WT/DS435/R/Suppl.1, WT/DS441/R/Suppl.1
WT/DS458/R/Suppl.1, WT/DS467/R/Suppl.1

28 June 2018

(18-4052)



Original: English

# AUSTRALIA – CERTAIN MEASURES CONCERNING TRADEMARKS, GEOGRAPHICAL INDICATIONS AND OTHER PLAIN PACKAGING REQUIREMENTS APPLICABLE TO TOBACCO PRODUCTS AND PACKAGING

REPORTS OF THE PANELS

*Appendices*

*SCI redacted, as indicated [[***]]*

This supplement contains Appendices A to E to the Reports of the Panels to be found in document WT/DS435/R, WT/DS441/R, WT/DS458/R, WT/DS467/R.

_____

**Table of Contents**

## LIST OF FIGURES

## APPENDIX A:

## POST-IMPLEMENTATION EVIDENCE ON NON-BEHAVIOURAL OUTCOMES OF THE TPP MEASURES

1. In these proceedings, the parties submitted as exhibits a number of peer-reviewed studies investigating the post-implementation impact of the TPP measures and enlarged GHWs on non-behavioural proximal outcomes, namely: (i) reduction in the appeal of tobacco products; (ii) increased effectiveness of GHWs; and (iii) reduction in the ability of the pack to mislead consumers about smoking harms.[1] This Appendix reviews and discusses this evidence, in the light of the relevant expert reports submitted by the parties.

2. Australia submits that the available post-implementation empirical studies on non-behavioural outcomes confirm that TPP and enlarged GHWs have (i) reduced appeal, (ii) increased the effectiveness of GHWs, and (iii) reduced the ability of packaging to mislead consumers about the harmful effects of tobacco products.[2]

3. Based on the review of these peer-reviewed papers, and in some cases the re-analysis of the data used in these papers, the Dominican Republic, Honduras and Indonesia argue that the TPP measures have not had the expected effects on the antecedents of behaviour posited in Australia's conceptual framework of the TPP measures.[3] In particular, they contend that beyond the obvious findings that the pack is less visually appealing and people more often notice the larger GHWs first, empirical evidence shows little or no effects of the policies on the antecedents of behaviour. They further claim that the variables relating to beliefs, attitudes and intentions towards smoking were almost entirely unaffected by the TPP measures.[4]

4. In addition, the experts of the Dominican Republic and Indonesia argue that some of the published empirical studies on Australia's TPP measures provide an inaccurate picture of the empirical evidence. They state that some of these papers failed to report the results for more than half of all the variables available in the survey dataset, which were overwhelmingly not statistically significant, suggesting no impact by plain packaging on these variables. The Dominican Republic and Indonesia contend that the authors of some of these published studies also failed to explain that a number of the reported statistically significant effects had vanished by the end of the first year of Australia's implementation of the TPP measures as a result of wear-out effects. The Dominican Republic and Indonesia further criticize these papers for failing to report the magnitude of the statistically significant effects. According to their experts, most of the reported statistically significant effects are small, suggesting that the TPP measures have little importance in shifting behaviour.[5]

5. Each peer-reviewed empirical paper discussed by the parties either addresses different questions or is based on specific survey data, or both. We discuss, for each proximal outcome, each survey data and corresponding paper separately, before turning to an overall assessment. We approach this assessment on the basis that our task is not to conduct our own econometric assessment of the TPP measures' impact on the proximal outcomes identified above but rather to

[1] Some of these papers also analyse more distant variables, such as beliefs, attitudes and intentions towards smoking and quitting as well as quitting attempts. These papers are also discussed in Appendix B.

[2] See Australia's first written submission, paras. 201-205; and comments on the complainants' responses to Panel question No. 146, para. 7.

[3] See Ajzen et al. Data Report, (Exhibit DOM/IDN-2); Ajzen et al. Second Data Report, (Exhibit DOM/IDN-4); Ajzen et al. Data Rebuttal Report, (Exhibit DOM/IDN-6); Ajzen et al. Second Data Rebuttal Report, (Exhibit DOM/IDN-8); Klick Supplemental Rebuttal Report, (Exhibit HND-122); and Klick Second Supplemental Rebuttal Report, (Exhibit HND-165).

[4] See Ajzen et al. Data Report, (Exhibit DOM/IDN-2), paras. 10-22.

[5] See Ajzen et al. Data Report, (Exhibit DOM/IDN-2), paras. 23-27.

examine, on the basis of the evidence before us, the overall robustness of the econometric evidence submitted by the parties in this respect.[6]

## 1 EVIDENCE RELATING TO THE APPEAL OF TOBACCO PRODUCTS SINCE THE ENTRY INTO FORCE OF THE TPP MEASURES

6. Four peer-reviewed papers have analysed empirically the impact of Australia's TPP measures on the appeal of tobacco products: (i) Wakefield et al. 2015; (ii) Dunlop et al. 2014; (iii) White et al. 2015a; and (iv) Miller et al. 2015.[7]

7. Different survey data were used by several of these published papers. Most survey data cover adult smokers, with the exception of a survey of students attending secondary schools.[8] Most of these peer-reviewed papers analyse only cigarette smokers, although some of these survey datasets also include information on cigar smokers. Only one peer-reviewed study analyses appeal-related outcomes in relation to cigar and cigarillo smokers.[9]

### 1.1 Datasets and related studies

#### 1.1.1 National Tobacco Plain Packaging Tracking Survey

8. In order to track the effect of the TPP measures, Australia's Department of Health and Ageing funded the National Tobacco Plain Packaging Tracking Survey (NTPPTS), a nationwide tracking survey conducted by the Cancer Council Victoria (CCV). The NTPPTS is a continuous cross-sectional baseline survey of about 100 interviews per week of current smokers and recent quitters aged 18 to 69 years, conducted from 9 April 2012 to 30 March 2014. A follow-up survey of baseline participants then took place approximately four weeks after the initial survey, with follow-up surveys conducted from 7 May 2012 to 4 May 2014.[10] The NTPPTS data have been used in several peer-reviewed papers published in the supplement to the *Tobacco Control* journal in 2015.[11]

9. Wakefield et al. 2015 use the NTPPTS data to investigate among adult smokers the impact of Australia's TPP measures on its three specific mechanisms, namely (i) the appeal of tobacco products, (ii) the effectiveness of GHWs, and (iii) the ability of packaging to mislead about smoking harms.[12] The authors estimate a logistic model, using baseline survey weights and controlling for individual characteristics, such as sex, age, highest educational attainment, nicotine dependence and socio-economic status.

10. Overall, Wakefield et al. 2015 conclude that Australia's TPP measures have statistically reduced the appeal of tobacco products for adult cigarette smokers. This statistically significant effect was sustained up to 12 months after implementation. In particular, the authors report a statistically significant increase in the proportion of adult smokers that disliked their pack, perceived lower pack appeal, lower cigarette quality, lower satisfaction and lower value, and disagreed that brands differed in prestige. However, the authors find that there was no statistically significant change in the proportion of adult smokers that perceived differences in the taste of different brands.

11. The Dominican Republic and Indonesia submitted an expert report by Ajzen et al.[13], also relied upon by Honduras[14], which reviews the accuracy and completeness of the findings reported

[6] For a similar approach, see Panel Reports, *US – COOL (Article 21.5 – Canada and Mexico)*, para. 7.183 (citing Panel Reports, *US – COOL*, para. 7.539).

[7] See Wakefield et al. 2015, (Exhibits AUS-206, DOM-306); Dunlop et al. 2014, (Exhibits AUS-207, HND-132, DOM-199); White et al. 2015a, (Exhibits AUS-186, DOM-235); Miller et al. 2015, (Exhibits AUS-102, DOM-315).

[8] See White et al. 2015a, (Exhibits AUS-186, DOM-235); and White et al. 2015b, (Exhibits HND-135, DOM-236, DOM-288).

[9] See Miller et al. 2015, (Exhibits AUS-102, DOM-315).

[10] See Dominican Republic's second written submission, paras. 380-429.

[11] See Ajzen et al. Data Report, (Exhibit DOM/IDN-2), para. 6.

[12] See Wakefield et al. 2015, (Exhibits AUS-206, DOM-306).

[13] Ajzen et al. Data Report, (Exhibit DOM/IDN-2).

in Wakefield et al. 2015. They reconsider the NTPPTS dataset and present the results of a logistic model for dichotomized outcome variables, a linear model for continuous outcome variables, and an ordered logit model for categorical outcome variables, using baseline survey weights and controlling for individual characteristics.

12. The Dominican Republic argues that the published studies greatly underreport the results of the NTPPTS survey, with a pattern of underreporting unfavourable results.[15] Ajzen et al. argue that Wakefield et al. 2015 failed to report the results of three other appeal-related outcome variables that were not statistically significant. According to Ajzen et al., the authors also failed to address the small magnitude of the observed statistically significant effects.[16] More generally, the Dominican Republic considered Australia's argument that the NTPPTS was much broader in scope than Wakefield et al. 2015 to be surprising, as this message was not expressed in Wakefield et al. 2015 or in the journal's editorial and the papers were presented as the first comprehensive evaluation of the TPP measures.[17]

13. Ajzen et al. conclude that the TPP measures have had very little impact on the mechanisms posited to underlie change in smoking behaviours because the most notable changes were only in pack appeal, with the effect on pack dislike moderate to strong, but the impact on product dislike, brand loyalty and identification was much smaller or statistically not significant. In particular, the authors find that the increase in the proportion of adult smokers that perceived lower pack appeal was statistically moderate to large but with partial evidence of wear-out effect. Ajzen et al. also report that the effects on perceived lower cigarette quality and satisfaction were statistically positive but small and without any wear-out effect. Similarly, they find that the effects on perceived lower value and brands' prestige were statistically very small but without any wear-out effect. They further confirm Wakefield et al.'s 2015 finding that there was no statistically significant change in the proportion of adult smokers that perceived differences in the taste of different brands. They reach a similar conclusion for three other variables, not reported by Wakefield et al. 2015, namely whether the smoker would stay loyal to a regular brand if the store ran out of it; agreed that they felt connected to others smoking their regular brand, and (very) often noticed others with their regular brand in the past month.[18]

14. Honduras submitted an expert report by Professor Klick, which examines the NTPPTS data on the effect of the TPP measures on the appeal of smoking.[19] Professor Klick submits that there are a host of items relating to the appeal of smoking that appear to have worsened following the introduction of the TPP measures, but which have been ignored by Wakefield et al. 2015.[20]

15. Professor Klick presents the results of an ordered probit model, which controls for the TPP measures (early TPP period and formal TPP period), gender, age, education, socio-economic status and a linear time trend.[21] He finds that the TPP measures are not associated with a decrease in the reported frequency of thoughts about enjoying smoking.[22]

16. According to Australia, the survey data are most suited to assessing changes in the specific mechanisms of the TPP measures. In that context, Australia contends that the positive findings reported by Ajzen et al. are completely consistent with Wakefield et al. 2015's findings, i.e. that

[14] See Honduras's second written submission, paras. 171-185. We note that Cuba states, "Cuba did not submit [this] expert report prepared by Professor Ajzen, and therefore no assumptions can be made as to Cuba's position concerning the effectiveness of plain packaging on the basis of that report." Cuba's response to Panel question No. 146.

