.37 “European Markets After 1992: Implications for Business Strategy,” Lexington, MA: Lexington Books, 1991, (with William C. Hightower).

Ex.3.38 “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.3.39 “New Products Over The Business Cycle,” *Journal of Product Innovation Management*, 7, December 1990.

Ex.3.40 “Diversification Strategy and Performance in Canadian Manufacturing Firms”, *Strategic Management Journal*, 11, September 1990 (with T. Nguyen and A. Seror).

Ex.3.41 “Rationally Determined Irrationality: An Extension of the Thesis of Rationality as Anti-Entropic,” *Journal of Economic Psychology*, 10, November 1989.

Ex.3.42 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 which 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) Bondy, S.J., Paglia, A. and M.J. Kaiserman (1996) Tobacco purchasing and marketing. In Stephens, T., Morin, M. (Eds). Health Canada. Youth Smoking Survey 1994: Technical Report. Ottawa, ON: Minister of Supply and Services Canada.

- (b) Freeman, B., Gartner, C., Hall, W., and S. Chapman. (2010). Forecasting future tobacco control policy: where to next? *Australian and New Zealand Journal of Public Health*, 34(5): 447-450.
- (c) Bansal-Travers, M., O'Connor, R. Fix, B. and K. Cummings (2011). What do cigarette pack colors communicate to smokers in the US? *American Journal of Preventative Medicine*, Vol. 40, No. 6, 683-689.
- (d) Yong, H-H., Borland, R., Cummings, K., Hammond, D., O'Connor, R., Hastings, G. and B. King (2011). Impact of the removal of misleading terms on cigarette pack on smoker's beliefs about 'light/mild' cigarettes: cross country comparisons. *Addiction*, Advance Publication, 1-10.
- (e) Mutti, S., Hammond, D., Borland, R., Cummings, M., O'Connor, R. and G. Fong (2011). Beyond light and mild: cigarette brand descriptors and perceptions of risk in the International Tobacco Control (ITC) Four Country Survey. *Addiction*. Advance Publication.
- (f) Ford, A., Moodie, C. and G. Hastings (2011). The role of packaging for consumer products: Understanding the move towards 'plain' tobacco packaging. *Addiction Research and Theory*. Early Online 1-9.
- (g) Currow, C. and A. Dossaix (2011). Plain packaging for tobacco products: Minimising the emotional attachment to a cigarette brand could help smokers quit. *BMJ* 2011; 343:d5693 doi.
- (h) Chapman, S. and B. Freeman (2011). From brand to bland – the demise of cigarette packaging. *BMJ* 2011; 343:d4376 doi.
- (i) Wakefield, M. (2011). Welcome to cardboard country: how plain packaging could change the subjective experience of smoking. *Tobacco Control* September 2011, Vol. 20 No. 5.

- (j) Brown, A., McNeill, A., Mons, U. and Guignard R. (2012). Do smokers in Europe think all cigarettes are equally harmful? *European Journal of Public Health*, Vol.22, Supplement 1, 2012, 35-40.

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) Swanson, M. (1997). Generic Packaging of Cigarettes. Thesis for Master of Public Health. University of Western Australia, Perth (unpublished).
- (b) Gallopel-Morvan, K., Moodie, C., and J. Rey (2010). Demarketing cigarettes through plain cigarette packaging (presentation at *Colloque de l'Association Française du Marketing*, Le Mans, 6-7 May 2010).
- (c) Hammond, D., and S. Daniel (2011). Plain Packaging and Smoking Susceptibility Among UK Youth (only available in abstract from Society of Nicotine and Tobacco Research, February 2011).
- (d) White, C., and D. Hammond (2011). The impact of cigarette pack design and plain packaging on female youth in Brazil: brand appeal & health-related perceptions (only available in abstract from Canadian Public Health Association 2011 Conference).
- (e) van Hal, G., Arts, M., Vriesacker, B., Fraeyman, J., and S. Roos (2011). Tobacco Plain Packaging: Perceptions of Flemish Teenagers. Unpublished manuscript.
- (f) Hammond, D., Thrasher, J., Fong, G., Wakefield, M., Bansal-Travers, M., Zanna, M., Sebrie, E., and M. Cummings (2011). International Packaging Study: United States. Unpublished summary of findings.

- (g) Hammond, D., and J. Thrasher (2011). International Packaging Study: Mexico. Unpublished summary of findings.