EXPERT REPORT

Research effects of standard packaging 2012 Australia

prof. Dr. Eddy KA Van Doorslaer
December 27, 2022

Introduction

In this report I have answered the questions and sub-questions that were put to me. To this end I have perused the file made available to me by Hogan Lovells (see attached inventory list).
In particular, my attention has focused on the most important documents on the subject. Those are:

Cochrane Database of Systematic Reviews 2017, Tobacco Packaging Design for Reducing Tobacco Use (see Appendix 6 to State's defense, January 14, 2020);

National Expertise Center Tobacco Control, part of the Trimbos Institute, Factsheet on generic tobacco packaging (plain packaging), June 2019 (see Appendix 3 to the State's defense, 14 January 2020);

Viscusi (2018), Assessing the Effect of Australian PP Regulation;

Dryden (2017), The effects of Standardized Packaging – an Empirical Analysis;

In addition, I have also completed the following studies:

Pascal A Diethelm, Timothy M Farley, "Refuting tobacco-industry funded research: empirical data shows a decline in smoking prevalence following the introduction of plain packaging in Australia," Tob. Prev. Cessation 2015;1(November):6 http://dx.doi.org/10.18332/tpc/60650

Kaul, A. and Wolf, M. "The (Possible) Effect of Plain Packaging on Smoking Prevalence in Australia: A Trend Analysis," University of Zurich Department of Economics Working Paper, June 2014

Tasneem Chipty, "Study of the Impact of the Tobacco Plain Packaging Measure on Smoking Prevalence in Australia," Report of Dr. Tasneem Chipty, January 24, 2016,

In response to my draft report, the following studies were also made available to me, which I also went through:

Diethelm PA, Farley TM. Re-analysing tobacco industry funded research on the effect of plain packaging on minors in Australia: Same data but different results. *Tob Prev Cessat.* 2017 Nov 16;3:130.

Underwood D, Sun S, Welters RAMHM. The effectiveness of plain packaging in discouraging tobacco consumption in Australia. *Nat Hum Behav.* 2020 Dec;4(12):1273-1284.

Expert report of Casmef and LUISS Business School, LUISS Guido Carli University, Italy, and Deloitte Financial Advisory, Italy., Analysis of the impact of plain packaging on smoking prevalence and tobacco consumption in Australia, 8 November 2019,

I also looked at this study, but not in detail because it does not directly relate to the questions asked to me:

Expert report of Casmef and LUISS Business School, LUISS Guido Carli University, Italy, and Deloitte Financial Advisory, Italy., Analysis of the impact of Plain Packaging on tobacco consumption in the UK and France, 19 July 2021,

I have read through most of the pieces where empirical analysis is presented and referenced, but I have not personally (re)analyzed the data used, although I was informed that I could obtain it. That was also not possible in the time and space allotted to me and, in my opinion, not necessary to reach my conclusions.

Although I have mostly limited myself to an explanation of my answer based on the critical analysis, I have also made unsolicited suggestions for further research or for possible refinement of the existing results in some areas.

In addition, I have also extensively studied the comments of the two parties (van Pels Rijcken (on behalf of the State) and HoganLovells (on behalf of BAT) on my draft report and have made a number of changes as a result. I have also commented on both comments. provided and also attached herewith (see 2 appendices with responses).

I assume that my advisory role has been fulfilled for the time being.

Yours faithfully,

Eddy Van Doorslaer
Professor of health Economics
Erasmus University Rotterdam

Rotterdam, December 9, 2012

Appendices: 1. EVD expert report (13 pages)

2. EVD's response to the State's response (6 pages)

3. EVD response to BAT response (9 pages)

QUESTIONS FROM THE COURT

(a) Are there statistical methods available that can be used to assess the causal effects of standard packaging on actual smoking behavior using real world data?

Answer: Yes, those methods exist and each has its advantages and disadvantages. There are a number of so-called quasi-experimental methods that make it possible to estimate and assess the causal effects of interventions without experimental prospective research.

It also seems almost impossible to set up a prospective randomized study in which some participants are randomly confronted with standard packaging and others are not. That is why many policy analyzes opt for quasi-experimental methods that try to imitate the conditions of an experimental study as closely as possible.

(b) If so, does that include the following static methods?:

(i) A before-after multivariate regression analysis, estimating the effects of standard packaging, taking into account other conditions that may influence smoking behavior, and using data/data from before and after the implementation of standard packaging.

Answer:

Yes, this is one of the methods available when it is not possible to find (or construct) an adequate control group against which to compare the group exposed to the intervention (in this case the standard packs of cigarettes in Australia since 2012).

A distinction must be made between before/after comparisons based on samples and surveys that follow *the same* people longitudinally (so-called panel studies) on the one hand, and studies that use new samples in successive periods on the other. From Professor Viscusi's description I can conclude that the RMSS data are repeated (annual) surveys in representative samples, but not among the same people. Under certain assumptions it is still possible to draw conclusions about eg the effects of the introduction of a change (such as standard packaging).

This assumption is that – after adjustment for possible differences in the composition of the samples drawn – the observed changes over time can still be analyzed as the actual changes in the population (in this case the Australian population). It is then not possible to observe these changes within the same persons because the same persons are only very rarely and by chance observed in more than 1 period. This means that the average smoking behavior before and after is observed, but not any changes in smoking behavior in the same people. The possibility of correcting personal characteristics that are not directly observable but constant over time by means of so-called "person fixed effects" is then lost. This is therefore not possible in the analyzes of successive cross-sections of the population.

