

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

30 November 2010

Table of Contents

1.	REPORT INTRODUCTION	2
2.	CRITERIA FOR EVALUATION	4
3.	SELECTION OF THE STUDIES	16
4.	REVIEW OF THE STUDIES	16
	CLASSIFICATION OF STUDIES	16
	DETAILED ANALYSIS OF THE STUDIES	17
	Hammond, et al. (2009)	17
	Hammond and Parkinson (2009)	24
	Moodie and Hastings (2009)	27
	Germain, et al. (2009)	31
	Doxey (2009).....	34
	Bansal-Travers, et al. (ND)	39
5.	CONCLUSION	41
6.	EXHIBIT ONE – QUALIFICATIONS OF PROFESSOR DEVINNEY	48
7.	EXHIBIT TWO – RESUME	50
	Examples of Research Support Received.....	53
	Professional Consulting Experience (Selection)	54
8.	EXHIBIT THREE – SELECTED PUBLICATIONS	55
9.	EXHIBIT FOUR – OTHER MATERIALS CONSIDERED.....	59
	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.....	59
	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	60

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. I have been involved most recently in an extensive set of research projects examining the degree to which social aspects of consumption influence behaviour.

1.2 Exhibits One to Three of this report set out in detail my professional qualifications, my current resume and a list of 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 specialized 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 the basis on which I have prepared this report below).

1.3 Prior to drafting this 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:¹

¹ Plain packaging has most recently been described by the European Commission in its 24 September 2010 Consultation on the Possible Revision of the Tobacco Products Directive 2001/37/EC as standardised tobacco packaging in which the “manufacturers would only be allowed to print brand and product names, the quantity of the product, health warnings and other mandatory information such as security markings. The package itself would be plain coloured (such as white, grey or plain cardboard)”..This consultation document is available at http://ec.europa.eu/health/tobacco/consultations/tobacco_cons_01_en.htm.

- (a) “Analysis of Consumer Survey Evidence Relevant to the UK Department of Health Consultation on the Future of Tobacco Control” dated 2 September 2008;²
- (b) “Analysis of Consumer Survey Evidence Relevant to the UK Department of Health Consultation on the Future of Tobacco Control – a supplemental report” dated 19 June 2009;³ and
- (c) “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 which considers, in an exhibit, “Tobacco Branding and Plain Packaging: The New Frontier of Tobacco Control?”, Hoek, Wilson and Louviere (2008) and “Effects of Dissuasive Packaging on Young People”, Hoek, Wong, Gendall, Louviere, Cong (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. Whilst this document is not relevant to the subject of plain packaging, I have read this document in terms of its relevance to the evaluation criteria adopted by Dr Keegan, which (as set out below) I also utilise in preparing this report.

In this report, I refer to these four documents collectively as “the Reports”.

1.4 This report addresses the extent to which publicly available consumer survey studies not considered in the Reports (which I refer to below as “the Studies”) 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.

² Available at http://www.jti.com/cr_home/industry_regulation.

³ Available at http://www.jti.com/cr_home/industry_regulation.

1.5 I note the following two points in this regard:

- (a) a number of surveys which are considered either in the Reports or this document seek to undertake research in respect of the extent to which respondents are “more likely to notice” mandatory health warnings on tobacco packs following the introduction of a plain packaging measure. Given that ‘noticing’ something does not necessarily translate into a change in respondent behaviour, in this report I comment on the extent to which on-pack health warnings connect the “visibility”, “prominence” or “salience” of health warnings to changes in the types of actual smoking behaviour identified above at paragraph 1.4(a) to (c); and
- (b) I am a colleague of Professor Jordan Louviere, one of the authors of the studies entitled “Plain Packaging: The New Frontier of Tobacco Control?”, Hoek, Wilson and Louviere (2008) and “Effects of dissuasive packaging on young people”, Hoek, Wong, Gendall, Louviere, Cong (2010). Given my personal and professional relationship with Professor Louviere who also teaches at UTS, I have not commented on these studies in this report. As noted above, the studies co-authored by Professor Louviere have been considered separately by Dr Keegan in his 2010 report referred to at paragraph 1.3 above.

2. CRITERIA FOR EVALUATION

2.1 Given that the Reports by Dr Keegan have gone into a detailed description of many of the evaluative criteria that I will also use, I simply note here my concurrence with the discussion of the content and relevance of these criteria to my evaluation. I will, however, as explained below, utilise several additional criteria that I believe are relevant to the evaluation of the Studies. A list of the other materials examined for the purposes of preparing this report is given in Exhibit Four.

2.2 In evaluating the consumer research studies a number of factors come into play in determining the validity of the study with respect to addressing the questions being set. For simplicity, I outline a summary of the criteria discussed in detail by Dr Keegan in the

Reports in Table 2 on page 12. This table also provides a brief definitional description of the criteria.

2.3 I would also note that in line with 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),⁴ the American Association for Public Opinion Research (AAPOR),⁵ the World Association for Public Opinion Research (WAPOR),⁶ the Council of American Research Organizations (CASRO),⁷ the International Statistical Institute (ISI),⁸ the Canadian Market Research and Intelligence Association (MRIA),⁹ and the European Society for Opinion and Marketing Research (ESOMAR).¹⁰

2.4 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. 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 cognizant of the

⁴ Market Research Society (2005), *MRS Code of Conduct*. <http://www.mrs.org.uk/standards/downloads/code2005.pdf>.

⁵ American Association for Public Opinion Research (2005), *AAPOR Code of Professional Ethics & Practices*. <http://www.aapor.org/aaporcodeofethics>.

⁶ World Association for Public Opinion Research, *WAPOR Code of Professional Ethics and Practices*, <http://www.unl.edu/WAPOR/ethics.html>.

⁷ Council of American Research Organizations (2008), *Code of Ethics and Standards for Survey Research*,. <http://www.casro.org/pdfs/CodeVertical-FINAL.pdf>.

⁸ International Statistical Institute (1985), *Declaration on Professional Ethics*. <http://isi.cbs.nl/ethics.htm>.

⁹ Market Research and Intelligence Association (2007), *Code of Conduct for Members*. <http://www.mria-arim.ca/STANDARDS/CODE2007.asp>.

¹⁰ 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.

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 cognizant of the degree to which the study itself enhances artificially the salience of the factors being studied.
- (c) Ultimately, the goal of policy related research is to examine the efficacy of a change in policy on a change in behaviour. Changes in behaviour themselves require the effected 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.

2.5 Points (a) and (b) in paragraph 2.4 relate specifically to what is known as the attitude-behaviour gap or the difference between 'stated' intentions and 'revealed' or actual purchases.¹¹ Good research will attempt to reduce this problem by focusing on three factors, which we will add to our list of criteria:¹²

- (a) **Incentive compatibility.** Incentive compatibility addresses the extent to which the methodology used by the researcher allows (or makes) subjects 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

¹¹ 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.

¹² 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.

reveal that behaviour, such as in the case of a new product). For example, it is well understood that forcing individuals to make a trade-off or asking them to pay a real price creates incentives that align better with their actual purchasing behaviour. Incentive compatibility is a particularly critical issue when asking:

- (a) 'intention' questions – e.g. "if faced with these alternatives, which would you choose?";
 - (b) 'speculative' questions – e.g. "how do you think a person faced with these alternatives would behave?"; and
 - (c) most types of 'self-report' questions – e.g. "how likely is it that you do/would engage in a behaviour".
- (b) **Inference of salience.** Inference of salience addresses the degree to which the sheer addition of a factor that would otherwise not be part of the consumer's decision is all of a sudden added into the mix. For example, asking individuals about newly added attributes to existing products – i.e., aspects of a product that consumers know do not currently exist in what is offered in the market – heightens the 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.

2.6 Point (a) in paragraph 2.4 forces us to ask very specifically how an individual chooses to do (or not do) something. Given the questions outlined in paragraph 1.4 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?

2.7 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)?

2.8 The point of this criterion is the linkage 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?

2.9 The experimental studies in this report examined 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.

2.10 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, which is effectively what we see in the Studies examined in this report, requires that individuals: (a) choose amongst a set of alternatives, (b) rank a set of alternatives, and/or (c) rate a set of alternatives.