[15] Dominican Republic's comments on Australia's response to Panel question No. 198, paras. 700-709.

[16] See Ajzen et al. Data Report, (Exhibit DOM/IDN-2), paras. 98-101.

[17] Dominican Republic's comments on Australia's response to Panel question No. 198, para. 705.

[18] See Ajzen et al. Data Report, (Exhibit DOM/IDN-2), paras. 89-97, 148-150, Appendix A, pp. 78-80.

[19] See Klick Supplemental Rebuttal Report, (Exhibit HND-122), paras. 54-73.

[20] See Klick Supplemental Rebuttal Report, (Exhibit HND-122), paras. 61-63.

[21] Professor Klick explains that, unlike Wakefield et al. 2015, he omits the measure of exposure to mass media anti-smoking messages, cigarette costliness and heaviness of smoking index, because these variables are endogenous. He further explains that the inclusion of these variables does not change the results he presents. See Klick Supplemental Rebuttal Report, (Exhibit HND-122), fn 35.

[22] See Klick Supplemental Rebuttal Report, (Exhibit HND-122), paras. 66-68. We note that Professor Klick considers the questions "how important is quitting to the person" and "how frequently people think about quitting" relevant to the appeal of tobacco. The results of these questions are discussed in Appendix B.

Australia's TPP measures reduce the appeal of tobacco products.[23] Australia further argues that because the scope of the NTPPTS was much broader than the specific and limited focus of Wakefield et al. 2015, the results of the study in question were reported appropriately and consistently. Australia also contends that the complainants' implication that unfavourable results were not reported is directly contradicted by the facts.[24]

### 1.1.2 Cancer Institute New South Wales Tracking Survey

17. The Cancer Institute New South Wales Tobacco Tracking Survey (CITTS) is a weekly tracking telephone survey of smokers and recent quitters (who quit in the past 12 months) involving approximately 50 interviews conducted per week throughout the year. The survey monitors smoking-related thoughts and behaviours among adult smokers and recent quitters in New South Wales.

18. Dunlop et al. 2014 uses the CITTS data to investigate the impact of Australia's TPP measures on two of the specific mechanisms: (1) decreasing the promotional appeal of packaging and (2) increasing the impact of health warnings.[25] The analysis covers 15,375 randomly selected adult smokers between April 2006 and May 2013 (i.e. six months after the introduction of the TPP measures). Adjusting for background trends, seasonality, antismoking advertising activity and cigarette costliness, the authors estimate autoregressive integrated moving average (ARIMA) models.

19. Overall, Dunlop et al. 2014 conclude that Australia's TPP measures have had an early statistically significant effect in reducing the promotional appeal of the packaging among adult smokers.[26] In particular, the authors find a significant increase in the proportion of adult smokers strongly disagreeing that the look of their cigarette pack is attractive, says something good about them, influences the brand they buy, makes their pack stand out, is fashionable and matches their style. According to Dunlop et al. 2014, changes in these appeal-related outcomes were maintained six months following the TPP measures' implementation.

### 1.1.3 Australian Secondary Students Alcohol Smoking and Drug Survey

20. The 2013 Australian Secondary Students Alcohol Smoking and Drug (ASSAD) survey extension is a follow-up survey of students attending secondary schools that participated in the 2011 ASSAD survey in Victoria and Queensland. In total 82 schools across both states participated. The 2013 survey extension was designed to compare attitudes to cigarette packaging before and after the introduction of Australia's TPP measures, and included questions about beliefs and attitudes about cigarette packaging, ratings of popular cigarette brands, noticing health warnings on cigarette packs, awareness of the specific harms of tobacco use, and perceptions of the prevalence of smoking and intention to smoke.[27]

21. White et al. 2015a use the responses from the ASSAD survey extension to analyse, among other things, the impact of plain packaging on the appeal of cigarette packs and brands among students aged 12-17 years.[28] The authors estimate generalized linear regression models and multinomial logistic regression models controlling for smoking status, age, sex, school education sector and state.

22. Overall, White et al. 2015a conclude that the TPP measures have reduced the appeal of cigarette packs among adolescents. In particular, they found a statistically significant decrease in the proportion of students, who had seen a cigarette pack in the previous six months, that rated positively the brand character and cigarette pack. The authors note that the effect on brand character and cigarette pack was even larger among younger smokers. White et al. 2015a also

[23] Australia's comments on the complainants' responses to Panel question No. 146, para. 7.
[24] Australia's response to Panel question No. 198, paras. 298-303.
[25] See Dunlop et al. 2014, (Exhibits AUS-207, HND-132, DOM-199).
[26] See Tobacco Plain Packaging PIR, (Exhibit AUS-624), paras. 77-80.
[27] See, e.g. Dominican Republic's second written submission, paras. 443-456.
[28] See White et al. 2015a, (Exhibits AUS-186, DOM-235).

report a statistically significant increase in the proportion of students, who had seen a cigarette pack in the previous six months, rating negatively cigarette packs.[29]

23. The Dominican Republic and Honduras argue that, in the absence of full access to the ASSAD survey data it requested, it is impossible to make an objective assessment of the findings in White et al. 2015a. The Dominican Republic submits that as its analysis of the NTPPTS data showed, the published results in White et al. 2015a may provide an unduly positive and inaccurate impression of the full dataset. The Dominican Republic argues that although the study concludes that there is a change in the visual appeal of packs, that reduced appeal does not have any meaningful impact on perceptions of the harmfulness of smoking, quit intentions and secondary indicators of quitting.[30] Professor Ajzen, in an expert report submitted by the Dominican Republic, Honduras and Indonesia, submits that the impact of Australia's TPP measures on visual appeal of tobacco products reported in White et al. 2015a and defined as "modest" by their authors makes it difficult for the change in appeal to carry all the way through the posited chain of effects to behaviour.[31]

### 1.1.4 Cigar and cigarillo smokers surveys

24. Miller et al. 2015 conducted individual interviews with ten regular premium cigar smokers, as well as two focus groups with premium cigarillo smokers and occasional premium cigar smokers (with a total of 14 smokers), and four focus groups with non-premium cigarillo smokers (with a total of 28 smokers), in February and March 2014, 15 months after the TPP measures became mandatory. In addition, in February and March 2014 the authors conducted an online survey of current cigar and cigarillo smokers. Of the original 56,589 people contacted, only 268 met all inclusion criteria, one of these criteria being current smokers of these products.[32] Rather than undertaking an econometric analysis, the authors present descriptive statistics of the results of these interviews, focus groups and the online survey.

25. According to Miller et al. 2015, there was incomplete exposure to the TPP measures on cigar and cigarillo smokers during the first 15 months following their implementation, with many premium cigar smokers purchasing fully branded cigars in boxes duty free or online and singles in non-compliant packaging. Reported exposure was seemingly highest among non-premium cigarillo smokers. However, the authors note that when exposure occurred, the TPP measures reduced perceived packaging appeal. In particular, they find that although changes in perceived taste, harm and value were minimal for experienced premium cigar smokers, they indicated some fear of being equated with cigarette smokers. Miller et al. 2015 also find that occasional premium cigar and premium cigarillo smokers with higher plain packaging exposure (gained by purchasing boxes rather than singles cigars) perceived that cigar/package appeal and value had declined. Similarly, they report that more non-premium cigarillo smokers affirmed that the perceived appeal, quality, taste, enjoyment and value had decreased. They also find that online survey participants stated that packaging appeal had diminished since the implementation of the TPP measures.[33]

26. Ajzen et al., in their expert report submitted by the Dominican Republic and Indonesia, contend that the results reported by Miller et al. 2015 suffers from several methodological shortcomings, such as the non-representativeness of survey participants, unsuitability of focus groups and interviews to draw causal inferences, failure to control for exposure to anti-smoking campaigns and changes in tobacco prices, and absence of "baseline" information collected before the implementation of the TPP measures. According to Ajzen et al., these methodological shortcomings severely limit any conclusions that can be drawn from the results.[34]

27. Ajzen et al. claim that even if Miller et al.'s 2015 results were taken at face value, the study revealed a few notable effects of the TPP measures that were clearly to be expected and unlikely to influence actual smoking behaviour, namely the changes in the appeal of the package,

[29] See White et al. 2015a, (Exhibits AUS-186, DOM-235).

[30] See Dominican Republic's second written submission, paras. 443-456; and comments on Australia's response to Panel question No. 196, para. 390.

[31] See Ajzen Report, (Exhibit DOM/HND/IDN-3), paras. 174-178.

[32] See Miller et al. 2015, (Exhibits AUS-102, DOM-315).

[33] See Miller et al. 2015, (Exhibits AUS-102, DOM-315).

[34] See Dominican Republic's second written submission, paras. 457-462. See Ajzen et al. Data Report, (Exhibit DOM/IDN-2), paras. 251-262.

increased noticeability of the health warnings, and perceived less value for money. However, according to Ajzen et al., the fact that tobacco products were seen as less value for money may have been due to the rise in their cost as a result of the December 2013 tax increase and other factors. In addition, they argue that Miller et al. 2015 failed to mention that the participants also reported no change with respect to smoking enjoyment. More generally, Ajzen et al. contend that smoking cigarettes and smoking cigars (or cigarillos) are different behaviours, and their determinants are also likely to differ. They conclude that consequently whatever effects the TPP measures are, or are not, found to have on cigarette-related cognitions and behaviours, such findings cannot be generalized to cigars.[35]

### 1.2 Analysis by the Panel

28. We note that among the peer-reviewed papers discussed by the parties, four studies analyse the impact of the TPP measures, as applied together with enlarged GHWs, on the appeal of tobacco products.