The latter would have been possible if (longitudinal) panel data had been collected (ie repeated measurements on the same sample of persons) but this is very difficult (due to dropout) and very expensive to collect and is therefore much rarer. It seems that the latter was not an option because there is not a sufficiently large and long-term panel data collection in the adult population of Australia available that can be used for this purpose.

(ii) **Differences-in-differences analysis,** in which a comparable population not affected by the introduction of standard packaging is identified and in which the smoking behavior of a group subject to standard packaging (the group under study) and a group not is affected by standard packaging (the control group) is followed. The analysis

is made using data/data from before and after the implementation of standard packaging.

Answer: Yes. In principle, this may be the most appropriate method to answer this causal question, provided that an adequate control group can be found and the main identifying condition of a DID analysis is met. This condition is that pre-intervention (sufficient) parallel trends can be observed in both the studied and control group. The unavoidable (and also non-testable) assumption is that, without intervention, these trends would have continued to run parallel in the two groups *after* the intervention date.

(iii) **A forecasting approach,** in which the relevant factors influencing smoking behavior are estimated on the basis of data/information from the period before the implementation of standard packaging, and then based on these estimates, a forecast is made of the extent to which in which people would smoke if standard packaging had not been introduced. The forecast is then compared to the actual smoking rate after the implementation of standard packaging.

Answer: Yes, this method is also an option, but I do not think it is the most suitable for this question. Prediction or prediction is not the same as establishing a causal relationship between an intervention (eg standard packaging) and an outcome (eg smoking behaviour). It is then still possible to predict the expected trajectory after intervention as follows: a model is first estimated that predicts smoking behavior on the basis of a number of predictive characteristics of persons or groups in the period before intervention. After the intervention, this model is used to predict what the expected smoking behavior would be for those characteristics based on the same observed characteristics, but given the estimated relationship before the intervention. The author Dr Dryden himself describes why this method – although it aims at the same thing and arrives at comparable outcomes – is inferior to the DID method (also used by him) if parallel trends can be assumed (common trends assumption).

(c) Should the following factors be considered when applying the statistical methods to assess the effects of standard packaging on smoking behaviour?
isolate:

i. The price of tobacco products, taking into account endogeneity?

Answer: Most definitely. A great deal of economic literature shows that the price of tobacco products is one of the most important determinants of smoking behaviour. Each method should therefore try to correct for this evolution of the price consumers have to pay (including excise duties) as far as possible when studying smoking behavior (both prevalence of smoking behavior and consumption of cigarettes).

An additional problem is that the average prices themselves can be causally influenced by smoking behavior and this gives rise to the problem of so-called endogeneity: one of the explanatory variables (the average cigarette price) is itself influenced by the dependent variable (cigarette consumption) in the fashion model. For example, Dryden's study shows (with a DID analysis) that standard packaging in Australia has reduced the average price of purchased cigarettes because some smokers have partly replaced the consumption of (more expensive) premium brands with (cheaper) non-premium brands. For this reason, the effect of price on smoking behavior can be better investigated via a so-called instrumental variable analysis, eg with an estimation in two steps (two-stage least squares). It means that not the

observed (endogenous) average price, but the *predicted* average price based on so-called (exogenous) instruments is included in the regression equation. This only works if the conditions for valid instruments are met (eg sufficient predictive power for the instrumented variable, but no direct influence on the dependent variable itself).

ii. The pre-existing trend in the analyzed data, including the possible non-linearity of the trend?;

Answer: Indeed, the most data 'fitting' approach to the secular time trend in smoking behavior appears to be crucial in this whole discussion, i.e. the change in smoking behavior over time that is not (fully) explained by the **explanatory** variables that are included in the model. The differences in outcomes between models with a linear trend versus those with a non-linear trend largely explain the differences in outcomes between some of the previous studies (such as Diethelm and Farley, 2015; Chipty 2016) and the studies by Viscusi and Dryden. The latter two authors convincingly demonstrate that a non-linear trend approach leads to model estimates that fit better with the actual observed evolution of smoking behavior in the period Dec 2012-Dec 2016. Viscusi believes that a quadratic trend fits better with his data, Dryden uses the most flexible way of trend estimation -- a series of both monthly and yearly indicator variables -- in its DID approach.

iii. Demographic factors, including age, education level and income?

Answer: Yes. The main reason why it is important to correct for these factors (and some others) is precisely because successive cross-sections of the Australian population are used. If these socio-demographic factors influence smoking behavior (and they do, statistically speaking) and if the sample composition varies according to these characteristics in successive samples, then adjustments should be made to avoid erroneously attributing any observed changes to (for example) policy interventions while they may be related to the aging or wealthier population.

Questions regarding the available data (data sets) and analyzes on the effects of standard packaging

(d) Do the studies on actual smoking behavior discussed in **the Cochrane Review' and the Trimbos Fact Sheet'** provide reliable evidence that standard packaging has led to a reduction in smoking prevalence (number of people who smoke) and tobacco consumption (number of tobacco products consumed) in Australia?

Answer: No. Both the Cochrane review (McNeill et al, 2017) and the Trimbos fact sheet are very cautious in drawing their conclusions regarding the evidence on whether standard packaging in Australia (and elsewhere) has led to a reduction in both the smoking prevalence and tobacco consumption. The (authoritative) Cochrane review (McNeill et al, 2017) -- which reviewed the 5 published studies available to date that had examined smoking reduction in Australia -- qualified all of these studies as "low quality" or "very low quality". It was perhaps for this reason that the conclusion of the review at the time was: *"the limited evidence we have from one study suggests that standardized packaging **can** lead to decreases in smoking prevalence". The ultimate conclusion was therefore: "standard packaging **has the potential** to decrease tobacco use prevalence".* This does not mean that it had actually (in Australia or elsewhere) been sufficiently demonstrated that it brought about a smoke reduction.