2.11 The set of alternatives – know 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.¹³ 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 that when looking at the choices that individuals are being asked to consider, that the attributes and features presented exhibit (as closely as possible) **orthogonality**.¹⁴ In product choice experiments the features of the product are broken into ‘attributes’ – such as brand, colour, or price – and the attributes into ‘levels’. For example, the ‘levels’ of the attribute *price* might be €1, €3, €5, €7 and the attributes of *colour* might be red, green and blue. Orthogonality implies that the experiment examining *price* x *colour* be set up with $4 \times 3 = 12$ alternatives as shown in Table 1 and that the individual be presented with a series

¹³ 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.

¹⁴ Perfect orthogonality is sometimes difficult to achieve and there are classes of non-orthogonal experimental designs. However, these are normally reserved for quite complex experiments and none of the Studies examined here would, in my expert opinion, be classified as being so complex as to require anything other than a basic experimental approach. See, e.g., Kuhfeld, W., Tobias, R. and M. Garrett (1994), “Efficient Experimental Design with Marketing Research Applications,” *J. of Marketing Research*, 31: 545-557.

of price x colour alternatives – as represented by the cells A-L below – such that the effects of each combination can be evaluated independently.¹⁵ 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.¹⁶ 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.

Table 1: An Example of a 4x3 Experiment

		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 that when looking at the choices that individuals are being asked to consider that the experiment exhibit level **balance**.¹⁷ This means that the

¹⁵ More specifically this is important when using linear regression models for estimation purposes. As noted by Kuhfeld, Tobias and 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.

¹⁶ 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.

¹⁷ 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

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.

2.12 Point (a) in paragraph 2.4 relates also to intentionality. Morwitz, Steckel and Gupta¹⁸ 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 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

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).

¹⁸ Morowitz, V., Steckel, J. and A. Gupta (2007) “When Do Purchase Intentions Predict Sales?” *International Journal of Forecasting*, 23: 347-364.

- (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).

2.13 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.

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 2.3.
<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

Evaluative Criteria	Short Description of Criteria and Its Relevance
	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 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
	Recall of non-recent past behaviour	Self-Reported Behavioural
	Prediction of future behaviour	Opinion / Attitudinal
Least reliable	Prediction of others’ future behaviour	Opinion / Attitudinal

Evaluative Criteria**Short Description of Criteria and Its Relevance**

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 generalization.

Age of respondents

Conducting research among minors presents particular issues that must be accounted for to ensure the reliability of the data collected.¹⁹ 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 reliable data.

Focus groups

Focus group studies are exploratory. They generate hypotheses rather than findings that can be generalized to a wider population.²⁰

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

¹⁹ Churchill, G and D. Iacobucci (2005), *Marketing Research: Methodological Foundations, 9th Edition*. Orlando, FL: Dryden, 387-390.

²⁰ Churchill, G and D. Iacobucci (2005), *Marketing Research: Methodological Foundations, 9th Edition*. Orlando, FL: Dryden, 81-85.

Evaluative Criteria	Short Description of Criteria and Its Relevance
5. Analysis	
Statistical significance	<p>It is imperative that authors refrain from projecting results that are not statistically significant to general populations or markets.²¹</p> <p>It is widely recognized 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 or error or other factors.²²</p>
Unsupported results or conclusions	<p>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.²³</p>
False comparison	<p>It is imperative that authors refrain from generating comparisons and drawing conclusions from comparisons that are not reflective of actual real life conditions.</p>
Cross cultural applicability	<p>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.</p> <p>This criterion is not applied to the studies in this report.</p>

²¹ International Statistical Institute (1985), *Declaration on Professional Ethics*. <http://isi.cbs.nl/ethics.htm>, paragraphs 1.3, 3.1.

²² Thompson, B. (1994), “The Concept of Statistical Significance Testing,” *Practical Assessment, Research & Evaluation*, 4:5-

²³ American Association for Public Opinion Research (2005), *AAPOR Code of Professional Ethics & Practices*. <http://www.aapor.org/aaporcodeofethics>, paragraph I.A.2, 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 the plain packaging in achieving the public policy goals given in paragraph 1.4 above were not considered in the formulation of my opinions. I have conducted my own literature review on the topics at issue, and am satisfied that my review has included all relevant studies as listed in paragraph 4.2.

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 As discussed above at paragraph 1.4, I have reviewed publicly available consumer studies and papers that relate specifically to the evaluation of the impact of plain packaging and were not evaluated previously by Dr Keegan in the Reports.

4.2 My review has therefore focused on several key primary research studies 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 1.4. The Studies are listed below in chronological order (from most recent to oldest) with a study-by-study analysis set out from paragraph 4.3 below.

- (a) Hammond, D., Dockrell, M., Arnott, D., Lee, A. and A. McNeill (2000). Cigarette pack design and perceptions of risk among UK adults and youth. *The European Journal of Public Health*, Advance Publication, 1-7. Hereinafter referred to as “Hammond, et al. (2009)”.

- (b) Hammond, D. and C. Parkinson (2009). The impact of cigarette package design on perceptions of risk. *The Journal of Public Health*, Advance Publication, 1-7. Hereinafter referred to as “Hammond and Parkinson (2009).”
- (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*, Advance Publication, Hereinafter referred to as “Moodie and Hastings (2009).”
- (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*, Advance Publication. Hereinafter referred to as “Germain, et al. (2009).”
- (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. “Doxey (2009).”
- (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. Hereinafter referred to as “Bansal-Travers et al. (ND²⁴).”

DETAILED ANALYSIS OF THE STUDIES

Hammond, et al. (2009)

4.3 This study reviews data obtained from 516 adult smokers and 806 minors aged 11 – 17 in the United Kingdom who participated in an online survey. Researchers assessed the degree to which adult smokers and minors perceived differences in tar, lower “health risk”, taste, and “attractiveness” of different packages. Its relevance to the plain packaging debate is seen in the conclusion that states: “*The findings also demonstrate that removing*

²⁴ By ND, I mean “no date”.

promotional information from the packs with a standardised appearance reduces false beliefs about the risks of cigarette brands. Perhaps most important, youth were less likely to identify brands in plain packages as more appealing if they were to try smoking.” However, the relevance of the differences in the constructs under investigation – the perceptions of tar, lower “health risk”, taste, and “attractiveness” – to changes in behaviour are not established. For example, the key construct of “attractiveness” was not defined by the authors of the study (respondents were simply asked: “*Of these two brands, which is the most attractive?*”) and why or how differences in “attractiveness” change behaviour is simply speculative.

Analysis

4.4 This study suffers from study design problems and empirical measurement issues, as well as a questionable sampling frame. It fails to account for the issue of prior knowledge and the degree to which a “false belief” is simply an artifact of the failure to sample non-smoking adults. Most relevant to the plain packaging debate is that the study does not provide any effective comparison amongst brands when all would be plain packaged. Overall, while it purports to examine the intentions of individuals, it does not have either a theoretical or empirical model of how those decisions are made, either ideally or in the practical reality of the marketplace.

4.5 In designing paired comparison choice studies such as this, it is important that the study be structured so that the experimental design meets accepted standards of efficiency, orthogonality and balance that allow for the proper statistical estimation (as noted in paragraph 2.10).²⁵ We can see this failure when looking at the issues of orthogonality and balance in particular. As noted earlier, orthogonality implies that the effects of the specific components (or attributes and levels, see below) of the packages on choice can be assessed independent of the alternatives. Balance implies that each attribute and level of the options considered appear the same number of times. For example, the brands

²⁵ Street, D. and L. Burgess (2007). *The Construction of Optimal Stated Choice Experiments*, Hoboken, NJ. I will discuss later that there are other classes of experimental designs that are not orthogonal. These are, however, quite complex and their discussion is peripheral to the evaluation of the studies being discussed in this report, none of which use proper experimental design approaches of any type.