29. A careful review of Wakefield et al. 2015 and Dunlop et al. 2014 and the econometric evidence submitted by the Dominican Republic, Honduras and Indonesia leads us to conclude that there is some empirical evidence suggesting that the TPP measures have reduced the appeal of tobacco products among adult cigarette smokers, in terms of pack dislike, product dislike, perceived lower quality, satisfaction and value, lower brands' prestige, and connection and identification.[36]

30. We further note that the Dominican Republic and Indonesia qualify the findings reported in Wakefield et al. 2015 by highlighting that for most of the appeal-related outcomes, although statistically significant, the impact is small, or very small in the case of perceived lower value and brand prestige, but without any evidence of wear-out effects. Only the impact on perceived lower pack appeal, which is found to be statistically moderate, shows some partial evidence of wear-out effect.[37] We also note that Ajzen et al.'s claim that Wakefield et al. 2015 did not report the results of three other outcome variables[38], which were not statistically significant, is actually only valid for the variable related to brand loyalty[39] and to a lesser extent to the variable related to brand connection. In fact, Ajzen et al. also find in the ordered logistic regression of the variable on brand connection a statistically significant increase in the proportion of adult smokers that disagree they feel connected to others smoking their regular brand. Similarly, Ajzen et al. report a decrease, albeit small, in the proportion of adult smokers that (very) often noticed others with their regular brand in the past month, which is statistically significant at 10% level in the logistic regression or 5% level in the linear regression and ordered logistic regression.[40]

31. We note that the empirical evidence available to us on the impact of the TPP measures on the perception of tobacco appeal among adolescents is limited to one peer-reviewed study by White et al. 2015a. In particular, White et al. 2015a suggest that the reduction in the appeal of cigarette packs and brands to adolescents, though modest, was statistically significant seven to 12 months after the introduction of the TPP measures. This result is consistent with the findings reported in Wakefield et al. 2015 and Dunlop et al. 2014.

---

[35] See Ajzen et al. Data Report, (Exhibit DOM/IDN-2), paras. 251-262.

[36] See Wakefield et al. 2015, (Exhibits AUS-206, DOM-306); Dunlop et al. 2014, (Exhibits AUS-207, HND-132, DOM-199); and Ajzen et al. Data Report, (Exhibit DOM/IDN-2), paras. 89-97, 148-150, Appendix A, pp. 78-80. We note that in his review of Dunlop et al.'s 2014 analysis, Professor Klick did not discuss and re-analyse the questions of the CITTS related to appeal. We also note that Professor Klick did not mention in his reports whether the commissioned Roy Morgan Research Survey also asked questions related to the appeal of tobacco products.

[37] See Ajzen et al. Data Report, (Exhibit DOM/IDN-2), paras. 89-97, 148-150, Appendix A, pp. 78-80.

[38] We are also not persuaded that the variable "thought about enjoying smoking" referred to by Professor Klick is directly relevant to assess the impact of the appeal of tobacco products. See para. 15 above. In fact, we note that this variable was discussed by Ajzen et al. Data Report, (Exhibit DOM/IDN-2) in the context of the balance between smoking enjoyment and concern. See Appendix B.

[39] We note that Ajzen et al. did not discuss the fact that in some specifications, such as the logistic regression of the brand loyalty variable, only three explanatory variables are statistically significant (besides the constant). See Ajzen et al. Data Report, (Exhibit DOM/IDN-2), back-up material.

[40] See Ajzen et al. Data Report, (Exhibit DOM/IDN-2), Appendix A, pp. 78-80.

32. Similarly, we note that specific empirical evidence before us on the impact of the TPP measures on appeal of cigars and cigarillos is limited to the peer-reviewed paper by Miller et al. 2015, in which a descriptive statistics analysis finds that occasional premium cigar and premium cigarillo smokers with higher TPP exposure, non-premium cigarillo smokers, and online survey participants reported reduced perceived appeal since the implementation of the TPP measures.[41] We note that many of the criticisms raised by Ajzen et al. are actually acknowledged by Miller et al. 2015. Specifically, the authors recognize that the primary limitations of the study are the representativeness of the samples and the accuracy of self-report measures, most notably recall. In the absence of other relevant data or study on cigars, it is however unclear to what extent the results would have changed if Miller et al. 2015 had explicitly accounted for exposure to anti-smoking campaigns and changes in tobacco prices or applied a different methodology. In that context, we note that although Ajzen et al. submit that the conclusions about the impact of the TPP measures on cigarettes cannot be generalized to cigars, they also recognize themselves that the impact of the TPP measures on appeal reported in Miller et al. 2015 was to be expected. In fact, we note that Miller et al.'s findings are consistent with the findings published in the peer-reviewed studies on adult cigarette smokers and adolescents reviewed above. We therefore see no basis to reject in its entirety Miller et al.'s study on the basis of Ajzen et al.'s criticism of it.[42]

## 2 EVIDENCE RELATING TO THE EFFECTIVENESS OF GHWS SINCE THE ENTRY INTO FORCE OF THE TPP MEASURES

33. Based on different datasets, five peer-reviewed papers have empirically investigated the impact of Australia's TPP measures on the effectiveness of GHWs: (i) Wakefield et al. 2015; (ii) Yong et al. 2015; (iii) Dunlop et al. 2014; (iv) White et al. 2015b; and (v) Miller et al. 2015.[43] An expert report prepared by Professor Klick and submitted by Ukraine also contains an analysis of the impact of the TPP measures on the effectiveness of GHWs.[44]

### 2.1 Datasets and related studies

#### 2.1.1 National Tobacco Plain Packaging Tracking Survey

34. In addition to analysing the impact on tobacco products appeal, Wakefield et al. 2015 use the NTPPTS data to investigate the impact of Australia's TPP measures on GHW effectiveness among adult smokers.[45] The authors estimate a logistic model, using baseline survey weights and controlling for individual characteristics.

35. Overall, Wakefield et al. 2015 conclude that Australia's TPP measures have had a statistically significant effect among adult cigarette smokers, generally sustained for up to 12 months after implementation, on increasing health warning effectiveness. In particular, the authors report a statistically significant increase in the proportion of adult smokers who noticed GHWs, attributed more motivation to quitting to GHWs, avoided specific GHWs when purchasing and covered their pack. However, they find that there was no statistically significant change in the proportion of adult smokers who perceived exaggeration of harms.[46]

36. Ajzen et al., in their expert report submitted by the Dominican Republic and Indonesia, argue that Wakefield et al. 2015's conclusion that the TPP measures are associated with consistent improvement in health warning effectiveness outcomes cannot withstand careful examination.

[41] See Miller et al. 2015, (Exhibits AUS-102, DOM-315).

[42] See Ajzen et al. Data Report, (Exhibit DOM/IDN-2), paras. 255-256.

[43] See Wakefield et al. 2015, (Exhibits AUS-206, DOM-306); Yong et al. 2015, (Exhibit DOM-382); Dunlop et al. 2014, (Exhibits AUS-207, HND-132, DOM-199); White et al. 2015b, (Exhibits HND-135, DOM-236, DOM-288); Miller et al. 2015, (Exhibits AUS-102, DOM-315).

[44] See section 1.6.6 in the main body of these Reports for a description of Ukraine's participation in these proceedings.

[45] See para. 8 above for a description of the NTPPTS data. See Wakefield et al. 2015, (Exhibits AUS-206, DOM-306).

[46] As discussed in Appendix B, we also note that Durkin et al. 2015 use the NTPPTS data to analyse the impact of the TPP measures on quitting-related variables and report greater increases in pack concealment and stubbing out cigarettes because of thoughts about the harm of smoking. See Appendix B and Durkin et al. 2015, (Exhibits AUS-215 (revised), DOM-305).

According to Ajzen et al., Wakefield et al. 2015 underreported the results of seven questions related to knowledge about diseases that were not statistically significant, did not pay attention to the magnitude of the statistically significant effects and failed to mention vanishing effects of some of these small effects.[47] The Dominican Republic considered Australia's claim that the NTPPTS was much broader in scope than Wakefield et al. 2015 to be surprising, as this message was not expressed in Wakefield et al. 2015 or in the journal's editorial in which the study was published.[48] Ajzen et al. reconsider the NTPPTS data and present the results of a logistic model for dichotomized outcome variables, linear model for continuous outcome variables and ordered logit model for categorical outcome variables, using baseline survey weights and controlling for individual characteristics.

37. Ajzen et al. conclude that the TPP measures have had a statistically moderate effect on the attention paid to the enlarged GHWs, but the impact of the TPP measures on concealing packs and requesting packs with different GHWs was statistically small. They also find that the increase in the proportion of adult smokers that attributed more motivation to quitting to GHWs was statistically small and subject to a wear-out effect. However, Ajzen et al. find that the TPP measures had no impact on most beliefs about the health risks of smoking, many of which were not published in Wakefield et al. 2015. In particular, Ajzen et al. report a statistically significant small increase in the proportion of adult smokers that freely recalled a disease on a current GHW and agreed that smoking causes blindness. For the remaining outcome variables related to beliefs about the health risks of smoking, Ajzen et al. find that there is no statistically significant change in the proportion of adult smokers that perceived exaggeration of harms and that agreed that there are diseases caused by smoking, that smoking causes harm to unborn babies, that lung cancer is an old age disease, and that smoking causes stroke, mouth cancer, bladder cancer, and gangrene.[49]

38. Similarly, Ajzen et al. argue that Wakefield et al. 2015 failed to report the results on the balance between smoking enjoyment and concern. In their view, this is all the more surprising given that, in Brennan et al. 2015, the same six authors use the NTPPTS data and hypothesize that the balance between smoking enjoyment and concern is "influenced by cigarette appeal, graphic health warning [GHW] effectiveness and perceived harm".[50] Ajzen et al. find no statistically significant change in the proportion of adult smokers that thought about enjoyment of smoking several/many times in the past month, were very/extremely concerned that smoking may affect health, and experienced more concern than enjoyment from/of smoking.[51] Overall, Ajzen et al. contend that the weakening of the policy across the chain of effects does not reflect a limitation of the NTPPTS dataset, as argued by Professor Chaloupka, but the theoretically expected decline in the impact of TPP across the chain of effects from pack appeal to behaviour.[52]

39. Professor Klick, in his expert report submitted by Honduras, reviews the NTPPTS data relating to the effect of the TPP measures on the effectiveness of GHWs.[53] Professor Klick contends that a large number of outcomes in the NTPPTS data, not reported by Wakefield et al. 2015, show that the TPP measures have not improved the effectiveness of GHWs.[54]

40. Professor Klick presents the results of an ordered probit model and a logistic model, which control for the TPP measures (early TPP period and formal TPP period), gender, age, education, socio-economic status and a linear time trend.[55] He finds that the TPP measures had no statistically significant impact on the concerns about the effect of smoking on health but had a