The conclusion of the Trimbos Institute Factsheet was also clear about this (p10) *"Several studies have been conducted into the effect on smoking prevalence, but no study can 'separate' the unique effect of the introduction of generic packaging from possible effects of other measures that are taken simultaneously. took place. There is therefore still **insufficient evidence** to state that the introduction of generic packaging has a directly demonstrable effect on the percentage of smokers in the population."*
In other words, these two reviews themselves indicated that at that time no reliable evidence was yet available regarding effects on the ultimate outcome: smoking behaviour.

> (e) Are the studies on non-behavioural impacts of
> standard packaging discussed in the Cochrane Review and/or the Trimbos Fact Sheet is able to demonstrate that standard packaging has reduced smoking prevalence (number of people smoking) and tobacco consumption (number of tobacco products consumed) in Australia, considering that the studies fail to analyze actual smoking behaviour?

Answer: No. These are not studies into the actual smoking or purchasing behavior of (potential) smokers, but into so-called *intermediate* or intermediate variables such as beliefs, knowledge and attitudes of those (potential) smokers with regard to smoking. To deduce from this that the actual smoking behavior is also changed as a result of the standard packagings through these mechanisms, (at least) two steps are necessary: (1) demonstrating that standard packagings have significantly influenced these variables and (2) demonstrating that changes in these intermediate variables in turn influence smoking behaviour.

With regard to (1), both Cochrane and Trimbos conclude that a large number of studies have been conducted and that some of them have found significant effects, but that the link with (2) is missing. If indeed these intermediate steps are essential in smoking decisions, which is the basis of much of this research, then this effect should also be reflected in the ultimate smoking behaviour. No strong evidence has been found for this. The most consistent evidence is that the cigarette packs are - unsurprisingly - rated as less attractive by potential users, but this does not seem to translate into a statistically observable decrease in smoking behaviour.
Viscusi is (re)analyzing some of the Australian data sets (such as the CITTS and NTPPTS) with which some of these studies have been conducted, sometimes with a significant extension of the study period (eg 37 months extra compared to Dunlop et al 2014). He does this for both the effects of smoking behavior and intermediate variables and finds (with a more extensive dataset and in my opinion also better methods) *no* statistically significant effect at all on the number of cigarettes smoked per day with the NTPPTS. Using the CITTS data, he finds mixed results: on the one hand a small but statistically significant increase of 1 cigarette per day, but on the other hand also a strange change in the composition: more daily smokers, fewer weekly smokers and more non-smokers who said the previous year to smoke. The additional studies therefore, in my opinion, do not provide convincing evidence of an influence on even the intermediate variables (which in turn are supposed to influence smoking behaviour).

> (f) Can the data from the Roy Morgan Single Source Survey ("RMSS") dataset from Australia be used to determine the causal effect of the 2012 Packaging Changes in Australia ("2012 Packaging Changes") on smoking prevalence (number of people smokes) using one or more of the statistical methods mentioned above?

Answer: Yes, provided that after adjustment for socio-demographic composition (via inclusion in the regression equation as control variables or co-variates), the annual samples can be regarded as representative of the adult Australian population for that year (and which month).
Changes over time observed in groups of different people can then still be used to estimate changes in those (different, but made comparable) groups.
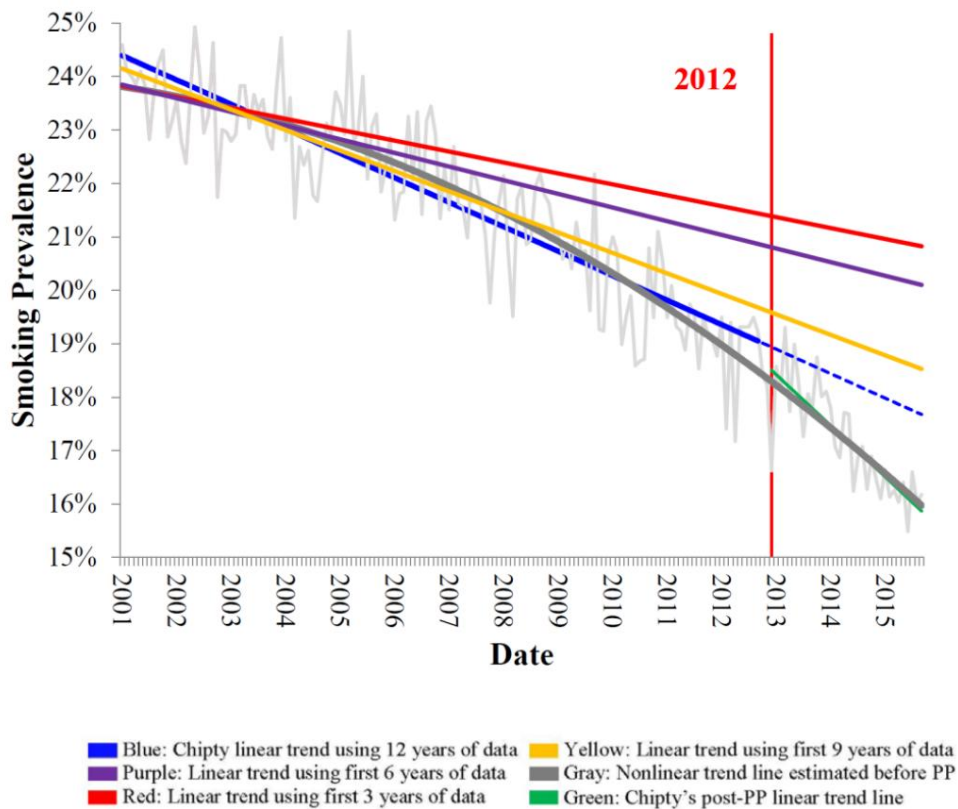
> (g) Is the methodology used by Professor Viscusi in his report to analyze the RMSS dataset a suitable method for elucidating the causal effects of the introduction of the

to isolate and analyze changes in packaging from 2012 in Australia on smoking prevalence (number of people who smoke)?