Marlboro, Richmond and Silk Cut appear only once each, Benson & Hedges (“B&H”) twice, Mayfair six times and Lambert and Butler 4 times in the 14 choices. A lack of balance makes it nearly impossible to estimate whether the effects are arising because of real differences or because of the number of times the attribute appears or because of other co-varying factors. For example, because the subjects never see a brown plain pack versus a white plain pack across brands, size, and so on, the researcher will not be able to assess with any degree of accuracy the difference between plain white and plain brown packages or whether the effects are confounded by brand, or price or availability or any other influences. This is made even more difficult by the addition of brown versus white plain packages. By adding ‘plain brown’ and ‘plain white’ as possible alternatives, you cannot tell whether it is the ‘white’ or ‘brown’ colours or their ‘plainness’ that are influencing choices. In simple terms, by adding a ‘brown’ plain pack and a ‘white’ plain pack element, the authors have introduced another variable, which, in my opinion, has not been accounted for in their study design.

4.6 In this study there are 5 package ‘attributes’ – brand, colour, smoothness, “light” and size. Brand has six ‘levels’ – Marlboro, Mayfair, L&B, Richmond, Silk Cut, and B&H. Colour has 8 levels – Gold, Red, Silver, Purple, Dark Grey, Dark Red, Plain White and Plain Brown. Smooth – labeled Smooth or not – Size – Regular and King – and Light – labeled Light or not – have two levels each. This would imply a $6 \times 8 \times 2 \times 2 \times 2 = 384$ design, which in theory is workable. However, this implies a very large number of possible product alternatives that have to be incorporated into the experimental structure. The authors neither explain the design they have chosen, nor do they justify it in any manner that is consistent with best practice experimental design science.²⁶ Nor, when they exclude potential options from the mix of alternatives they present to their subjects, do they justify why those alternatives are being excluded. Overall, giving the individual only 14 choices ensures that: (a) the design is very inefficient statistically and unlikely to create an estimateable decision model and (b) is unbalanced as the various attributes/levels appear an unequal number of times (as noted above). The implication is that we can say nothing definitive about the nature of the choices that would arise because we cannot

²⁶ Street, D. and L. Burgess (2007). *The Construction of Optimal Stated Choice Experiments*, Hoboken, NJ.

isolate the specific effects of the components of the packaging. Nor without a properly structured design can one justify general statements as to the impact of specific colours, brands or labels.

4.7 The science underlying the application of experimental methods to consumer choice can be extremely complicated, but, putting it in simple terms to assist a lay reader of this report, there are two key points to note. First, for the reasons set out in paragraphs 4.5 and 4.6 above, the authors have not accounted for all the pack design variables they have introduced and which could influence decision making and would allow the researcher to make valid statements about the influence of the composition of the package on the decisions being made. For example, I note in paragraphs 4.5 and 4.6 that basic aspects of experimental design science – orthogonality and balance – are being ignored. Researchers can, in certain circumstances, apply what are known as non-orthogonal and non-balanced designs, which adds even more complexity to the study and requires that quite sophisticated statistical analyses be applied that account for the nature of the experimental design. However, in only one case – the Germain, et al. (2009) study to be discussed later at paragraphs 4.32 to 4.37 – was there any attempt to discuss the nature of the design chosen and why it was chosen. This calls into question the plain packaging findings of the Hammond, et al. (2009) study as the results could be arising for any number of effects that the researchers failed to control for in the structure of their experiments. Secondly, all of the possible permutations that are relevant to meaningful conclusions should form part of the study design. In paragraph 4.6, I noted that there were 384 possible alternatives that could potentially arise from the mixture of ‘attributes’ and ‘levels’ in this study. This does not mean that each person needed to see all 384 alternatives or even that all of the 384 permutations needed to be in the experiment. However, the experimental design had to be structured statistically so that the researcher was confident that they could get as much of the information that was possible given that there were 384 alternatives. This is what is known as the ‘efficiency’ of the experimental design. Very simply, it asks, how many and which of the 384 possible alternatives do people have to see so that the researcher gets enough information to make a statistically valid guess at the decision model being used.

4.8 There is also the issue of the appropriateness of the sample. It is unclear why only adults who smoked were the relevant comparator group to minors who both did and did not smoke. This is particularly intriguing given that one of the operative measures was “false beliefs”, which appear to be strongly linked to whether or not the individual smokes. So a critical comparison is missing and that is between non-smoker adults and non-smoker teens. This is telling in that table 2 in the paper reveals that there is no difference in the number of cigarettes smoked a day and perceptions of package differences, while the only consistent finding for minors in table 3 in the paper is that smokers were more likely to say there was a difference. This could be arising simply due to the fact that smokers were more knowledgeable (or believed they need to make what appeared to be a ‘knowledgeable’ choice).

4.9 The questions designed for the study make assumptions about a respondent’s knowledge. Asking questions about “which brand do you think would have the *most* tar?” assumes that the individual has knowledge of the meaning of ‘tar’ and an ability to determine the degree of tar yield for each brand. Although perhaps more relevant for smokers, it is less likely that non-smokers would have knowledge of what tar was. The same is true of the other questions, where there is an assumption that the question itself is relevant and should be answered. As noted in paragraph 4.8, we see a clear link between the statement that there is a difference and experience with cigarettes. This may simply be related to the fact that those with less knowledge would exhibit higher variance in their responses (hence the effect would appear to be smaller).

4.10 The questions asked of the subjects are leading, but in a subtle way. Asking about ‘tar’ will cognitively alert subjects to the importance of ‘tar’ independent of whether or not they consider it relevant. As noted in paragraph 4.9 above, this is most likely to be most relevant to those who need this ‘priming’ to determine what it is that they should be evaluating. In other words, the sheer fact that a researcher is asking the question signals to the respondent that the question is important. Hence, leading questions draw attention to, and can cause the subject to attach more importance to, something that they may not have initially viewed as being important. One common solution to this is to ask a series of

'clearing' questions that reduce the likelihood that respondents attach too much importance to any one question being asked.

4.11 The issues discussed in paragraphs 4.9 and 4.10 bring to the fore the problem of what are known as "pseudo opinions". The act of asking about an issue 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. For example, Bishop et al. (1980)²⁷ utilized 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 (p. 202)."

Hence, it is entirely conceivable that the subjects in these studies are expressing "pseudo opinions" about vague policy options that have little relationship to their preferences and would not reflect what would motivate their actual behaviour in more realistic circumstances.

4.12 The questions ultimately do not represent actual decisions. For example, the conclusions that are drawn are based upon a belief that "lower tar", "smoother taste" and "package attractiveness" are the operative decision making criteria, without any validation that that is indeed the case in reality (an incentive compatibility issue). Nor do we know the degree to which they serve as effective differentiators for consumers in their choice. In other words, we do not know the ordering of their importance in making a decision and how they might stand up against other salient attributes, such as price, brand, availability, peer group effects and so on. It is best practice in choice experiments to span the domain

²⁷ Bishop, G., Oldendick, R., Tuchfarber, A. and S. Bennet (1980), "Pseudo-Opinions on Public Affairs," *Public Opinion Quarterly*, 44: 198-209.

of all the relevant attributes the consumer considers in making a choice,²⁸ most notably price (which is excluded from this study) and also product availability. Otherwise, one risks biasing the results by making those factors that the researcher highlights as the most important in making a choice.²⁹

4.13 The choice questions do not represent actual choices, and as a result, force the subject into a situation where they are speculating about “trying to reduce” smoking and/or “making it easier to quit”. Again, we do not know anything about the hypothesized decision model related to smoking reduction or cessation and whether the questions are increasing the salience of the packaging as a driver of this behaviour (otherwise, why would the research be asking the subject the question?). In other words, we do not know what it is that the individual is actually thinking when attempting to link the act of “trying to reduce” smoking to the package; in other words what the cognitive process is that is driving the answer to the question. This, again, is an example where a “pseudo opinion” (discussed above at paragraph 4.11) could arise due to the fact that the individual is being asked to speculate on something for which there may be no relationship at all.

4.14 The statistical analysis is inappropriate. With paired comparison choice tasks the appropriate modeling structure is a logit.³⁰ The dependent variable would simply be a binary or multinomial choice – which option(s) was(were) chosen – and the independent variables would be the experimental design. The individual level factors – such as gender, age, social status, smoking status and so on – would be incorporated as covariates.³¹ This

²⁸ See, for example, Train, K. (2003). *Discrete Choice Methods with Simulation*. Cambridge, UK: Cambridge University Press, who notes that the choice sets presented to people should be “*exhaustive, in that all possible alternatives are included*” (p. 15).