[47] See Ajzen et al. Data Report, (Exhibit DOM/IDN-2), paras. 115-122.
[48] Dominican Republic's comments on Australia's response to Panel question No. 198, para. 705.
[49] See Ajzen et al. Data Report, (Exhibit DOM/IDN-2), paras. 102-114, 151-155, Appendix A, pp. 81-83.
[50] See para. 8 above for a description of the NTPPTS data. Ajzen et al. Data Report, (Exhibit DOM/IDN-2), paras. 138-142.
[51] See Ajzen et al. Data Report, (Exhibit DOM/IDN-2), paras. 132-137, 159-162, Appendix A, pp. 87-89.
[52] Ajzen et al. Data Rebuttal Report, (Exhibit DOM/IDN-6), para. 11 and Table 1.
[53] See Klick Supplemental Rebuttal Report, (Exhibit HND-122), paras. 74-81.
[54] See Klick Supplemental Rebuttal Report, (Exhibit HND-122), paras. 74-75.
[55] Professor Klick explains that unlike Wakefield et al. 2015 he omits the measure of exposure to mass media anti-smoking messages, cigarette costliness and heaviness of smoking index, because these variables are endogenous. He further explains that the inclusion of these variables does not change the results he presented. See Klick Supplemental Rebuttal Report, (Exhibit HND-122), fn 35.

negative and statistically significant in respondents' awareness of the causal relationship between smoking and mouth cancer.[56]

41. Professor Chaloupka, in an expert report submitted by Australia, contends that the pattern of results reported by Ajzen et al. is consistent with the strengths and limitations of the NTPPTS data. Australia further claims that the results of Wakefield et al. 2015, whose focus is more specific and limited than the much broader scope of the NTPPTS data, were reported appropriately and consistently. According to Australia the complainants' implication that unfavourable results were not reported is directly contradicted by the facts.[57] In particular, Professor Chaloupka argues that, as expected, Ajzen et al. found consistent and statistically significant effects for the impact of the TPP measures on the most proximal outcomes, such as noticing and avoiding health warnings, but the impact is smaller, less statistically significant and less consistent as the focus shifts to less proximal outcomes, such as health knowledge and perceptions of the health risks, when looking at the impact in the overall sample of smokers and recent quitters.[58] Professor Chaloupka further contends that the NTPPTS data cannot be used to assess the impact of TPP on the population most likely to be influenced by the measure, namely, never users who might have taken up tobacco use in the absence of TPP. According to Professor Chaloupka, the NTPPTS data can also not measure the impact of TPP on relapse among former smokers who are not categorised as "recent quitters".[59]

### 2.1.2 International Tobacco Control Policy Evaluation Project

42. The International Tobacco Control (ITC) Policy Evaluation Project (ITC Project) is a longitudinal cohort survey to assess the impact, and identify the determinants of, effective tobacco control policies in more than 20 countries, including Australia. The ITC Project covers a number of issues related to GHWs, including attention to health warnings, salience of health warnings and the effect of health warnings on consumers' thoughts, behaviours and intentions to quitting. The wave survey prior to the implementation of the TPP measures was conducted between September 2011 and February 2012, while the wave after the implementation was conducted between February and May 2013.

43. Yong et al. 2015 use data from the Australian component of the ITC Project to assess the impact of the TPP measures' GHW effectiveness. The authors estimate various generalised estimating equation (GEE) models controlling for age, sex, income, education, cigarettes per day, past year quit attempts, survey mode (phone vs. web) and wave of recruitment.[60]

44. Yong et al. 2015 find a statistically significant increase in the proportion of adult smokers who increased their attentional orientation towards health warning labels, noticed them more, experienced cognitive reactions with respect to health warning labels and avoided health warning labels. However, the authors report no statistically significant change in the proportion of adult smokers who read health warning labels and forego cigarettes. Yong et al. 2015 further find that the subgroup of respondents that switched from initially focusing away to focusing on the health warning labels following the introduction of the TPP measures, noticed and read the health warning labels more, thought more about the harmful effects of smoking and avoided health warning labels, but did not forego cigarettes. Conversely, Yong et al. 2015 show that the subgroup of respondents who chose to focus away from the health warning labels, noticed them less, experienced less cognitive reactions, and avoided the health warning labels less.[61]

45. Ajzen et al., in their expert report submitted by the Dominican Republic and Indonesia, claim that they were unable to present a comprehensive assessment of the ITC dataset, including

[56] See Klick Supplemental Rebuttal Report, (Exhibit HND-122), paras. 74-78 and 80-81. We note that Professor Klick considers the question "how often the person stub out due to thoughts about the harm of smoking" to be relevant to the effectiveness of GHWs. The results of this question are discussed in Appendix B.

[57] Australia's response to Panel question No. 198, paras. 298-303.

[58] See Chaloupka Rebuttal Report, (Exhibit AUS-582), para. 9; and Chaloupka Third Rebuttal Report, (Exhibit AUS-604), paras. 2-7.

[59] Chaloupka Rebuttal Report, (Exhibit AUS-582), para.3.

[60] See Yong et al. 2015, (Exhibit DOM-382).

[61] Yong et al. 2015 (Exhibit DOM-382) note that the mechanism for this apparent reactance, i.e. the minority of smokers stimulated to shift from initially focusing on to focusing away from the warnings and reporting a reduction in avoidance, is not clear.

correcting for the possibility of false positive findings due to multiple hypothesis testing (i.e. statistically significant results might have occurred by chance), controlling for anti-smoking advertising in mass media and testing for wear-out effects, because they only had access to a small subset of the data.[62] Similarly, Ajzen et al. criticise Yong et al. 2015 for failing to justify their selection of outcome variables available in the ITC dataset and having chosen not to report the results of downstream variables and actual smoking behaviour and other questions related to warning labels. According to Ajzen et al., Yong et al. 2015 also failed to report effect sizes, explore wear-out effects and correct for multiple hypothesis testing.[63] Ajzen et al. replicate the analysis of Yong et al. 2015 by re-estimating GEE models controlling for the survey mode (phone vs. web) and wave of recruitment, as well as respondents' age, sex, income, education, cigarettes per day consumed and past year quit attempts.

46. Overall, Ajzen et al. conclude that the results based on the ITC data are similar to those based on the NTPPTS dataset, namely that the TPP measures have had a mixed and overall weak impact on GHW effectiveness. In particular, Ajzen et al. find that, although the TPP measures have had a moderate positive and statistically significant effect on increasing smokers' attentional orientation towards health warning labels and a statistically significant small positive effect on the noticeability of the enlarged GHWs, smokers did not actually read the GHWs more. Similarly, they report that the TPP measures have had a statistically significant small positive impact on cognitive reactions and a statistically significant moderate positive impact on avoiding health warning labels.[64]

47. Australia contends that Ajzen et al. correctly conclude that the TPP measures have significantly increased attention paid to GHWs, noticeability of GHWs, cognitive reactions to GHWs, and avoidance of GHWs.[65] Professor Chaloupka further submits that Ajzen et al. fail to recognize that the impact of plain packaging should be smaller for the less proximal outcomes, such as knowledge about the health consequences of tobacco use, when one looks at the impact in the overall sample of smokers and recent quitters, because one would not expect that a smoker whose likelihood of noticing health warnings did not increase following plain packaging would show any change in his/her knowledge about the health consequences of smoking.[66]

48. Ajzen et al. argue that Australia and its experts do not contest the accuracy of the analytical approach they adopted or the results they obtained and do not challenge their serious criticisms of Yong et al. 2015.[67] They further contend that Professor Chaloupka's assertion that the more distal outcomes will be less affected by the policy than the most proximal outcomes is unfounded. In their view, this weakening of the policy across the chain of effect does not reflect a limitation of the data, but the theoretically expected decline in the impact of plain packaging across the chain of effect from pack appeal to behaviour.[68]

### 2.1.3 Cancer Institute New South Wales Tobacco Tracking Survey

49. As well as analysing the impact of the TPP measures on the appeal of tobacco products, Dunlop et al. 2014 use the CITTS data to investigate their impact on GHW effectiveness among a large group of randomly selected adult smokers between April 2006 and May 2013.[69] The authors present the results of ARIMA models controlling for background trends, seasonality, anti-smoking advertising activity and cigarette costliness.

50. Overall, Dunlop et al. 2014 conclude that Australia's TPP measures have had an early statistically significant effect on increasing effectiveness of health warnings among adult

[62] See Dominican Republic's comments on Australia's response to Panel question No. 196, paras. 409-414; and Ajzen et al. Second Data Report, (Exhibit DOM/IDN-4), paras. 12-19.

[63] See Ajzen et al. Second Data Report, (Exhibit DOM/IDN-4), paras. 30-37 and 53-58.

[64] See Ajzen et al. Second Data Report, (Exhibit DOM/IDN-4), paras. 22-27. We note that Ajzen et al. use the terms "intermediate", "moderate", and "medium" interchangeably to interpret effect sizes of their estimates. See, e.g. Ajzen et al. Data Report, (Exhibit DOM/IDN-2), paras. 90 and 119, and Appendix B, pp. 100-102.

[65] See Australia's response to Panel question No. 196, paras. 225 and 237.

[66] See Chaloupka Rebuttal Report, (Exhibit AUS-582), para. 4.

[67] See Ajzen et al. Second Data Rebuttal Report, (Exhibit DOM/IDN-8), paras. 13-16.

[68] See Ajzen et al. Data Rebuttal Report, (Exhibit DOM/IDN-6), paras. 39-45.

[69] See para. 17 above for a description of the CITTS data. See Dunlop et al. 2014, (Exhibits AUS-207, HND-132, DOM-199).

smokers.[70] In particular, the authors report that two to three months following the introduction of the TPP measures, the absolute proportion of adult smokers having strong cognitive ("graphic warnings encourage me to stop smoking"), emotional ("with the graphic warnings, each time I get a cigarette out I worry that I should not be smoking") and avoidant ("they make me feel that I should hide or cover my packet from the view of others") responses to on-pack GHWs has increased. They find, however, that the impact of the TPP measures on smokers' responses to the salience of GHWs ("the only thing I notice on my cigarette pack is the graphic warnings") was positive but statistically not significant.

51. Professor Klick, in an expert report submitted by Honduras, re-examines the impact of the TPP measures on GHW effectiveness using more recent CITTS data. Unlike Dunlop et al. 2014, Professor Klick restricts the sample to be more evenly spread between the pre- and post-implementation periods, namely from January 2009 to December 2014. In addition, rather than looking only at "strong agreement" with a statement, as Dunlop et al. 2014 did, Professor Klick also analyses "strong disagreement" with a statement. Professor Klick presents the results of a logistic regression controlling for annual time trend, week of survey and individual characteristics, and submits that comparable results may be found using the ARIMA models.