Answer: Yes, given the absence of a control group AFTER the introduction of standard packaging (henceforth abbreviated as SV) (which were introduced throughout Australia) it is only possible to investigate what the expected development would have been if the trends from *before* the change had continued *after* implementation. This requires a method that attempts to estimate what the most likely development in Australian smoking behavior would have been without that introduction. With a regression equation that explains the probability that someone identifies themselves as a smoker in a survey, it is then estimated what the development in smoking behavior would have been without the intervention, taking into account factors such as (i) other determinants of smoking behavior that have changed over time, such as the average price or excise tax on tobacco, or other measures of smoking control policy, or (ii) changes due to changes due to sampling by adjusting for socio-demographic characteristics of the respondents such as level of education, age, income, place of residence, ed That is what this method aims to do and it is also almost identical to the method used by Dr Chipty in her report of an analysis of the Roy Morgan Single Source Survey (RMSS) data (Chipty, 2016) for the Australian government and included in the Australian Post-Implementation Report of 2016. It then estimated that the "2012 Packaging Changes" had led to a significant drop in smoking prevalence by 0.55 percentage points.

Viscusi's analysis includes both a re-analysis of Chipty's regression results (using the same method, data and period) and an extension of it by adding an additional 15 months of observations (until December 2016). In his analysis, he manages to replicate Chipty's earlier results (for the same period) and then adds some additional analyses, both on the same analysis period and on the extended period. In contrast to Chipty, Viscusi concludes that the standard packaging (SV) introduction in 2012 did **not** lead to a statistically significant change in the percentage of smokers. He shows – in my opinion very convincingly and clearly – that this is due to two striking differences in the method used: (1) the way in which (changes in) the price of cigarettes is taken into account and (2) the way in which the time trend in smoking behavior (the smoking prevalence) is estimated.

Chipty used for (1) a series of so-called indicator variables that indicate the excise duty increases on cigarettes and for (2) a so-called linear time trend. As soon as one of these two is replaced by Viscusi's preference variables, the significant effect disappears. For (1) he uses a *continuous* price variable that indicates the average price of cigarettes (and some variations thereof) taking into account general price inflation and excise tax increases. This is a better way to include price in the analysis that the consumer faces when purchasing than the discrete changes in excise duties. For (2) he shows that a non-linear (quadratic) trend statistically fits the actual observed data much better than a linear trend. I also find his (extensive) justification of those alternative choices (in the model specification) convincing. First, the price of cigarettes is a very important determinant of smoking behavior and its continuous variables approximate this evolution much better than Chipty's tax indicator variables. And secondly, he also shows very clearly graphically that the decreasing time trend in smoking behavior is not linear at all in Australia according to these data. See the figure reproduced below of p32 in Viscusi which shows both (monthly) observed smoking prevalence and estimated time trends.

Blue: Chipty linear trend using 12 years of data — Yellow: Linear trend using first 9 years of data
Purple: Linear trend using first 6 years of data — Gray: Nonlinear trend line estimated before PP
Red: Linear trend using first 3 years of data — Green: Chipty's post-PP linear trend line
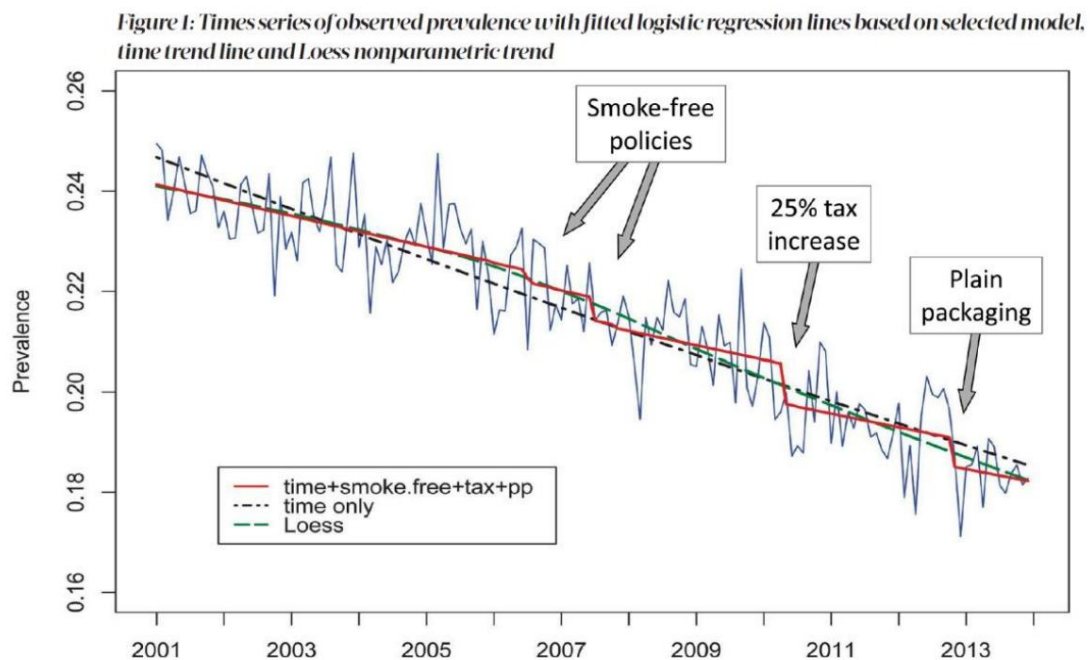
Bron: Viscusi (2018), p. 32

I agree with Viscusi that a better measurement of the existing (secular) trend in smoking prevalence via a non-linear trend is preferable to the linear trend. The non-linear trend is more flexible and fits visually and statistically better with the observed data than the linear trend. Indeed, the estimated effect of Chipty is mainly a result of not adequately estimating this time trend: smoking prevalence was already falling (increasingly) before the introduction of SV, and the decline was not (statistically significant) accelerated by it . The jump and kink from the blue straight line to the green straight line is only caused by the imposition of that linear trend on the data.