²⁹ Hensher, D., Rose, J. and W. Greene (2005), *Applied Choice Analysis: A Primer*, Cambridge, UK: Cambridge University Press.

³⁰ A logit model estimates the probability that an option is chosen based upon the levels of the attributes that option possesses plus a random component. As noted by Train (2003), it is “by far the easiest and most widely used discrete choice model. ... Its popularity is due to the fact that the formula for the choice probabilities takes a closed form and is readily interpretable.” The structure, history and appropriateness of this model is discussed by McFadden in his 2000 Nobel Prize Lecture; McFadden, D. (2001), “Economic Choices,” *American Economic Review*, 91:351-378. For rating scales it is more appropriate to use a multinomial probit model (see, Haaijer, R. and M. Wedel (2007) for the difference in the statistical models).

³¹ How these ‘covariates’ are added into the model is will be based upon how they are hypothesized to influence a decision. Train (2003) and Henscher, et al. (2005) show how different statistical models

would allow for a more immediate comparison of the influence of the package attributes and their levels and would also account for the heterogeneity amongst individuals. It would also allow for a direct modeling of the decision making structure of the individual and would hence alleviate the need for making inferences from indirect questioning.

Conclusion

4.15 This study applies an inadequately designed paired comparison choice study and an inappropriate statistical analysis. The researchers also do not have a sampling frame that allows them to draw the conclusions that they do as they are comparing knowledgeable adult smokers with both knowledgeable and unknowledgeable non-adult smokers and non-smokers. 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 1.4 above.

Hammond and Parkinson (2009)

4.16 This study does not speak to plain packaging per se but is included in this assessment for two reasons. First, despite its subject matter, it draws a number of unsupported conclusions with respect to plain packaging. Second, its structure and design are similar to Hammond, et al. (2009) evaluated in paragraphs 4.3 to 4.15 and allows us to examine the validity of the general approach in this and the prior study.

4.17 The main argument of this paper is that “*current regulations have failed to remove misleading information from tobacco packaging*” based on the finding that individuals are inappropriately associating specific words as evidenced by the correlations between different researcher-created scales.

Analysis

4.18 The authors recruited 603 respondents aged 18 years and over, smokers and non-smokers, between January and March 2007 from shopping malls in Ontario, Canada. As

need to be used to account for the different ways in which factors such as age, social economic status or other demographics are accounted for in choice models.

discussed above, this study utilises the same sort of paired comparison choice study as Hammond, et al. (2009) and suffers from the same issues associated with an inadequate experimental design, the salience of the packaging being overstated, poor question construction, and a failure to model the data appropriately.

4.19 The packages designed by the researchers for this study have no clear attributes and levels so it is difficult to understand what the actual experimental design was or should have been. What is obvious is that the 9 choice sets presented are unbalanced and the experimental design inefficient, not allowing for an appropriate statistical analysis. For example:

- (a) Light, Full-Flavour and Regular appear in 2 of 9 sets but not in a manner that allows us to determine the independent effects of any of the labels. For example there are Light versus Full Flavour and Light versus Ultra Light options, but this tells us nothing of how Ultra Light would compare with Full Flavour as no such comparison is ever made.
- (b) Mild, Smooth, Silver, Ultra Light, “6”, “10”, White Symbol, Grey Symbol, Lighter Colour, Darker Colour, and Charcoal Picture all appear once meaning that we can not determine the effect of these attribute levels since there is only one comparison for each.
- (c) Of the brands, Kent appears six times, Mayfair twice and Richmond once revealing again a poorly designed set of options with excessive imbalance in the brand alternatives presented.

4.20 Another severe limitation of this study is the bias induced by creating scales based on what the researchers believed should be true rather than any theoretical model or via pre-testing procedures that would justify their conclusions. As noted on page 3: “*prior to the study, one package from each pair was identified as the package most likely to be rated as higher tar, smoother taste and lower health risks, based on a priori hypotheses*”. Exactly what these a priori hypotheses were is never explained. This is a particularly worrying issue for two reasons. First, as noted by the researchers, they biased the

presentation of the package alternatives by having what they considered to be the “*higher tar, smoother taste and lower health risk*” alternatives always being “*listed first for each pair*”. Proper designs would require this to be randomized or structured in a manner that does not potentially bias any responses. Second, the conclusions of the paper are based entirely on the correlations seen amongst these artificially created researcher-based scales. For example, the statement: “*These findings raise important questions about taste versus health descriptors in cigarette packaging*” (page 7) is based entirely on the correlations amongst the *a priori* scales for which there are no justifications given.

4.21 It is more worrying when one compares results between Hammond, et al. (2009) and this study and a number of inconsistencies that appear to be sampling related. First, this study was conducted with adults and one conclusion is that “*smokers were significantly more likely to perceive differences in taste, tar delivery and health risk*” (page 6). However, Hammond, et al.’s (2009) table 2 discussed in paragraph 4.7, shows no effects of difference perceptions from smoking intensity. This reinforces the conclusion that leaving out non-smoking adults from that sample makes the comparisons invalid. Second, the percentage of adults who indicate “no difference” in Hammond, et al. (2009) hovers around 60% over all the packages examined. However, in this study the numbers average less than 10%. It is difficult without any hypotheses to understand such large differences and why this sample would choose the “no difference” option so little. One possible conclusion is that the structure of the study is itself an ‘artifact’ that is influencing the choices being made by the subjects.

Conclusion

4.22 Based on the above analysis, this study provides no valid evidence to support the propositions that the packaging information is “*misleading*” and “*deceptive*” and that “*these terms [health and taste descriptors] are equivalent in the minds of many smokers when used on packaging.*” The section of the paper “What this study adds” is speculative and inappropriately extrapolates findings to the plain packaging issue, when these questions were not themselves studied specifically. Put simply, this study does not provide reliable evidence of the type set out in paragraph 1.4 above.

Moodie and Hastings (2009)

4.23 This paper presents a small sample diary panel examination of cigarette packaging. Although it did not specifically examine plain packaging, it attempts to draw a series of conclusions about plain packaging based upon its results. Specifically, it argues at page 12 that: *“Tobacco packaging is no longer the ‘silent salesman’ it once was, it now has to shout out loudly. The increased volume is used to defy advertising bans and drown out health warnings. The only consistent and effective policy response is to mandate generic packaging.”* However, the study never address this issue and the researchers are drawing conclusions that reveal a lack of objectivity and a disregard for what their own evidence is revealing. In addition, the study suffers from some basic methodological flaws.

Analysis

4.24 The most glaring limitation of the study is its sample. The sample includes only “on average” 20 individuals, which is significantly below what would be considered acceptable for drawing generalizable conclusions. There are well-established procedures for determining an appropriate sample size that relate to the variance of the underlying phenomenon being studied, the heterogeneity of the population being sampled, and the size of the hypothesized effects in the researcher’s model.³² Ultimately, the goal of finding an appropriate sample is to see the hypothesized ‘signal’ – which is the phenomenon of interest – above the ‘noise’ of the population. This is critical to the validity of a study’s conclusions because as noted by Lenth:³³

The study must be of adequate size, relative to the goals of the study. It must be “big enough” that an effect of such magnitude as to be of scientific significance will also be statistically significant. It is just as important, however, that the study not be “too big” where an effect of little scientific importance is nevertheless statistically detectable. Sample size is important for economic reasons: An undersized study can be a waste of resources for not having the capability to

³² Cochran, W. (1977), *Sampling Techniques, 3rd Edition*, Hoboken, NJ: Wiley.

³³ Lenth, R. (2001), “Some Practical Guidelines for Effective Sample Size Determination,” *The American Statistician*, 55: 187-193.

produce useful results, while an oversized one uses more resources than are necessary. (p. 187).