52. Professor Klick finds that the TPP measures have had a positive and statistically significant impact on adult smokers' cognitive, emotional, avoidant and salience responses when the outcome variables are defined as "strong agreement" with a given statement. However, when the outcome variables are defined as "strong disagreement" with a given statement, the results are reversed and the impact of the TPP measures on adult smokers' cognitive, emotional, avoidant and salience responses is found to be negative and statistically significant. Professor Klick submits that this situation of contradictory outcomes likely results from consumer perceptions and intentions being ill-conceived, in a context where there is no cost to providing inaccurate or even inconsistent answers.[71]

53. Professor Klick further submits that Dunlop et al. 2014 have cherry-picked questions to analyse GHW effectiveness and failed to examine other equally relevant questions, whose results contradict the authors' conclusions. In particular, Professor Klick finds that the TPP measures had a positive and statistically significant impact on the probability that the respondent strongly agrees with the statements "I do not look at the warnings each time the smoker gets a cigarette out" and "the graphic warnings are exaggerated". In addition, Professor Klick finds a negative but not statistically significant impact of the TPP measures on the probability that the respondent strongly agrees with the statement "I have read the detailed information on the warning labels".

54. Australia's expert Professor Chaloupka submits that the CITTS dataset has features, including many of the same features of the NTPPTS data, which limit its utility in fully assessing the impact of plain packaging on the proximal and distal outcomes that are likely to be affected by the TPP measures.[72] Professor Chaloupka claims that the CITTS dataset does not include young people and never smokers and, therefore, cannot assess the impact of the TPP measures on the population most likely to be influenced by the measure. Professor Chaloupka further submits that any analysis that uses responses to questions, which are only asked of smokers, is likely to be considerably understating the effects of the TPP measures because recent quitters, who seem most likely to have responded positively to these questions and to have already been influenced by the TPP measures, are not included in the analysis. Professor Chaloupka contends also that Professor Klick's analyses fail to appropriately account for significant changes in the CITTS's sampling methods implemented in 2013, resulting in an increase in the share of the sample accounted for by younger people and by men.

### 2.1.4 Australian Secondary Students Alcohol Smoking and Drug Survey[73]

55. In another paper, White et al. 2015b use the 2013 ASSAD survey dataset to examine the impact of Australia's TPP measures on the effectiveness of GHWs among students aged

[70] See Tobacco Plain Packaging PIR, (Exhibit AUS-624), paras. 77-80.
[71] See Klick Second Supplemental Rebuttal Report, (Exhibit HND-165), paras. 35-58.
[72] See Chaloupka Second Rebuttal Report, (Exhibit AUS-590), paras. 24-30.
[73] See Dominican Republic's second written submission, paras. 443-456.

12-17 years in Victoria and Queensland.[74] The authors estimate linear and logistic regression models controlling for individual characteristics, such as age, sex, school type, state and smoking status.

56. Overall, White et al. 2015b conclude that Australia's TPP measures have increased awareness among adolescents of bladder cancer, blindness and smoking as leading cause of death. However, the authors find that the TPP measures had no impact on adolescents' other health beliefs and cognitive processing of warning information. In particular, they find no change over time in responses to the statements that smoking is addictive, is toxic (from tobacco smoke), clogs arteries, harms unborn babies, increases the risk of having a heart attack, doubles the risk of stroke, and causes mouth cancer, diseases of the gums, kidney disease, lung cancer, emphysema, and diseases in toes and fingers (gangrene).[75] Similarly, the authors report that there is no statistically significant change in the proportion of adolescents that read GHWs, paid close attention to GHWs, thought about GHWs, and talked about GHWs.[76]

57. The Dominican Republic argues that without having full access to the data, it is impossible to make an objective assessment of the findings reported in White et al. 2015b. The Dominican Republic submits that, as the analysis of the NTPPTS data showed, the published results in White et al. 2015b may provide an unduly positive and inaccurate impression of the full dataset.[77] Professor Ajzen, in an expert report submitted by the Dominican Republic, Honduras and Indonesia, further submits that the change in knowledge about bladder cancer could very well be attributable to a confounding factor, namely new information contained in new health warnings.[78] More generally, the Dominican Republic notes that White et al. 2015b conclude that a period of one year following the implementation of the TPP measures was not too short to detect effects on adolescents. The Dominican Republic also submit that White et al. 2015b recognize that a "process of habituation" means that these initially weak effects are likely to weaken further over time.[79]

### 2.1.5 Cigar and cigarillo smokers surveys

58. As part of their analysis of cigar and cigarillo smokers, Miller et al. 2015 review the responses to questions related to GHW effectiveness obtained from individual interviews, focus groups and an online survey conducted 15 months after the introduction of the TPP measures.[80]

59. Miller et al. 2015 find that exposure of cigar and cigarillo smokers to the TPP measures was incomplete during the first 15 months following the implementation of the TPP measures because they purchased fully branded cigars in boxes duty free or online and singles in non-compliant packaging. They note, however, that when exposure occurred, the TPP measures increased the noticeability of health warnings. In particular, the authors report that premium cigar smokers who were exposed had noticed and were concerned by the health warnings, tried to avoid them and felt more like "dirty smokers". Similarly, they find that occasional premium cigar and premium cigarillo smokers with higher plain packaging exposure (gained by purchasing boxes rather than singles) and online survey participants reported having noticed GHWs more. They also note that non-premium cigarillo smokers reported high plain packaging exposure, somewhat increased perceived harm, as well as greater noticeability of GHWs and concealment of packs.

60. Ajzen et al., in their expert report submitted by the Dominican Republic and Indonesia, claim that the results reported by Miller et al. 2015 suffer from several serious methodological shortcomings, including the non-representativeness of survey participants, impossibility of drawing causal inferences, failure to control for confounding factors, and absence of information collected

[74] See White et al. 2015b, (Exhibits HND-135, DOM-236, DOM-288). See para. 20 above for a description of the ASSAD survey.

[75] We note that White et al. 2015b report a statistically significant at 10% level increase in responses to the statement that smoking causes mouth cancer.

[76] See White et al. 2015b, (Exhibits HND-135, DOM-236, DOM-288).

[77] See Dominican Republic's second written submission, paras. 443-456; and comments on Australia's response to Panel question No. 196, para. 390.

[78] See Ajzen Report, (Exhibit DOM/HND/IDN-3), para. 178.

[79] See Dominican Republic's second written submission, paras. 451-455.

[80] See para. 24 above for a description of the cigar and cigarillo smokers surveys data. See Miller et al. 2015, (Exhibits AUS-102, DOM-315).

during the pre-implementation period of the TPP measures, which severely limits any conclusions that can be drawn from the analysis.[81]

61. According to Ajzen et al., even if the results published by Miller et al. 2015 are taken at face value, the few notable effects of the TPP measures revealed in the paper, including the increased noticeability of health warnings, are clearly to be expected and unlikely to influence actual smoking behaviour. Ajzen et al. further claim that Miller et al. 2015 did not mention that the participants also reported no change with respect to concerns that smoking may damage health, stubbing out cigars or cigarillos and stopping smoking. More generally, Ajzen et al. contend that whatever effects the TPP measures are, or are not, found to have on cigarette-related cognitions and behaviours, such findings cannot be generalized to cigars. This, they submit, is because smoking cigarettes and cigars (or cigarillos) are different behaviours, and their determinants are also likely to differ.[82]

### 2.1.6 Commissioned Roy Morgan Research Survey (Australia and New Zealand)

62. Roy Morgan Research dataset, commissioned by Professor Klick, is a survey of individuals who were current or former (in the past 12 months) smokers in both Australia and New Zealand undertaken using random digit dialling sampling techniques. The first wave of the survey was completed prior to the implementation of Australia's TPP measures between 2 November 2012 and 26 November 2012 in Australia and between 8 November 2012 and 1 December 2012 in New Zealand. Subsequent waves were carried out at three-month intervals up until February 2014. The survey probed the respondents' experience with attempts to quit and their intention to do so in the future as well as other attempts to change some aspect of their smoking behaviour.[83]

63. Professor Klick uses the Roy Morgan Research data to estimate a difference-in-difference logit model that explains non-behavioural outcomes related to GHWs in Australia and New Zealand, controlling for the TPP measures, an Australian baseline variable, a common baseline variable and a post-TPP implementation time-period.

64. Overall, Professor Klick concludes that the TPP measures have had no statistically significant effect on smokers' beliefs about the health effects of smoking and the degree to which they notice warnings on their cigarette packages in Australia relative to New Zealand.[84] In particular, Professor Klick finds there was no statistically significant TPP impact in Australia relative to New Zealand on the likelihood that the respondent answered positively on whether smoking is a major issue, a minor or major issue, and harmful to the heart, stomach, mouth, bladder, throat, sight, skin, or teeth. In addition, Professor Klick finds that the TPP measures have had no statistically significant impact in Australia relative to New Zealand on placing a cover over the cigarettes, placing the cigarettes in a different container and keeping the pack out of sight.[85]

65. Although Australia's expert, Dr Chipty, does not address directly Professor Klick's analysis of non-behavioural outcomes based on the Roy Morgan Research data, a number of criticisms raised by Dr Chipty regarding Professor Klick's difference-in-difference analysis of smoking incidence also apply to the analysis of non-behavioural outcomes. Dr Chipty argues that Professor Klick's analysis of the Roy Morgan Research data is invalid because Professor Klick's commissioned Roy Morgan survey does not contain a pre-period and is incapable of distinguishing which respondents had noticed plain packaging. Furthermore, Dr Chipty considers that New Zealand is an invalid counterfactual for the purposes of studying the effects of plain packaging, because of New Zealand's January 2013 excise tax increase, one month after the introduction of TPP measures.[86]

---

[81] See Dominican Republic's second written submission, paras. 457-462; and Ajzen et al. Data Report, (Exhibit DOM/IDN-2), paras. 251-262.

[82] See Ajzen et al. Data Report, (Exhibit DOM/IDN-2), paras. 251-262.

[83] See Klick Report, (Exhibit UKR-5), pp. 6-8. See section 1.6.6 in the main body of these Reports for a description of Ukraine's participation in these proceedings.

[84] See Klick Report, (Exhibit UKR-5), pp. 18, 47-54.

[85] See Klick Report, (Exhibit UKR-5), pp. 18, 54-56.

[86] See Chipty Rebuttal Report, (Exhibit AUS-535) (SCI), paras. 54-71.

### 2.2 Analysis by the Panel

66. We note that a slightly larger number of peer-reviewed studies before us (six in total) analyse the impact of the TPP measures and enlarged GHWs on outcome variables related to GHW effectiveness, compared to the number of studies before us addressing the impact on appeal of tobacco products.