Viscusi also demonstrates, and again convincingly in my opinion, that the increasingly strong increase in the average price of cigarettes (partly as a result of the increased excise duties) is partly an explanation for the increasingly rapid decline in smoking in Australia. With an additional so-called instrumental variable (IV) analysis of the effect of price on smoking, he also shows that price does indeed have an effect on smoking if there are doubts about the endogeneity of price with regard to smoking behaviour. This is quite likely if there were reverse causality or simultaneity, as also found in Dryden's analyzes (see later in this report).

In his report, Viscusi also pays attention to the findings of Diethelm and Farley (2015) who analyze the estimated monthly smoking percentages based on the same RMSS data but for a much shorter (1-year) period (until Dec 2013). They do this (like Chipty later) on the basis of a fairly similar regression analysis (but with monthly averages, not individual data) and also conclude that the SV introduction led to a statistically significant reduction in smoking prevalence of 3.5% in Australia led. This analysis is basically inferior to Chipty's for three reasons: (1) the monthly estimates (derived visually from a graph by Kaul and Wolf 2015) were not corrected for the different composition of the (weekly) sampling. Although approximately 4500 respondents are interviewed per month, there can be considerable differences

lead in the composition of that sample to important factors such as age, education, income, place of residence, etc.). (2) the analysis includes 15 post-intervention months less than Chipty and Viscusi, which means that the impact of SV can be less accurately estimated. And above all: (3) here too a linear trend is estimated that is not compared to a non-linear trend. But otherwise it is an interesting precursor to the (improved) Chipty study because this analysis – albeit with linear trends per period – also concludes that three of the four investigated tobacco control measures would have had a statistically significant smoke-reducing effect: smoke-free zones: -4.8%; 25% excise tax increase: -4.5%; and standard packaging -3.7%). Strangely enough, no significant effect was found for the fourth measure (health warnings on the packaging).

Although inferior to Chipty and Viscusi, I include their graph here for didactic reasons.

Figure 1: Times series of observed prevalence with fitted logistic regression lines based on selected model, time trend line and Loess nonparametric trend

Diethelm P.. et al. Tob. Prev. Cessation 2015; 1(November): 6
http://dx.doi.org/10.18332/tpc/60650

Source: Diethelm and Farley (2015), figure 1

This graph also clearly shows that smoking prevalence in Australia is not declining linearly, but the authors nevertheless opt to describe the course over time in a linear manner for each period and - partly because of this - can attribute the decreases to the successive interventions. It goes without saying that this result could be very different if the deviations were measured against a non-linear (eg quadratic) trend. And that is exactly what Viscusi has shown later, with better (because individual) data and over a longer period (2001-Dec 2016). Viscusi also presents a number of robustness analyzes that use different research choices but confirm the main results.

This leads me to conclude that Professor Viscusi's analysis provides the best estimate of the effects of the 2012 SV introduction in Australia known to me at the time of my draft report. As a result of my draft report, I was made aware of some later publications

studies1 such as the study by Underwood et al (2020) (which was not commissioned by industry and was published in Nature Human Behaviour, as well as LUISS and Deloitte studies of 2019 and 2021 which were commissioned by BAT). I will sometimes make use of this in both the replies to the reactions of both parties and in the report below.

I have only two additional comments on this analysis.

1. *Indirect effect via price reduction.* Dr Dryden's analyzes show that the introduction of SV in Australia has also had an impact on average cigarette prices. By correcting for price developments, it is not impossible that this reduces the influence of SV, because corrections have already been made for so-called intermediate variables: for example, if the effect of SV were to manifest itself mainly through a price reduction, which in turn increases smoking prevalence (as Viscusi demonstrates) then perhaps wrongly the (indirect) effect of the SV would no longer be found. I have tried to deduce this from the Table in Viscusi p. 25, but this shows that this may not be the case. In all specifications with or without the price variables, the SV effect is not significant, regardless of the trend (linear or nonlinear). The SV effect is halved, but is not significant in any comparison (with the exception of the Chipty specification with the excise tax indicator variables and the linear trend). This shows that the correct specification of the trend seems much more important than the price-versus-excise adjustment.

2. *Effect of health warnings.* An interesting incidental finding (which is less relevant here) is that both in the analyzes by Diethelm and Farley and by Viscusi, the impact of the introduction of mandatory health warnings on packaging is also estimated to be nil. I found this astonishing. If a smoke-reducing effect can be expected from the appearance of cigarette packaging, then I would expect this much sooner from those (larger) health warnings than (merely) from the standard packaging. But that aside.

(h) Are the results of Professor Viscusi's preference analyzes — that the expected effect of standard packaging in Australia **is indistinguishable statistically from zero** — accurate with a reasonable degree of scientific certainty (in other words, a test for the hypothesis of zero effect using of the conventional value of 5% for type 1 errors)?