The sampling frame used in this study is based on “*age (21-30, 31-35), gender [male, female], and social class (ABC1, C2DE), weighted in favour of low income groups*” (p. 4). This would imply 8 cells $2(\text{age}) \times 2(\text{gender}) \times 2(\text{social class})$. With only 20 panel members this means that each cell has “on average” $20/8 = 2.5$ individuals. It is not clear what the researchers mean by “*weighted in favour of low income groups*” but one must assume that several of the sampled groups have “on average” 3 individuals and the others 2, that “averages 20”. Given that the population of smokers is in the millions it is inappropriate to make general statements based upon 2 or 3 people in specific demographic groups. For example, the opinions or observations of 2 or 3 individuals representing 21-30 year old males from the ABC1 social class are hardly scientifically or even logically valid as measures of what a larger population is observing or believing.

4.25 It is unclear how this panel was formed (other than via age, social status, and gender) and what “on average” means. It is unclear whether one person filled in the diary between 2002 and 2007 or whether many different individuals filled in the diary. If it was the former (one person filling in the diary for 6 years) then, at best, the panel used could be considered in line with focus group research, where the intention is not to draw conclusions but to help formulate hypotheses and to aid in the structuring of larger scale and more statistically valid research. If it was the latter (and many different individuals filled in diaries), then one is faced with the fact that the data is not comparable longitudinally since the basis on which the diary is being filled in (e.g., where the person shops, how much they smoke, whether they notice things more than others or not, and so on) will change.

4.26 The entire point of diary panels is that they provide you with information that is “within the individual” over a long period of time. Hence, we do not know whether any conclusions that would be drawn across time are due to actually changes being recognised or different individuals doing the recognising. Given the sample limitations, the drawing of definitive conclusions is unwarranted.

4.27 There is also a salience issue in that individuals are being asked to fill in diaries that focused on “tobacco marketing” (a term used, but not defined in the study) without making it clear what it was that they were specifically asking the panel to notice or asking them to include other items of relevance (such as pricing). All that we know from the information given in the study is that “*the diary provided the panel with the opportunity to record any examples of tobacco marketing encountered*” (p. 4). Hence, given that they chose to be involved in the study they will, almost by definition, be more likely to notice marketing aspects as this is what they are instructed to do. Whether or not they would do this in reality is not assessable (hence we also have an issue of whether the diary is creating incentives to notice marketing when in reality it would be ignored).

4.28 It would have been more relevant to: (a) ask them to fill in a diary that dealt with other aspects of tobacco purchasing – so as to reduce the salience of marketing per se – and (b) potentially include other product categories for which calibration of their responses were possible so that the extent to which the individual’s tendency to notice any marketing by any product is determined, and (c) asking the panel members to indicate the retail outlets (or location) in which they observed the “tobacco marketing”. The researchers could then cross reference whether what was observed by the panel member matched with independent verification of what was available at a sample of the retail outlets they indicated.

4.29 The approaches outlined in 4.28 would ensure that at least the panelist’s self-reports were subject to methods that were closer to industry best practice when it came to validation. Absent this, the researchers could have utilized independent evaluators to determine whether or not their conclusions as to the meaning of the panel comments were independent of any researcher bias. As noted by Byers and Wilcox:³⁴ “*whichever type of analysis is employed, ... [data and results] should be submitted to another researcher for validation. Cross-validation will enhance the objectivity and reliability of the research.*” (p. 75).

³⁴ Byers, P. and J. Wilcox (1991), “Focus Groups: A Qualitative Opportunity for Researchers,” *J. of Business Communication*, 28: 63-78.

4.30 Most worrying overall is the lack of objectivity of the researchers. This arises when statements made in one part of the paper are contradicted later on. For example, on page 8 of the document it is stated: *“The panel were highly price sensitive, with almost all references to packaging relating to price-marking and value”*. On page 11 it is stated: *“The smokers’ panel seldom mentioned image and innovation based packaging ...”* which is followed by an attempt to explain this away. However, in the discussion (page 11), the researchers conclude:

Packaging is used differentially to communicate value,, created desirable brand imagery, ..., and stimulate interest in all tobacco categories through innovation.

However, their own earlier statements would lead to the opposite conclusion. For the most part, consumers focus on value and price and ignore other aspects of the product. It is possible that they notice other aspects of the product – what the researcher’s call “tobacco marketing” – because they were asked by the researchers to do this. However, whether their noticing mattered to what they would purchase is never investigated or validated. It is simply assumed that because the panel member noticed the package when asked to, that would be information relevant to their purchasing decision.

Conclusion

4.31 In the end, the researchers jump from a small sample series of self report panel diaries to quite grand conclusions that “package marketing” is effective (which is opposite to what the self reports seem to indicate, and for which they have no objective measures) in influencing behaviour (for which they have no measures at all, either objective or subjective). They also conclude that the purpose of use of tobacco packaging is to *“defy advertising bans and drown out health warnings”* (again for which there is no evidence at all, since the panel members were not asked to indicate the degree to which they failed to notice health warnings, and no other information is provided as to the supposed intentions behind the packaging).

Germain, et al. (2009)

4.32 In this study, researchers assessed the degree to which those between the ages of 14 and 17 rated different plain packaging alternatives and existing alternatives on package and smoker image. They utilised a 3x5 between-subjects design where the subject rated only one of the 15 alternatives possible (a 3x5 design will create 15 alternatives). These alternatives were created by varying the brand (3 conditions) – between Winfield, Peter Jackson and Longbeach – and the package design (5 conditions) – between the existing package and plain packets with decreasing size of the brand font (3 conditions) and one condition with the health warning increased to 80% of the pack. The researchers conclude (at page 6) that *“when a cigarette pack is progressively stripped of its colour, imagery, and branded fonts, adolescents perceive the packs as less appealing.”*

Analysis

4.33 A major flaw of the study is its set up. It is desirable in survey and experimental research to ensure that the purpose of the study is shielded from participants to the greatest extent possible. The ideal study is the double-blind study, where both the person executing the study and the participant does not know the purpose of the study. However, with careful research procedures single blind studies – where the participant does not know the purpose of the study but the study executer does – can work as well. Without some degree of such ‘blinding’ there is the possibility that the participant will be biased because of the way in which they see their participation in the study.³⁵ In this study, two factors create potential biases. First, the study was identified as being conducted by the Cancer Council of Victoria, a fact that would immediately identify the study as one that was associated with policy prescriptions in line with the sponsoring organisation; e.g., such as increased regulatory intervention. A second point reinforces this: *“Parents were informed that the study results would help to guide the development of tobacco control policies in Australia ...”*. It is conceivable that (a) it would be easy for a parent to relay the purpose of the study to their child, thereby influencing that child’s

³⁵ Bannie, E. (2008), *The Basics of Social Research*, 4th Edition, Belmont, CA: Thomson.

responses, and (b) that parents who wanted to influence their child in some way either would or would not permit the child to be involved in the research.³⁶

4.34 It is also not clear from the study whether or not the categorisations of smoking status are valid representations of actual behaviour. Although the researchers acknowledge that the sample is not representative of the population, it is important to attempt some sort of validity check on whether or not the self-reported smoking proclivities are likely to be valid. For example, the number of self-reported “established smokers” in their sample amounts to approximately 17 percent with those saying they are “non susceptible non-smokers” averaging around 45%. A simple check would be to examine how this matches up with the actual behaviour of adolescents in the population sampled. This is important, as it is generally understood that self-report surveys of this type led to significant under-reporting.³⁷

4.35 The practical empirical importance of paragraph 4.34 is seen in their figure 2 (on page 5), which shows that perceptions differ almost completely between established smokers and all others. Established smokers effectively do not differentiate at all amongst the different packs while non-smokers (of any category) appear to differentiate. What this implies is that established smokers are uninfluenced by the ‘plain’ and ‘unplain’ (unbranded and branded) pack alternatives. However, it may also imply that the less knowledgeable non-smokers (who do not differ in their ratings by the categorisation of type of non-smoker) are reacting to packaging design because the study itself is making the packet salient.

4.36 This study concentrated only on the perceptions of packs without any consideration for the link between those perceptions and the ultimate behavioural

³⁶ The WAPOR Code of Professional Ethics and Practices (<http://www.unl.edu/WAPOR/ethics.html>) indicates that “respondents shall be informed of the sponsor of a survey unless the researcher and sponsor believe this will bias responses.” It is clear that in this case such identification can induce such a bias because of (a) the social orientation of the sponsor and (b) the fact that the study is being identified as a way to influence public policy in a specific direction.