67. A careful review of Wakefield et al. 2015 and Yong et al. 2015 and the econometric evidence submitted by the Dominican Republic, Honduras and Indonesia leads us to conclude that there is some empirical evidence that suggests that the TPP measures have improved the GHW effectiveness among adult cigarette smokers by increasing the noticeability of GHWs, attention towards them, avoidance of health warnings labels, pack concealment, request for a pack with a different GHW and attribution to the motivation to quit to GHWs (cognitive responses).[87] We note that although Ajzen et al. qualify these impacts as small or moderate in the case of attention and avoidant responses, they are still statistically significant, with no evidence of wear-out effects, except in the case of the cognitive responses.[88]

68. However, we note that the empirical evidence on the impact of the TPP measures on knowledge about health risks is more nuanced and statistically not significant as regards the balance between smoking enjoyment and concern. Ajzen et al. and Professor Klick contend that the impact of the TPP measures on GHW effectiveness is mixed because the impacts on reading the GHW, perceived exaggeration of harms and knowledge about specific risks are not statistically significant.[89] We note in this respect that while the analysis suggests that the TPP measures seem to have a limited impact on recalling specific smoking risks, the results also suggest that the TPP measures have had a small but positive and statistically significant impact on recalling a disease on a current GHW and on believing that smoking causes blindness without evidence of a wear-out effect.[90] In that context, we find that Ajzen et al.'s claim that Wakefield et al. 2015 underreported non-statistically significant results should be qualified.[91] First, we note that Wakefield et al. 2015 did not report some results. However, following their own analysis of the unreported results with respect to questions assessing whether the TPP measures and enlarged GHWs increased knowledge about diseases caused by smoking, Ajzen et al. acknowledge that the TPP measures had a "very small effect" on respondents' ability to freely recall a disease on a current GHW and their belief that smoking causes blindness.[92] Second, Ajzen et al. also find an increase, albeit very small, in the proportion of adult smokers who believe smoking causes bladder cancer, which is statistically significant at the 10% level in the logistic regression. Similarly, Ajzen et al. report in the linear regression and ordered logistic regression a very small but statistically significant (at the 10% level) increase in the proportion of adult smokers who disagree that lung cancer is a disease smokers only get in old age.[93] Third, and more generally, we agree with Professor Chaloupka that the impact of the TPP measures is likely to be smaller for the less proximal outcomes, when looking at the impact in an overall survey sample composed of smokers and recent quitters.[94]

69. A careful review of the analysis of the CITTS data reported by Dunlop et al. 2014 and Professor Klick confirms the findings reported in Wakefield et al. 2015 and Yong et al. 2015 that the TPP measures had, in the two to three months following their introduction, a positive and

[87] See Wakefield et al. 2015, (Exhibits AUS-206, DOM-306); Dunlop et al. 2014, (Exhibits AUS-207, HND-132, DOM-199); Yong et al. 2015, (Exhibit DOM-382); and Ajzen et al. Data Report, (Exhibit DOM/IDN-2), paras. 89-97, 148-150, Appendix A, pp. 78-80. We also note that in his review of Dunlop et al. 2014's analysis, Professor Klick did not discuss and re-analyse the questions of the CITTS related to appeal. We also note that Professor Klick did not mention in his reports whether the commissioned Roy Morgan Research survey also asked questions related to the appeal of tobacco products.

[88] See Ajzen et al. Data Report, (Exhibit DOM/IDN-2), paras. 102-114, 151-155, Appendix A, pp. 81-83; and Ajzen et al. Second Data Report, (Exhibit DOM/IDN-4), paras. 22-27.

[89] See Ajzen et al. Data Report, (Exhibit DOM/IDN-2), paras. 102-114, 151-155, Appendix A, pp. 81-83. See also Ajzen et al. Second Data Report, (Exhibit DOM/IDN 4), para. 22.

[90] See Ajzen et al. Data Report, (Exhibit DOM/IDN-2), Appendix A, pp. 81-83.

[91] See Ajzen et al. Data Report, (Exhibit DOM/IDN-2), para. 116.

[92] See Ajzen et al. Data Report, (Exhibit DOM/IDN-2), paras. 112 and 116-117.

[93] See Ajzen et al. Data Report, (Exhibit DOM/IDN-2), pp. 81-83. We also note that Ajzen et al. report in the logistic regression results table a very small but statistically significant at the 10% level decrease in the proportion of adult smokers that agree that smoking causes mouth cancer.

[94] See Chaloupka Rebuttal Report, (Exhibit AUS-582), para. 9.

statistically significant impact on adult smokers' cognitive, emotional and avoidant reactions to GHWs.[95] We are not persuaded that the econometric results presented by Professor Klick can be taken at face value and provide sufficient basis to dismiss Dunlop et al.'s results. A comparison between Professor Klick's re-analysis with Dunlop et al. 2014 is not straightforward, not only because the sample period is different but more importantly because the estimation method is different. Professor Klick submits that using Dunlop et al. 2014 would yield similar results.[96] Yet, Professor Klick reports a statistically significant impact of the TPP measures on smokers' responses to GHW salience, while in Dunlop et al. 2015 the impact is found to be not significant.[97] It is therefore unclear to what extent Professor Klick's findings are the result of his model specification (i.e. not controlling for exposure to advertising and tobacco costliness) or a different sample period or both. It is also unclear why Professor Klick decided to discard data from April 2006 to December 2008. Likewise, it is unclear if Professor Klick's findings are affected by the changes in the survey's sampling methods. Professor Klick also did not provide an explanation as to why the TPP measures would decrease GHW effectiveness, as suggested by some of his counter-intuitive results.

70. As noted above in our review of the post-implementation studies before us on appeal, only one peer-reviewed study by White et al. 2015b reports empirical evidence of the impact of the TPP measures on adolescents. In the context of GHW effectiveness, we note that White et al. 2015b suggest that one year after their implementation, the TPP measures have had a limited impact on adolescents' beliefs about the health risks of smoking and no impact on their cognitive processing of the GHW information.[98] The authors report that acknowledgement of negative health effects of smoking among Australian adolescents remains high. This could explain why the TPP measures did not increase adolescents' health beliefs, except as regards bladder cancer, mouth cancer, blindness and smoking as a leading cause of death.

71. Post-implementation empirical evidence on cigar and cigarillo smokers is also limited to a single peer-reviewed paper, by Miller et al. 2015, who present a descriptive statistics analysis of various personal interviews, focus groups and an online survey. The authors find that 15 months after the introduction of the TPP measures, cigar and cigarillo smokers exposed to plain packaging reported greater noticeability of the GHWs and in a few cases greater concerns about the health warnings, avoidance of graphic health labels and pack concealment.[99] As pointed out by Ajzen et al., the evidence on cigar and cigarillo smokers' health beliefs and cognitive responses are more mixed. As explained in our review of the post-implementation studies on appeal, many of the criticisms raised by Ajzen et al. regarding Miller et al.'s methodology, such as the representativeness of the samples and accuracy of self-report measures, have been identified as limitations by the authors themselves. It is however unclear to what extent the results would have changed if Miller et al. 2015 had explicitly accounted for confounding factors or applied a different methodology, noting that no other relevant data on cigars have been provided by the parties. That being said, we note that although Ajzen et al. submit that conclusions about the impact of the TPP measures on cigarettes cannot be generalized to cigars, they also recognize that the impact of the TPP measures on increased noticeability of the health warnings on cigars was to be expected.[100] Therefore, we see no basis to reject in its entirety Miller et al.'s study on the basis of Ajzen et al.'s. criticism of it, noting that the overall findings reported in Miller et al. 2015 are to some extent in line with the results reported in the other peer-reviewed papers analysing the impact of GHW effectiveness on adult cigarette smokers reviewed above.

72. Finally, after a careful review, we question the robustness of Professor Klick's analysis of the commissioned Roy Morgan Research data. We note that unlike in the context of his empirical analysis of smoking incidence based on the Roy Morgan Research data, Professor Klick did not re-estimate his model specification to address some of the criticisms raised by Dr Chipty.[101] We also question the validity of Professor Klick's difference-in-difference analysis, for two main

[95] See para. 17 above for a description of the CITTS data. See Dunlop et al. 2014, (Exhibits AUS-207, HND-132, DOM-199).

[96] See Klick Second Supplemental Rebuttal Report, (Exhibit HND-165), para. 42.

[97] See Klick Second Supplemental Rebuttal Report, (Exhibit HND-165), pp. 18 and 21-23.

[98] See White et al. 2015b, (Exhibits HND-135, DOM-236, DOM-288).

[99] See Miller et al. 2015, (Exhibits AUS-102, DOM-315).

[100] See, Ajzen et al. Data Report, (Exhibit DOM/IDN-2), paras. 255-256.

[101] We are aware that most of Dr Chipty's critics referred to Professor Klick's analysis of smoking prevalence, but some of these critics apply to the difference-in-difference methodology in general.

reasons. First, the structure of the commissioned data prevents Professor Klick from accurately identifying the respondents who had noticed plain packs in the pre-period. This is because the question about noticing changes in the packaging was not asked to all respondents, making the pre-period covered by the commissioned data, in our view, very questionable.[102] Second, it is unclear how controlling for the excise tax increase that took place in New Zealand between Waves 1 and 2 and Waves 5 and 6 of the commissioned survey would have changed the results.