Answer: Yes, that is what Viscusi thinks, and as I stated above, his study seems to be generating the most credible results for the time period under review (Dec 2012-Dec 2016) at this point. Viscusi's analysis was repeated in LUISS and Deloitte (2019) for a longer period of 5 years post-intervention (until December 2017) and those estimates also produced a non-significant effect.

It is only a pity that these results themselves have not been published. If that had been the case, these results could also have been included in any updates to reviews such as the Cochrane review. Ultimately, the peer review that accompanies this is also an important touchstone, although I agree with his remark that seldom if ever that peer review, nor of the journals,

---

1 Expert report of Casmef and LUISS Business School, LUISS Guido Carli University, Italy, and Deloitte Financial Advisory, Italy., Analysis of the impact of plain packaging on smoking prevalence and tobacco consumption in Australia, dated 8 November 2019, available at https://businessschool.luiss.it/news/tobacco-consumption-in-australia/ ("LUISS and Deloitte (2019) report").
Underwood D, Sun S, Welters RAMHM. The effectiveness of plain packaging in discouraging tobacco consumption in Australia. Nat Hum Behav. 2020 Dec;4(12):1273-1284. doi: 10.1038/s41562-020-00940-6. Epub 2020 Sep 21. Erratum in: Nat Hum Behav. 2020 Oct 2; PMID: 32958901, available at https://www.nature.com/articles/s41562-020-00940-6
Expert report of Casmef and LUISS Business School, LUISS Guido Carli University, Italy, and Deloitte Financial Advisory, Italy., Analysis of the impact of Plain Packaging on tobacco consumption in the UK and France, dated 19 July 2021, available at https://businessschool.luiss.it/en/be-inspired-blog/analysis-of-the-impact-of-plain-packaging-on-tobacco-consumption in-the-uk-and-france/

nor from a body such as the Cochrane review, gives rise to replication through reanalysis of the results, for example, although it is increasingly the case for the most prominent economic and medical journals that the underlying data must also be made publicly available.

(i)   Can the available ***retail scanner data*** in **Australia** be used to analyze the causal effects of the **2012 Packaging Changes in Australia** on tobacco consumption (number of tobacco products consumed) using one or more of the above statistical methods?

Answer: Yes, that is an opportunity to examine tobacco consumption because they provide the best monthly estimates of smoking behavior through the study of purchasing behavior. Because this does not concern samples, but the actual observed total sales per month, this is a longitudinal analysis of measurements in the same research population (the Australian adult population).

(j) Are the methodologies used by Mr Dryden in his report to analyze retail sales data (retail scanner data) appropriate methods to assess the causal effects of the introduction of the 2012 Packaging Changes in Australia on tobacco consumption (number tobacco products that are consumed) and analyze it?

Answer: Yes, the so-called difference-in-differences method is suitable if it can be made plausible that pre-intervention there were joint ('common') or parallel trends in the two populations to be compared and the evolutions over time are compared after adjustment for expected differences. I have the following comments about this:

1. *Parallel pre-intervention trends.* The most important condition to be fulfilled in order to perform a credible DID is the assumption of parallel trends in the intervention (Australia) and control group (New Zealand). The problem is that this can only be tested prior to the intervention. With the help of a co-integration, Dryden shows that the (rather long) period from 1970 to 2010 does indeed show a reasonably comparable development in the consumption pattern of both countries. But the crucial period for the DID is the 4-year pre-intervention period used in the regression (2009-Nov 2012). The parallel trends in the monthly data are shown in Figure 4. Graphically, this evolution indeed appears to be very parallel, but no separate statistical test has been performed. One possibility is, for example, to test whether the coefficients for the time indicators (monthly or quarterly) do not differ significantly in both countries (taking into account the large difference in total consumption). It would also be more useful to present this graph and test for the dependent variable that is also used in the regression model (i.e. not for total consumption but for per capita consumption). Assuming that the demographic evolution of the adult population (20+) was not so different in the two countries, this should not make too much of a difference. This is crucial for the DID effect estimates.

2. *Price evolution.* Dryden also examines whether SV has had an effect on the average price evolution. No pre-intervention parallel trends are considered for this additional DID analysis. Footnote 43 explains that this was because no long time series is available (such as the 40 years for consumption), but again ignores the fact that the crucial period for the parallel trends is the period Jan 2009-Nov 2012. A graph could be made for this and, if necessary, a statistical test could also be carried out. It is missing in this document and did not allow us to assess the acceptability of that assumption. As a result, the finding of an average price decrease effect of SV based on this DID is less well substantiated than that for an increase in average consumption. However, it is supported by the findings of Underwood et al (2020) who, based on other analyzes of other data, also conclude that as a result of SV, average tobacco consumption has increased while expenditure on tobacco per household has decreased. They conclude that "the substitution effect is the only argument that fits this pattern of findings"

(p1280). They also conclude that this is only possible by replacing more expensive cigarette brands with cheaper ones.

(k) Are the following results of Mr. Dryden's differences-in-differences analyses, as included in his report, accurate with a reasonable degree of scientific certainty:

i. Standard packs are associated with **an increase in per capita cigarette consumption** in Australia when compared to what would have occurred if standard packs had not been introduced (in other words, against the counterfactual analysis).

Answer: Yes, Dryden's DID analysis finds a small but statistically significant increase in tobacco consumption of approximately 2.2 to 3.5% per year in Australia compared to New Zealand. In reality, this is actually a less strong decrease in consumption than in New Zealand, because average consumption in both countries falls during this period. This finding is reliable assuming that the trends in Australia and New Zealand would have continued to run parallel if this change had NOT been introduced. Dryden also provides the necessary supporting analyzes for this (see answer to question j (1)).
His DID model (as shown in comparison in B3 p 78 of the Annex) most flexibly takes into account the observed time trends in both countries through so-called month and year fixed effects.