³⁷ Adams, J., Parkinson, L., Sanson-Fisher, R., and R. Walsh (2008), “Enhancing Self-Report of Adolescent Smoking: The Effects of Bogus Pipeline and Anonymity,” *Addictive Behaviors*, 33: 1291-1296, reported that 25% to 26% of Australian minors reported smoking in the last week and nearly 40% in the last month when they were potentially subjected to a biological validation that would have confirmed whether or not they had actually smoked.

outcomes attempting to be achieved by various regulators proposing plain packaging for tobacco products (paragraph 1.4 above). Hence, the conclusions must ultimately be considered as limited by the degree to which such perceptions have any influence on reducing the likelihood of smoking uptake or cessation by minors. Ultimately, the question is whether or not these perceptions: (a) should be included in a model of decision making that relates to the public policy goals outlined in paragraph 1.4; and (b) the extent to which they matter materially when considered against other relevant alternatives to reduce minors smoking – such as price or availability or other factors that can be expected to influence smoking uptake, cessation or purchasing. For example, in the case of availability, there is some evidence³⁸ that intervention that influences a minor's accessibility to cigarettes can be an effective deterrent to smoking uptake by minors when enforced.³⁹ Hence, the subjects may perceive a difference but that difference may have absolutely no influence on actual decisions when other critical factors are put onto the table. Since the study never examines these other factors we have no idea how important perceptions really are.

Conclusion

4.37 The researchers note that the design of their study allows them to determine the value of specific elements of the pack on perceptions; however, what they cannot say is the degree to which such perceptions ultimately lead to specific choices in the context in which realistic purchases are being made and the decision to smoke or not to smoke is being made. This is ultimately a salience or perceived importance issue. Because the study cannot blind the subject to its intent and because the researchers focus very specifically on the packaging facets alone – e.g., things like font size – they ignore the overall decision model that the consumer would be expected to use when making actual purchases. Such a decision model would, at a minimum, be expected to include price,

³⁸ Stead, L. and T. Lancaster (2000), "Systematic Review of Interventions for Preventing Tobacco Sales to Minors," *Tobacco Control*, 9: 169-217; Stead, L. and T. Lancaster (2008), *Interventions for Preventing Tobacco Sales to Minors (Review)*, The Cochrane Collaboration.

³⁹ I am aware that, for example, several states in Australia have criminalised the act of purchase of tobacco products by minors. In addition, I note that all of Australia's six states, as well as the Australian Capital Territory and the Northern Territory, have enacted prohibitions relating to proxy purchase.

product/brand availability, and purchasing context. This is important if the evidence from the study is to be used to form part of sound policy making decisions. By not accounting for the 'excluded' elements of the decision model, the researcher may significantly overstate the importance of the things that were studied, while significantly understating those facets that were excluded. Hence, a variable that may actually be inconsequential to actual purchase decision making risks having its importance overstated when the role of other variables are not fully taken into account. So even if the researchers discovered that, at the margin and holding everything else constant, something like font size mattered, this says absolutely nothing about whether font size matters when other factors are taken into account. Another way to think about this is that if I were asked whether I prefer a BMW or a FIAT I certainly would say that I prefer the BMW. Or if I were asked would I prefer a black car to a yellow car, I would say that I preferred the black car. However, this does not mean that I would choose a €50,000 BMW over a €15,000 FIAT. Or a €50,000 black BMW over a €45,000 yellow BMW.

Doxey (2009)

4.38 This study was a postgraduate student dissertation submitted to the University of Waterloo and supervised by David Hammond (an Assistant Professor at that University). The structure of the study was similar to those of Hammond, et al. (2009) and Hammond and Parkinson (2009) and, fundamentally, suffers from most of the major experimental design and statistical analysis limitations of those studies. The study was not specifically related to plain packaging – concentrating more on perceptions of packages by females – but includes as one of its research questions, “*to what extent does 'plain' packaging reduce perceptions of brand appeal, positive attitudes toward smoking, perceptions of health risk, and beliefs about smoking and weight control among young women?*” (page 14). A total of 512 females between the ages of 18 and 25 were recruited throughout Canada through a market research service.

Analysis

4.39 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.

4.40 This issue of the perceptions are 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?*” – and health risks – “*How would the health risks of these cigarettes compare to other cigarette brands?*” – are (a) assuming that the individual is competent to understand the meaning of “tar delivery” and (b) define “health risks” in a manner that is comparable between individuals (in 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 persons, 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.

4.41 The research uses 5-point scales to measure her 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 were 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:⁴⁰ “*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 entity. Failing this, the conceptual definition of the construct will be inadequate for indicating how the construct should be (operationally)*”

⁴⁰ Rossiter, J. (2002), “the C-OAR-SE Procedure for Sale Development in Marketing”, *International Journal of Research in Marketing*, 19: 305-335.

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 – what 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 then the measure is conceptually invalid.

4.42 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: “*In your opinion, how appealing would this brand of cigarettes be to young women your age compared to other brands on the market?*”. 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: “*How would the health risks of these cigarettes compare to other cigarette brands?*”. 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 has no idea of the “health risks” of a brand (and is willing to say so) in the same category as someone who views the risks as high!

4.43 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 create 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 8 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 well-established formulation described in paragraph 4.40, 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 but intermediate measures are completely confusing. If someone rated the 8 brands as a 4 on "tar" this could arise because 4 of the brands were perceived to have a "lot less tar" and 4 of the brands were not rated "Don't know" or all 4 of the brands were perceived to have "a little more tar" and 4 "a little less tar". In other words, there is absolutely no way to determine what actual perceptions led to the intermediate scores. Mean statistics are provided for these indices in Doxey (2009)'s tables 4, 8, 12, and 16 and what we see is, as a matter of accepted norms for such assessment, concerning. 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".

4.44 As in Hammond, et al. (2009) the experimental conditions are 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, as noted earlier in paragraph 2.10 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.

4.45 This criticism of the experimental design means that there is no logical link between the design and the statistical model being estimated. In other words, the vast majority of the analyses are based upon pairwise comparisons (such as "standard" versus "plain") 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 this study, we are left with the result that we can say absolutely nothing generalisable about any of the results of the analysis.

4.46 This last point is relevant to conclusions drawn with respect to plain packaging. The researcher purports that her findings showed that plain packaging “*reduced brand appeal and perceived taste, beliefs about an association between smoking and weight control, and beliefs that smokers possess more positive personality and physical traits*” (p. 64). But, it is equally possible that her findings show that plain packages lead to many more “don’t knows” and “no differences” as these are scored in exactly the same way as lower brand appeal, less taste and so on. The fragility of these findings and also those of Hammond and Parkinson (2009) are seen in the finding of Doxey that “*while previous research [e.g., Hammond and Parkinson] has also demonstrated that plain packs reduce false beliefs about the health risks of smoking, we do not replicate this finding in the current study. This could be the result of the methodological limitation of not having side-by-side comparison packs, or, the choice of white background for the plain packs*” (p. 64). There is an alternative explanation. Because the studies discussed used flawed experimental designs, generalisable conclusions are impossible. Neither this study, nor those discussed earlier, applied proper methods nor did they have a theoretical behavioural model as a basis of their design and statistical analysis. Hence, the results are, at best, specific to the circumstances investigated and do not provide valid enough findings to make broader conclusions.

Conclusion

4.47 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 modeling is atheoretical and

does not align with any behavioural model. As with many of the other studies examined the results are all perception based and provide no evidence of the veracity of the findings to actual choices.

Bansal-Travers, et al. (ND)

4.48 This study is effectively a replication and extension of Hammond, et al. (2009) and Hammond and Parkinson (2009). Its conclusions with respect to plain packaging are that *“plain packaging may reduce many of the erroneous misperceptions of risk communicated through the design features on cigarette packs”* (page 2).