## 3 EVIDENCE RELATING TO THE ABILITY OF TOBACCO PACKAGING TO MISLEAD CONSUMERS ABOUT THE HARMFUL EFFECTS OF SMOKING SINCE THE ENTRY INTO FORCE OF THE TPP MEASURES

73. Two peer-reviewed papers analyse empirically the impact of Australia's TPP measures on the ability of the pack of tobacco products to mislead consumers among adult cigarette smokers and adolescents, respectively: (i) Wakefield et al. 2015; and (ii) White et al. 2015a.[103]

### 3.1 Datasets and related studies

#### 3.1.1 National Tobacco Plain Packaging Tracking Survey

74. In addition to analysing the impact of the TPP measures on appeal and GHWs effectiveness, Wakefield et al. 2015 uses the NTPPTS data to investigate their impact among adult smokers on tobacco packaging's ability to mislead consumers about smoking harms.[104] The authors present the results of a logistic model, using baseline survey weights and controlling for individual characteristics.[105]

75. Wakefield et al. 2015 find that, 12 months after their implementation, the TPP measures have had a positive and statistically significant effect on the proportion of adult smokers who believed that brands do not differ in harmfulness. However, the authors report no statistically significant change in the proportion of adult smokers who rated their current cigarette or tobacco as more harmful compared to a year ago and who believed brand variants do not differ in strength.[106]

76. Ajzen et al. in their expert report submitted by the Dominican Republic and Indonesia, submit that although Wakefield et al. 2015 acknowledges that the impact of the TPP measures on reducing the ability of packaging to mislead was mixed, the picture they presented must be considered incomplete because they failed to mention an important wear-out effect and to report statistically not significant results.[107] In particular, Ajzen et al. find that the increase in the proportion of adult smokers who believed that brands did not differ in harmfulness was very small and showed a wear-out effect. In addition, Ajzen et al. report no statistically significant change in the proportion of adult smokers that perceived their cigarettes are more harmful compared to a year ago, believed that brand variants do not differ in strength and agreed they had trouble believing that their regular brand of cigarettes is harmful.[108] More generally, the Dominican Republic considered Australia's claim that the NTPPTS was much broader in scope than Wakefield et al. 2015 to be surprising, as this message was not expressed in Wakefield et al. 2015 or in the journal's editorial in which the study was published.[109]

77. Professor Klick, in his expert report submitted by Honduras, reviews the NTPPTS data relating to the effect of the TPP measures on the ability of packaging to mislead consumers about the harm of smoking.[110] Professor Klick contends that there are other outcomes in the NTPPTS data, not

[102] See Chipty Rebuttal Report, (Exhibit AUS-535) (SCI), paras. 54-62.
[103] See Wakefield et al. 2015, (Exhibits AUS-206, DOM-306); and White et al. 2015a, (Exhibits AUS-186, DOM-235).
[104] See para. 8 above for a description of the NTPPTS data. See Wakefield et al. 2015, (Exhibits AUS-206, DOM-306).
[105] See para. 8 above for a description of the NTPPTS data.
[106] See para. 8 above for a description of the NTPPTS data. See Wakefield et al. 2015, (Exhibits AUS-206, DOM-306).
[107] See Ajzen et al. Data Report, (Exhibit DOM/IDN-2), paras. 128-132.
[108] See Ajzen et al. Data Report, (Exhibit DOM/IDN-2), paras. 123-127 and 156-158, Appendix A, pp. 84-86.
[109] Dominican Republic's comments on Australia's response to Panel question No. 198, para. 705.
[110] See Klick Supplemental Rebuttal Report, (Exhibit HND-122), paras. 82-86.

reported by Wakefield et al. 2015, which go against their conclusion that the TPP measures are achieving their goal regarding the ability of packaging to mislead consumers about smoking harms.[111]

78. Professor Klick presents the results of an ordered probit model, which controls for the TPP measures (early TPP period and formal TPP period), gender, age, education, socio-economic status and a linear time trend.[112] He finds that the TPP measures had no statistically significant impact on the respondents' belief about the harmfulness of the regular brand.[113]

79. Australia contends that the findings reported by Ajzen et al. are completely consistent with the findings of Wakefield et al. 2015 regarding the impact of the TPP measures on reducing the ability of packaging to mislead consumers about the harmful effects of smoking.[114] More generally, Australia argues that the results of Wakefield et al. 2015 were reported appropriately since the scope of the NTPPTS was much broader than the specific and limited focus of Wakefield et al. 2015. Australia further claims that the complainants' implication that unfavourable results were not reported is directly contradicted by the facts.[115]

### 3.1.2 Australian Secondary Students Alcohol Smoking and Drug Survey

80. In addition to assessing the impact of Australia's TPP measures on the appeal of tobacco products, White et al. 2015a employ the 2013 ASSAD survey extension to examine the impact of the TPP measures on the ability of tobacco product packaging to mislead about smoking harms among students aged 12-17 years.[116] The authors present the results of several generalized linear regression models and multinomial logistic regression models controlling for individual characteristics.

81. Overall, White et al. 2015a conclude that the impact of Australia's TPP measures on reducing the ability of cigarette packaging to mislead young smokers regarding the harmful effects of smoking is mixed. In particular, the authors report a statistically significant increase in the proportion of students who had seen a cigarette pack in the previous six months and disagreed that some brands are more addictive than others and that some cigarette brands contain more harmful substances than others. However, the authors find that the number of students agreeing that some cigarette brands are easier to smoke than others has increased, while there was no change over time in responses to the statement "some cigarette brands are easier to quit than others". White et al. 2015a conclude that further research is needed to determine whether continued exposure to standardized packs leads adolescents to develop more uncertainty or disagreement regarding brand differences in ease of smoking and quitting, perceived addictiveness and harms.[117]

82. As explained previously, the Dominican Republic argues that in the absence of full access to the dataset, it is impossible to make an objective assessment of the findings in White et al. 2015a. According to the Dominican Republic, the published results in White et al. 2015a may provide an unduly positive and inaccurate impression of the full dataset.[118] The Dominican Republic notes that White et al. 2015a acknowledge that the impact of plain packaging and enlarged GHWs on

[111] See Klick Supplemental Rebuttal Report, (Exhibit HND-122), para. 84.

[112] Professor Klick explains that, unlike Wakefield et al. 2015, he omits the measure of exposure to mass media anti-smoking messages, cigarette costliness and heaviness of smoking index, because these variables are endogenous. He further explains that the inclusion of these variables does not change the results he presented. See Klick Supplemental Rebuttal Report, (Exhibit HND-122), fn 35.

[113] See Klick Supplemental Rebuttal Report, (Exhibit HND-122), paras. 85-86. We note that in Professor Klick's model specification with an early TPP period and a formal TPP period, the impact of the early TPP period variable is positive and statistically significant, suggesting that the respondents are not more convinced that their regular brand is harmful.

[114] Australia's comments on the complainants' responses to Panel question No. 146, para. 7.

[115] Australia's response to Panel question No. 198, paras. 298-303.

[116] See para. 20 above for a description of the ASSAD survey dataset. See also White et al. 2015a, (Exhibits AUS-186, DOM-235).

[117] See White et al. 2015a, (Exhibits AUS-186, DOM-235).

[118] See Dominican Republic's second written submission, paras. 443-456; and comments on Australia's response to Panel question No. 196, para. 390.

deception are "mixed" and that the authors refrain from concluding that the pack changes have reduced deception.[119]

### 3.2 Analysis by the Panel

83. Among the various peer-reviewed studies discussed by the parties, two assess the impact of the TPP measures and enlarged GHWs on the ability of the packaging of tobacco products to mislead consumers about the harmfulness of smoking among adult cigarette smokers and adolescents.

84. A careful review of Wakefield et al. 2015 paper and the econometric evidence submitted by the Dominican Republic and Indonesia suggests that the impact of the TPP measures on the ability of the pack to mislead is much more mixed. Wakefield et al. 2015 only find a statistically significant impact of the TPP measures in reducing the belief that brands differ in harmfulness, but no impact on the belief that there is no difference in strength level across brands and that the current tobacco product is perceived as more harmful than a year ago.[120] While Ajzen et al. confirm these results, they also qualify the impact on the belief about difference in harmfulness across brands as being small and subject of wear-out.[121] Ajzen et al. and Professor Klick further claim that Wakefield et al. 2015 failed to report the result associated with the question on the harmfulness of their own brand, which according to them is statistically not significant. We note, however, that the results of the linear regression and ordered logistic regression suggest a small decrease in the proportion of adult smokers who do not have trouble believing that their brand is harmful, which is statistically significant at the 10% level.[122]

85. We note that the empirical evidence reported in Wakefield et al. 2015 regarding the mixed impact of the TPP measures on the packaging ability to mislead among adult smokers is supported by White et al. 2015b in the case of adolescents. A review of White et al. 2015b suggests that, one year after their implementation, the TPP measures have had some impact in reducing adolescents' beliefs in difference in addiction and harmfulness across brands. However, the authors find that the TPP measures have had no impact on adolescents' belief that some brands are easier to quit than others and, surprisingly, a positive impact on the belief that some cigarette brands are easier to smoke than others.[123]

## 4 OVERALL CONCLUSION ON POST-IMPLEMENTATION EVIDENCE ON NON-BEHAVIOURAL OUTCOMES

86. Overall, based on the studies and expert reports before us and discussed above, the empirical evidence available to us regarding the impact of the TPP measures, together with enlarged GHWs, since their entry into force, on the proximal outcomes of interest suggests that:

a. The TPP measures and enlarged GHWs have statistically significantly reduced the appeal of cigarettes among adult smokers.

b. The TPP measures and enlarged GHWs have statistically significantly increased GHWs' effectiveness on the noticeability of health warnings, avoidance of graphic health labels and pack concealment among adult cigarette smoker, albeit modestly for some outcomes, while the impact of the TPP measures and enlarged GHWs on adult cigarette smokers' health beliefs is relatively more limited and nuanced.

c. The TPP measures and enlarged GHWs have had a more mixed and limited impact on the ability of the pack to mislead adult cigarette smokers about the harmful effects of smoking.

---

[119] See Dominican Republic's comments on Australia's response to Panel question No. 200, paras. 793-794; and Ajzen Report, (Exhibit DOM/HND/IDN-3), paras. 174-178.

[120] See Wakefield et al. 2015, (Exhibits AUS-206, DOM-306).

[121] See Ajzen et al. Data Report, (Exhibit DOM/IDN-2), paras. 123-127, 156-158, Appendix A, pp. 84-86.

[122] See Ajzen et al. Data Report, (Exhibit DOM/IDN-2), Appendix A, pp. 85-86.

[123] See White et al. 2015b, (Exhibits HND-135, DOM-236, DOM-288).

d. While the TPP measures (together with enlarged GHWs) have contributed statistically significantly in reducing the appeal of cigarettes among adolescents, the impact of the TPP measures (with enlarged GHWs) on adolescents' health beliefs and cognitive processing of warning information on cigarettes packs is much more limited. Similarly, the impact of the TPP measures (with enlarged GHWs) on the ability of the pack to mislead adolescents about the harmful effects of smoking is more mixed and limited.

e. There has been a decrease in perceived packaging appeal when cigar and cigarillo smokers were exposed to the TPP measures (and enlarged GHWs). In addition, there has been an increase in the noticeability of health warnings and pack concealment among cigar and cigarillo smokers but the evidence is mixed regarding health beliefs.

87. No post-implementation empirical evidence has been presented to us on the impact of the TPP measures on the ability of the pack to mislead cigarillo and cigar smokers about smoking harms.

## APPENDIX B:

## EVIDENCE ON QUITTING-RELATED OUTCOMES AND OTHER DISTAL OUTCOMES SINCE THE ENTRY INTO FORCE OF THE TPP MEASURES

1. Earlier in these Reports, we referred to behaviours such as calling a Quitline and concealing a pack in public as "smoking-related behaviours". Australia's experts have referred to changes in quit intentions and quit attempts as "distal" outcomes of the TPP measures.[1] The Dominican Republic and Indonesia have similarly referred to quit intentions, secondary quit indicators, and quit attempts as "distal" outcomes and "antecedents of smoking behaviour".[2] We focus our discussion in this Appendix on the post-implementation evidence submitted by the parties with respect to these variables.