The significant increase in per capita cigarette consumption is also supported by the later findings of both Underwood et al (2020) (with different data) and LUISS and Deloitte (2019) (with the same data). I will not go into these studies in detail here, but LUISS and Deloitte (2019) use almost the same methods as Dryden (but observe an extra year of data, so until Dec 2017). The study by Underwood et al (2020) also estimates increased consumption, but not with the same method. Although this study has been published in a journal with a very good peer review reputation *(Nature Human Behaviour),* I myself have the following 3 comments: (i) parallel pre-intervention trends are only displayed graphically and do not always look parallel, (ii) tobacco prices are not included in the model and (iii) only linear trends in the annual data are used, something that both Viscusi and Dryden indicate does not fit well with the data.

For these reasons, I think that the Dryden study provides more reliable evidence than Underwood et al (2020), although they both find an increase in cigarette consumption.

ii. Standard packaging is associated with **a reduction in the average retail price** paid by consumers for cigarettes in Australia, relative to the counterfactual analysis.

Answer: Yes, using a similar DID method, Dryden finds that the standard packaging has statistically significantly reduced the average retail price paid for cigarettes in Australia (by 2-2.5%) compared to the development in New Zealand (which serves as a comparative benchmark) . In this analysis, little or no supporting information is presented for parallel pre-intervention trends in average prices in both countries. This would weaken the findings because that assumption is crucial for this method. In a later study by LUISS and Deloitte, this co-integration analysis of the price variables in both countries is performed and the trends appear to be very similar.

In addition, Underwood et al (2020) also argue that only a reduced average sales price for cigarettes is consistent with their findings. They find that the SV has reduced average tobacco spending per household while increasing average consumption of tobacco. This

is only possible if the average price paid for tobacco has fallen. They too – like Dryden – interpret this finding as evidence for the substitution of more expensive cigarette brands by cheaper brands.

iii. The introduction of standard packaging is accompanied by **a statistically significant acceleration of a so-called "down-trading" trend in Australia,** with consumers shifting from premium brands to cheaper, non-premium brands.

Answer: Yes, this is indeed the conclusion of this analysis. A DID analysis is also used for this in comparison with New Zealand like the previous two. So the downtrading is examined by comparing the price evolution per brand between the two countries. The method tests whether the price decrease found could have been a result of *down-trading* or the replacement of premium by non-premium brands. Indeed, it is difficult to see how SV could lead to an increase in consumption unless an average price decrease has followed, which in turn could be a consequence of the switch from smoking from more expensive to cheaper brands.

The more recently published study by Underwood et al (2020) also finds – albeit with different (annual) data and in my opinion inferior methods (see my comment above) – that the introduction of SV in Australia has led to increased average tobacco consumption but decreased average expenses per household. According to them, this can only be explained by the substitution of more expensive brands of cigarettes by cheaper brands.

(l) Are the results of Mr Dryden's alternative analyzes of data from Australia alone — that standard packaging is associated with an **increase** in cigarette consumption as contained in his report - accurate with a reasonable degree of scientific certainty.

Answer: Yes. These 'alternative' analyzes are accurate, but they are only intended to complement the previously presented DID analyses, which Druden calls his preferred analyses. They have been added mainly to enable some comparison with the methods used in Chipty's study and broadly confirm the own DID findings, but not those of Chipty. Dryden himself clearly and correctly indicates why all these methods are inferior to the DID analysis and should therefore be given less weight.

(m) Do the empirical analyzes of real world data from Australia presented by Professor Viscusi and Mr Dryden provide more reliable evidence/analysis of the impact of Australia's 2012 Packaging Changes on actual smoking behavior than the analyzes and studies of actual smoking behavior (smoking prevalence: number of people who smoke and/or tobacco consumption: number of tobacco products consumed) analyzed in the Cochrane Review and/or the Trimbos factsheet (taking into account that in the studies of Professor Viscusi and the Mr Dryden more recent data and a larger dataset has been analyzed and that more extensive statistical methods have been used to isolate the impact of Australia's 2012 Packaging Changes)?

Answer: Yes, I think so, for the reasons I have already explained above. In scientific research we usually speak of stronger evidence rather than of more reliable evidence. The results presented in these two studies use better methods as well as more extensive or different data and therefore – unless evidence to the contrary – represent the best available evidence. Four years of data before and 4 years of data after the intervention seem to be the best basis for comparison to investigate the effect of SV 2012.

A question that should always be asked in the case of non-significant results is whether the analysis had sufficient statistical power to demonstrate that relationship. Again, the Viscusi and Dryden studies, when they appeared, were the studies that analyzed the most extensive data set, and therefore also had to be able to determine the statistical significance of relatively smaller changes.

(than previously conducted studies such as those by Chipty). In other words, they had greater statistical power to test even small differences.

In addition, some of these analyzes have now apparently been repeated (in LUISS and Deloitte, 2019) for an even longer post-intervention period (of 5 years) and these confirm the earlier results. This also reinforces confidence in the earlier findings.

I understand that similar studies have now been carried out with a comparable DID method for other (European) countries that have introduced the standard packaging (in LUISS and Deloitte, 2021), but it seems unlikely that they will have a greater static capacity than the Viscusi study because standard packaging was introduced there more recently. I was given this study, but I did not go through it in detail because I was not asked to do so directly.