Analysis

4.49 A total of 197 adult smokers and 200 adult non-smokers were interviewed in June to July 2009 outside a mall in Buffalo, New York. As with Hammond, et al. (2009), Hammond and Parkinson (2009) and Doxey (2009), this study suffers from the failure to utilise acceptable experimental design approaches associated with studying choice. Given that this has been discussed extensively above in paragraphs 4.5, 4.19 and 4.41-4.42, it is just noted again that the structure of this study also lacks an efficient, orthogonal and balanced design that would allow a proper statistical analysis of the design components of the packaging.

4.50 This study also asks individuals to respond to questions where there is no indication that they have the knowledge to respond. As before there are concerns about asking individuals about the tar yields of different brand types – *“which pack would you expect to deliver the most tar if you were to smoke it?”* – but also speculative questions for which the average subject may have no knowledge or expertise – e.g., *“Between the two packs, which do you think would most appeal to youth under 18 years old?”*.

4.51 Along with the design issues, there is only one plain package alternative given and it is given in comparison to a basic red package (see below).



What this means is that any statements with respect to plain packaging can only be made with respect to that one comparison alone. No other comparison to any other aspect of package design would be valid as no individuals compared the plain package to any other combination of alternatives (as one would have with a proper orthogonal design). Hence, the statement – *“These results suggest that plain packing may reduce many of the erroneous perceptions of risk communicated through design features on cigarette packs”* – is, in my expert opinion, fallacious as the only comparison that is made is between one plain pack and one basic pack. To infer that one can then make general statements about design ‘features’ plural is unjustified by the structure of the study or its findings.

4.52 Also, as with the other studies using similar methods, all of the conclusions are based upon perceptions, not behaviours. Hence, we have no indication of the degree to which these perceptions matter to actual purchasing or how they would stack up against other product features that are no doubt more salient – such as price and availability.

4.53 Finally, this study had subjects only choose amongst one plain package against one basic alternative. If plain packaging was introduced the market choices would be all plain packaging. There is no indication in this study as to what the choice would be in that

market reality, nor the degree to which the perceptions would matter when the alternatives against which they have subjects comparing are no longer available.

Conclusion

4.54 This study is effectively a replication and extension of Hammond, et al. (2009) and Hammond and Parkinson (2009) and, as such, exhibits nearly all of the same evidential limitations. There is no behavioural model and all the conclusions are speculative extrapolations from perceptions. In addition, the experimental structure does not allow for any generalised statements to be made with respect to plain packaging on smoking cessation, uptake or reduction.

5. CONCLUSION

5.1 In section 4 of this report I have provided a number of detailed comments about specific aspects of the individual studies that relate to plain packaging. In this section of the report, I will summarize and expand upon that analysis by looking specifically at the criteria I discussed in paragraphs 2.5, 2.10, 2.11 and 2.12. Unlike section 4, which focused on each study individually, I will focus on them as a group and relate them back to the specific criteria.

5.2 The criteria outlined in paragraph 2.5 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 none of the studies examined 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. Because the studies all focused on packaging absent other salient attributes of the products (such as price) and factors that would potentially lead to an achievement

of the policy goals outlined in paragraph 1.4, 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 examined also failed to provide experimental or situational contexts that created scenarios in which the individual would be applying the decision model that they used when making purchasing decisions or decisions that related to the policy goals outlined in paragraph 1.4. The fact that all of the studies focused exclusively on package design created a situation where packaging features dominated the decision 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).
- (c) Finally, it is my conclusion that, as the Studies concentrated entirely on stated preferences and attitudinal measures, one cannot assume any predictive accuracy with respect to actual purchasing behaviour or the intended policy goals outlined in paragraph 1.4. 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. 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 and the operative measures outlined in paragraph 1.4, smoking uptake, smoking reduction, or smoking cessation. One would have to make a leap of faith about the importance of the various measures used in these studies to relate them to actual behaviours.

5.3 The criteria outlined in paragraph 2.10 relates specifically to the experimental studies examined. In the case of Hammond, et al. (2009), Hammond and Parkinson (2009), Doxey (2009) and Bansal-Travers, et al. (ND), 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. Three of the

studies – Hammond, et al. (2009), Hammond and Parkinson (2009), and Bansal-Travers, et al. (ND) – apply what are effectively the same, or very fundamentally similar, flawed approaches to the structure of their experiments. The fourth study – Doxey (2009) – also uses a similar flawed design approach but compounds this with a number of significant data manipulation and statistical errors (see, for example, my additional concerns regarding of Doxey (2009) set out paragraph 4.42 above). By contrast, Germain, et al. (2009) utilises a simple experimental design that is more efficient, balanced and orthogonal, but lacks validity because of their focus exclusively on packaging independent of the other attributes of the product. As the goal of choice experiments is to determine the decision model underlying choice, it is important not just to have a well-designed experimental structure but to include in the experimental conditions all the factors relevant to the determination of the different decisions individuals would reasonably be expected to make.

5.4 The criteria outlined in paragraph 2.11 help us understand the degree to which stated intentions align better with actual behaviour.⁴¹ Again, these criteria are relevant to the experimental studies just discussed in paragraph 5.3. As noted above at paragraph 2.11, stated intentions align better to actual behavior 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.
- (b) They are for durable rather than non-durable goods.⁴² Individuals are more variable in how they purchase non-durables and use more deliberative decision

⁴¹ Morowitz, V., Steckel, J. and A. Gupta (2007) “When Do Purchase Intentions Predict Sales?” *International Journal of Forecasting*, 23: 347-364.

⁴² 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.

- models for durable products. Cigarettes are a non-durable good generally, I assume, subject to frequent purchasing. Hence, this might imply that stated intentions are less accurate for cigarette consumption.
- (c) They are for short-term horizon decisions rather 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 1.4 – 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 5.4(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 over 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 studies considered choice amongst several alternatives, except Doxey (2009), who had subjects evaluate packets one at a time, and Germain, et al. (2009), who had subjects evaluate only one packet each.

Overall, it is my conclusion that none of the studies met enough of these criteria whereby even if their ‘stated intention’ measures were derived in a valid manner (which paragraph 5.2 and 5.3 concludes is not the case), one could argue that they would provide potentially valid predictions of actual purchasing behaviour.

5.5 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 – which in many cases assumed knowledge or led the respondent, were based on assumptions about the relevance of the construct in question to actual decisions, or where the construct was created in a manner that was invalid (as in Doxey’s (2009) various indices). The Germain, et al. (2009) study potentially biased those involved by identifying the purpose of the study as changing policy and the sponsor as the Cancer Council of Victoria. In all the Studies, the narrow focus of the tasks involved – either package choice or the filling in of a panel diary – immediately identified the purpose of the study as related to packaging. In the case of the

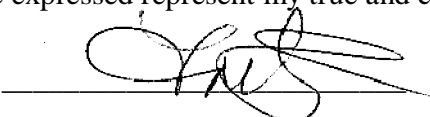
- Moodie and Hastings (2009), the researchers draw conclusions that appear to be predetermined and inconsistent with a more disinterested evaluation of their panel respondents' protocols. Their failure to have these protocols assessed by independent evaluators could also be considered a source of bias. In all of these experimental studies, there is an issue of response reliability as none of the studies examine actual behavior nor do they attempt to link their results to actual behavior.
- (b) The sampling frame in these studies varied. Hastings and Moodie (2009) panel was discussed in paragraphs 4.24 and 4.25 as being inappropriately constructed and a potentially serious source of bias. Their results were more in line with what one would expect of a focus group and few conclusions of a generalisable nature could be drawn from their panel. Hammond, et al. (2009), Hammond and Parkinson (2009), Doxey (2009) and Bansal-Travers, et al. (ND) all used non-random sampling, hence one cannot make population predictions based on their findings. Germain, et al. (2009) used a slightly more sophisticated sampling approach but again their sample is not constructed to generate a generalisable population prediction. In all of these cases this lack of an ability to make predictions about either the general population or the population of minors (both smokers and non-smokers), in my expert opinion, renders their findings irrelevant to the policy questions outlined in paragraph 1.4.
- (c) The analysis in these studies suffers from significant validity issues. Most particularly, the experimental design limitations outlined in paragraph 5.3 and discussed throughout section 4 imply that the statistical analyses seen in Hammond, et al. (2009), Hammond and Parkinson (2009), Doxey (2009) and Bansal-Travers, et al. (ND) must be considered to be invalid. Their failure to structure their experiments correctly implies that their ability to make generalisable conclusions is lost. All of these studies, as well as Germain, et al. (2009), jump from attitudinal results relating to intentions and views on "package attractiveness" to policy conclusions as to the veracity of plain packaging as a regulatory initiative. However, none actually test only plain packaging alternatives – which is the purchasing reality consumers would face – nor do they consider the other product

features and contexts which influence cigarette purchasing, such as price or availability. Hence, in my view, they are making conclusions based on false comparisons. In addition, none of their analyses deal directly with the key constructs of smoking uptake, reduction and cessation noted in paragraph 1.4 in a manner that goes beyond speculation on the part of the subjects in their studies. Finally, as noted in paragraphs 4.29 and 5.5(a) Moodie and Hastings (2009) draw unsubstantiated conclusions that packaging matters when their panelists clearly note that what is most noticeable to them is price.

5.6 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.

5.7 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: 30 November 2010

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.

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, and 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 his field 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 Associate Editor of the

Academy of Management Perspectives and am Co-Editor of 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 recognized by numerous organizations 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 Decisionmaking 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), 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, 2009–2011. 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.

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

Associate Editor, Academy of Management Perspectives, 2006–present

Associate Editor, Australian Journal of Management, 1995–2005

Associate Editor, Management Science, 1988–1990

Consulting Editor, Journal of International Business Studies, 2010–present

Editorial Board, Strategic Management Journal, 2007–present

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

Editorial Board, Global Strategy Journal, 2010–present

Editorial Board, Strategic Organization, 2006–present

Editorial Board, Corporate Governance: An International Review, 2007–present
 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).

Professional Consulting Experience (Selection)

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

8. EXHIBIT THREE – SELECTED PUBLICATIONS

Ex.3.1 “Using Frontier Analysis to Evaluate Company Performance,” *British Journal of Management*, 21, 4, December 2010 (with G. Yip & G. Johnson).

Ex.3.2 *The Myth of the Ethical Consumer*, Cambridge: Cambridge University Press, 2010 (with P. Auger and G. Eckhardt).

Ex.3.3 *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.3.4 “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.3.5 “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.3.6 “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.3.7 “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.3.8 “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.3.9 “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.3.10 “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.3.11 “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.3.12 “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.3.13 “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.3.14 “The Other CSR,” *Stanford Social Innovation Review*, Fall 2006 (with P. Auger, G. Eckhardt & T. Birtchnell).

Ex.3.15 “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.3.16 “Modular Strategies: B2B Technology and Architectural Knowledge,” *California Management Review*, 47, 4, Summer 2005 (with P. Richard).

Ex.3.17 “Client and Agency Mental Models in Evaluating Advertising,” *International Journal of Advertising*, 24, 1, 2005 (with M. Collins & G. Dowling).

Ex.3.18 “A New Perspective on the Integration-Responsiveness Pressures Confronting Multinational Firms,” *Management International Review*, 44, SI1, 2004 (with D. Midgley & S. Venaik). Reprinted in: J.A. Krug and J.D. Daniels, *Multinational Enterprise Theory*, Thousand Oaks, CA, 2007.

Ex.3.19 “What Will Consumers Pay for Social Product Features?” (with P. Auger, J. Louviere & P. Burke), *Journal of Business Ethics*, 42, 3, 2003.

Ex.3.20 “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.3.21 *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.3.22 “E-Business: Revolution, Evolution or Hype?” *California Management Review*, 44, 1, Fall 2001 (with D. Midgley, T. Coltman & A. Latekefu).

Ex.3.23 “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.3.24 “Understanding Institutional Designs Within Marketing Value Systems,” *Journal of Marketing*, 63, Special Issue, 1999 (with S. Carson, G. John & G. Dowling).

Ex.3.25 “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.3.26 “A Formal Model of Trust Based on Outcomes,” *Academy of Management Review*, 23, 3, July 1998 (with R. Bhattacharya & M. Pillutla).

Ex.3.27 *The Essence of Corporate Strategy: Theory for Modern Decision Making*, Sydney: Allen & Unwin, 1997 (with Jeremy Davis).

Ex.3.28 “How Well Do Patents Measure New Product Activity?” *Economic Letters*, 41, April 1993.

Ex.3.29 “New Products and Financial Risk Changes,” *Journal of Product Innovation Management*, 9, September 1992.

Ex.3.30 “New Product Innovations and Stock Price Performance,” *Journal of Business Finance & Accounting*, 19, September 1992 (with P. Chaney).

Ex.3.31 “The Return to Advertising Expenditure,” *Marketing Letters*, 3, May 1992, (with B. Cooil).

Ex.3.32 *EUROPEAN MARKETS AFTER 1992: Implications for Business Strategy*, Lexington, MA: Lexington Books, 1991, (with William C. Hightower).

Ex.3.33 “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.3.34 “New Products Over The Business Cycle,” *Journal of Product Innovation Management*, 7, December 1990.

Ex.3.35 “Diversification Strategy and Performance in Canadian Manufacturing Firms”, *Strategic Management Journal*, 11, September 1990 (with T. Nguyen and A. Seror).

Ex.3.36 “Rationally Determined Irrationality: An Extension of the Thesis of Rationality as Anti-Entropic,” *Journal of Economic Psychology*, 10, November 1989.

Ex.3.37 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 1.4, 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) Freeman, B., Chapman, S., Matthew Rimmer, R. (2008), The case for the plain packaging of tobacco products. *Addiction*, 103, 580-590.
- (b) Moodie, C., Hastings, G., Ford, A. (2009), A brief review of plain packaging research for tobacco products: Report prepared for the UK Department of Health.
- (c) Freeman, B., Chapman, S. (2009), Open source marketing: Camel cigarette brand marketing in the Web 2.0 world, *Tobacco Control*; 18:212-217.

- (d) Wilson, N., Weeraskera, D., Peace, J., Edwards, R., Thomson, G., Devlin, M. (2009), Misperceptions of lights abound: National Survey Data. *BMC Public Health*, 9, 126.
- (e) European Network of Smoking Prevention (2009). Plain tobacco product packaging as a means to protect young people and adult consumers.
- (f) Padilla, J., Watson, N. (2010). A critical review of the literature on generic packaging for cigarettes: A report for Philip Morris International.
- (g) Thrasher, J., Hammond, D., Arillo-Santillan, E.A. (2010). The alchemy of Marlboro: transforming “light” into “mild”. *Tobacco Control*; 19:342-343.
- (h) International Union Against Tuberculosis and Disease Control (2010). Tobacco Packaging and Labelling: Technical Guide.

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) Gallopel-Morvan, K. (2008). Impact of visual tobacco warnings and plain packaging on French people: Results of a qualitative survey (Part 1 – December 2008), Report for the French National Cancer Institute.
- (b) Gallopel-Morvan, K. (2009). Impact of visual tobacco warnings and plain packaging on French people: Results of a qualitative survey (Part 2 –June 2009), Report for the French National Cancer Institute.
- (c) Thrasher, J., Rousu, M., Hammond, D. (2009) Estimates of reduction in demand associated with different cigarette package warning label formats: An experimental auction among adult smokers in the U.S. (Unpublished manuscript).
- (d) SIRUS (Norway) (2009). Evaluation of pack design: Implications for tobacco control. (Unpublished).

- (e) Hammond, D. (2010). Plain packaging regulations for tobacco products: the impact of standardizing color and design of cigarette packs. *Revista de Salud Publica de Mexico* (In Press).
- (f) Hammond, D. , Doxey, J., Bansal-Travers, M. (2010). Impact of female-oriented cigarette packaging in the United States. Submitted to the *American Journal of Public Health* (In Press).
- (g) Rey, J.M. (2010). Effects of visual warnings and plain packaging on decreased demand of tobacco products: preliminary results (Manuscript in preparation).
- (h) Fraeyman, J. (2010). Ongoing research in Belgium examining youth attitudes to plain packaging. (Unpublished research).
- (i) International Tobacco Control Project (On-going). Seven Country Study on Plain Packaging. (Unpublished research).
- (j) Webb, L. (2010). Plain Packaging Research conducted at the University of Auckland (Unpublished research).