(n) Do the empirical analyzes of real world data from Australia presented by Professor Viscusi and Mr Dryden provide more reliable evidence of the impact of Australia's 2012 Packaging Changes on actual smoking behavior (smoking prevalence: number of people smoking and/ or tobacco consumption: number of tobacco products consumed), than the evidence from the surveys and experimental studies that do not analyze the actual smoking behavior of the population and which have been analyzed in the Cochrane Review and/or the Trimbos fact sheet?

Answer: Yes. The evidence provided in the Viscusi and Dryden studies provides more credible evidence of the impact of the 2012 Packaging Changes in Australia on smoking behavior and tobacco consumption than the studies included in the Cochrane and Trimbos reviews. Standard packaging has not demonstrably led to an accelerated decrease in smoking prevalence in Australia (than in New Zealand), but it has led to increased tobacco consumption as a result of the average price decrease. Dryden shows that the latter is a consequence of the substitution of more expensive brands by cheaper brands.
Two additional comments on these studies:

1. *Mediation.* The results of research into effects on so-called intermediate variables (such as attitudes, beliefs, quit calls, etc.) are all useful in explaining the mechanism if a decrease had been observed or demonstrated. They could then contribute to an answer to questions such as how smoking decisions were changed and whether, for example, there were fewer starters or more quitters. But if no overall effect has been established, there is also less point in further investigating the mechanisms that may underlie this.
In the absence of a clear link between possibly changed knowledge, beliefs or attitudes and behavioral change, this becomes somewhat less relevant.

2. *Opposite heterogeneous effects.* In the draft report I indicated that it might be interesting to look for what we call heterogeneous effects. Perhaps the smoking behavior of some subgroups has been positively influenced by SV 2012 and that of other groups negatively. It could then be that positive and negative effects cancel each other out, so that no effects are observed at all. This could be investigated with a more specific analysis per age group. This may provide additional information because most smokers start at a young age (16-22 years) while quitters are more spread over the entire life cycle. It is now clear that this research has already been carried out in the LUISS and Deloitte (2019) study: in appendix A2.1 and Table 10 they present an analysis for under 18 year olds and find – as for the entire population – no effects of SV on the smoking prevalence of this subgroup.

(o) Taking into account the above factors, the expert believes that the empirical analyzes of real world data from Australia as presented by Professor Viscusi and Mr
Dryden provide more reliable evidence to analyze whether standard packaging would reduce actual smoking behavior in the Netherlands (smoking prevalence: number of people

smoking and/or tobacco consumption: number of tobacco products consumed) than the Cochrane Review and/or the Trimbos factsheet?

Answer:

Yes. In any case, the two studies mentioned use the best available data and the most appropriate methods to investigate this. Therefore, in my opinion, they do indeed currently provide the best available evidence on the impact of Australia's 2012 Packaging Changes on smoking prevalence and consumption.

Meanwhile - in the response to my draft report - I was made aware of the fact that LUISS and Deloitte (2019), in an assignment for BAT - have redone the analyzes of both Viscusi and Dryden using the same methods but with an extra year of data added (so until December 2017). These analyzes do not change the evidence much and confirm the previous results: SV has not had a statistically significant effect on smoking prevalence but has increased average tobacco consumption in Australia.

Similar analyzes are not (yet) available for the Netherlands and will not be possible in the very near future because the standard packaging in the Netherlands was only introduced in 2020 and it is still too early to investigate the effect, if the data is already available. would be. To the extent that the Australian findings are transferable to the Dutch context (different context, different period), in my opinion they indeed currently constitute the best available evidence on which to base policy.

Meanwhile, in the response to the draft report, I was made aware of the fact that a new study is also available, also by LUISS and Deloitte (2021) and also commissioned by BAT, which has also used the DID method applied by Dryden to evaluate the introduction of standard packaging (SV) for France (from 1 January 2017) and the United Kingdom (from 20 May 2017). The researchers use a similar DID regression approach to compare France to Italy and the UK to Germany. They find that SV in France led to a 5% increase in cigarette consumption compared to Italy up to October 2020. (see section 9.3 and Table 5 of the LUISS and Deloitte (2021) report). No effect is found in the DID regression for the comparison of the evolution in the UK with Germany. The methodology of both DID studies is almost identical to the methods previously used by Dryden and also by LUISS and Deloitte (2019) in Dryden's reanalysis. Since the detailed examination of these newer studies was not among the questions I was originally asked to comment on, I will refrain from further comment and limit myself further to my conclusions about the studies about which I have been asked questions.

Because the studies by Viscusi (2018) and Dryden (2017) were not yet available – they are still not published or otherwise available in the public domain – the Cochrane review and Trimbos fact sheet could not yet include and evaluate them. It should be noted that also the (authoritative)
Cochrane review (McNeill et al, 2017) qualified all 5 published studies available until then that had investigated smoking reduction in Australia as "low quality" or "very low quality". It was perhaps for this reason that the conclusion of the review at the time was that *"the limited evidence we have from one study suggests that standardized packaging **can** lead to decreases in smoking prevalence" and the ultimate conclusion was that "standard packaging **has the potential** to decrease tobacco use prevalence".*
That is not to say that it had actually been sufficiently demonstrated (in Australia or elsewhere) that it brought about a smoke reduction.
The conclusion of the Trimbos Institute Factsheet was also clear (p10) *"Some studies have been done into the effect on smoking prevalence, but no study can 'separate' the unique effect of the introduction of generic packaging from possible effects of other measures that took place simultaneously. . There is therefore still **insufficient evidence** to state that the introduction of generic packaging has a directly demonstrable effect on the percentage of smokers in the population."*

Thus, these two conclusions (both Trimbos and Cochrane) do **not** conflict with the conclusions of Viscusi or Dryden. They do conflict with Chipty's conclusion for the PIR of Australia.