2. Australia submits that several peer-reviewed studies, which focus on quitting-related cognitions, pack concealment, and quit attempts, confirm that the TPP measures have resulted in increased calls to Quitline and the number of quit attempts.[3] Australia further submits that the features of certain survey data are most suited to detecting changes in proximal outcomes (i.e. tobacco products appeal, GHW effectiveness, and ability of packaging to mislead) than in more distal variables, such as intentions and quitting-related behaviours.[4]

3. The Dominican Republic and Indonesia submit that the TPP measures have not had the expected effects on the antecedents of behaviour posited by Australia's conceptual framework of the TPP measures. In particular, the Dominican Republic and Indonesia contend that beyond the obvious findings that the pack is less visually appealing and people more often notice the larger GHW first, empirical evidence shows little or no effects of the policies on the antecedents of behaviour. The Dominican Republic's, Honduras's and Indonesia's experts further argue that the empirical evidence shows that the TPP measures have had no impact on variables relating to quitting and relapse.[5] The Dominican Republic and Indonesia submitted expert reports dedicated to reviewing a series of peer-reviewed papers assessing the impact of the TPP measures on quitting-related outcomes. In some cases, the experts of the Dominican Republic, Honduras and Indonesia also re-analysed the data used in the studies.[6]

4. In addition, the experts of the Dominican Republic and Indonesia contend that some of the published empirical studies relied upon by Australia provide an inaccurate picture of the empirical evidence. According to them, some of these papers failed to report the results for more than half of all variables available in the survey dataset, which were overwhelmingly not statistically significant, suggesting no impact of plain packaging on these variables. The Dominican Republic and Indonesia contend that the authors of some of these studies failed also to explain that a number of the reported statistically significant effects had "vanished" by the end of the first year of the TPP measures' implementation as a result of wear-out effects. The Dominican Republic and Indonesia further criticize these papers for failing to report the effects size of the statistically significant effects. According to their experts, most of the reported statistically significant effects are small, suggesting that they have little importance in shifting behaviour. The Dominican Republic and Indonesia also criticize the authors of one of the studies for having removed the effects of a non-existent daily trend in survey responses, which has distorted the analysis by finding wrongly statistically significant effects.[7]

---

[1] See Chaloupka Rebuttal Report, (Exhibit AUS-582), para. 2; and NTPPTS Technical Report, (Exhibits AUS-570, HND-124, DOM-307), pp. 6 and 21.

[2] Ajzen et al. Data Rebuttal Report, (Exhibit DOM/IDN-6), paras. 27 and 35.

[3] Australia's second written submission, para. 464.

[4] Australia's comments on the complainants' responses to Panel question No. 197, paras. 371-375.

[5] See Ajzen et al. Data Report, (Exhibit DOM/IDN-2), paras. 10-22.

[6] See Ajzen et al. Data Report, (Exhibit DOM/IDN-2); Ajzen et al. Second Data Report, (Exhibit DOM/IDN-4); Ajzen et al. Data Rebuttal Report, (Exhibit DOM/IDN-6); Ajzen et al. Second Data Rebuttal Report, (Exhibit DOM/IDN-8); and Klick Report, (Exhibit UKR-5).

[7] See Ajzen et al. Data Report, (Exhibit DOM/IDN-2), paras. 23-27.

5. For each main type of outcome, namely quit intention, pack concealment and quit attempt, we discuss the survey data and corresponding papers separately, before turning to an overall assessment of the evidence before us. As in respect of the evidence relating to proximal outcomes addressed in Appendix A, we approach this assessment on the basis that our task is not to conduct our own econometric assessment of the impact of the TPP measures on the relevant outcomes but rather to examine, on the basis of the evidence before us, the overall robustness of the econometric evidence submitted by the parties in this respect.[8]

## 1 EVIDENCE RELATING TO QUITTING-RELATED COGNITIONS SINCE THE ENTRY INTO FORCE OF THE TPP MEASURES

6. Four peer-reviewed papers empirically analyse the impact of Australia's TPP measures on quitting-related cognitions: (i) Durkin et al. 2015; (ii) Yong et al. 2015; (iii) White et al. 2015b; and (iv) Miller et al. 2015.[9] An expert report prepared by Professor Klick, submitted by Ukraine, also contains a study analysing the impact on quitting-related cognitions.[10]

### 1.1 Datasets and related studies

#### 1.1.1 National Tobacco Plain Packaging Tracking Survey

7. As described in Appendix A, the NTPPTS, funded by Australia's Department of Health and Ageing, is a nationwide tracking survey conducted from 9 April 2012 to 30 March 2014 with a follow-up survey of baseline participants from 7 May 2012 to 4 May 2014.[11] Durkin et al. 2015 employ the NTPPTS data to examine the effects of the TPP measures on short-term changes in quit intentions. The authors estimate various logistic models based on the respondents' experiences regarding quit intentions over the one-month follow-up period for the cohorts surveyed before the implementation of the TPP measures, over the transition period to TPP, and during the first year of implementation. The authors adjusted the models for baseline levels of the outcome and controlled for the date of the follow-up survey, number of days between baseline and follow-up surveys, anti-smoking television advertising in the three months prior to the follow-up survey, change in cigarette costliness, sex, age, education, socio-economic status, and addiction level.[12] Durkin et al. 2015 also present the results of unadjusted models, where the only control variable is the baseline response.

8. Durkin et al. 2015 find that the TPP measures were associated with increased rates of quitting cognitions among adult cigarette smokers. In particular, compared to the referent group of smokers having completed their follow-up survey in the pre-TPP period, smokers who were followed-up with in the late transition period reported greater increases in their intentions to quit in the next month. However, the authors find that there was no statistically significant change in the proportion of adult smokers who were followed-up with in the first year of implementation of the TPP measures who thought daily about quitting in the past week, intended to quit in the next month or set a firm date to quit in the next month compared with the pre-TPP period.

9. Ajzen et al., in their expert report submitted by the Dominican Republic and Indonesia, use the NTPPTS data to review the findings reported in Durkin et al. 2015. They argue that Durkin et al.'s 2015 conclusions are unwarranted and contradicted by the data. According to Ajzen et al., Durkin et al. 2015 overstate their findings because the TPP measures have had no impact on the majority of the quitting-related measures.[13] More generally, Ajzen et al. contend that Durkin et al. 2015 have adopted a very different and unorthodox comparison methodology compared to Wakefield et al. 2015, despite being the same six authors using the same dataset.

---

[8] For a similar approach, see Panel Reports, *US – COOL (Article 21.5 – Canada and Mexico)*, para. 7.183 (citing Panel Reports, *US – COOL*, para. 7.539).

[9] See Durkin et al. 2015, (Exhibits AUS-215 (revised), DOM-305); Yong et al. 2015, (Exhibit DOM-382); White et al. 2015b, (Exhibits HND-135, DOM-236, DOM-288); and Miller et al. 2015, (Exhibits AUS-102, DOM-315).

[10] See section 1.6.6 in the main body of these Reports for a description of Ukraine's participation in these proceedings.

[11] Dominican Republic's second written submission, paras. 380-429.

[12] See Durkin et al. 2015, (Exhibits AUS-215 (revised), DOM-305).

[13] See Ajzen et al. Data Report, (Exhibit DOM/IDN-2), paras. 174-179.

10. Ajzen et al. submit that Durkin et al.'s analytical approach suffers from three serious methodological shortcomings. First, unlike Wakefield et al. 2015, Durkin et al. 2015 control for a non-existent daily trend that distorts the estimates of the impact of the TPP measures by conflating the contribution to behavioural change due to the daily trend on the one hand, and the TPP measures on the other. When the daily trend is removed from Durkin's own analysis, each of the significant results reported in Durkin et al. 2015 disappears.[14] Second, Durkin et al. 2015 failed to show that the statistically significant results experienced wear-out effects, with none of the results still significant in the final quarter of the first year following the introduction of the TPP measures.[15] Third, Durkin et al. 2015 did not correct for multiple significances tests that would have shown that none of the effects reported as being statistically significant by Durkin et al. 2015 remain statistically significant.[16] Ajzen et al. further argue that Durkin et al.'s claim that relying on changes in participants' responses from month to month minimizes the influence of sampling variation and increase the power of the statistical tests has been mathematically proven to be wrong.[17]

11. Ajzen et al. present the results of a logistic model for dichotomized outcome variables, linear model for continuous outcome variables and ordered logit model for categorical outcome variables, controlling for the same covariates as Durkin et al. 2015 except for the date of the follow-up interview. They find no statistically significant change in the proportion of adult smokers who were followed-up in the first year of the TPP measures who thought daily about quitting in the past week, intended to quit in the next month or set a firm date to quit in the next month.[18] In addition, Ajzen et al. claim that Durkin et al. 2015 failed to report the result showing that the TPP measures did not bring about any increase in respondents' ability to perceive the importance of staying quit.[19]

12. Professor Klick, in his expert report submitted by Honduras, looks at the NTPPTS data on the effect of the TPP measures on quitting-related cognitions and intentions.[20] Professor Klick claims that Wakefield et al. 2015 did not present the results of some variables related to quitting that did not improve or might have even gotten worse with the TPP measures.[21]

13. Professor Klick presents the results of an ordered probit model, which controls for the TPP measures (early TPP period and formal TPP period), gender, age, education, socio-economic status and a linear time trend.[22] He finds that there is no evidence that the TPP measures increased the importance of quitting in respondents' mind. He reports, however, a negative and statistically significant impact of the TPP measures on the respondents' thoughts about quitting.[23]

14. Professor Klick further contends that Durkin et al.'s 2015 results of the unadjusted models, which show that none of the quitting-related cognition variables are statistically distinguishable from the pre-TPP period, must be given appropriate weight and considered as at least equally valid compared with the results of the adjusted models, for various reasons.[24] First, Professor Klick claims that accounting for the respondent's characteristics does not seem necessary, because they should be captured by the individual's baseline response. Second, he argues that the variables for the individual's heaviness of smoking and cigarette price are likely to be endogenous, which could bias the estimates of the TPP measures. Third, he claims that the variable for the exposure to mass media anti-smoking messages, which is computed over a fixed three-month time-period, measures different things for different respondents and controlling for days between interviews

[14] See Ajzen et al. Data Report, (Exhibit DOM/IDN-2), paras. 188-192.
[15] See Ajzen et al. Data Report, (Exhibit DOM/IDN-2), paras. 193-195.
[16] See Ajzen et al. Data Report, (Exhibit DOM/IDN-2), paras. 196-198.
[17] See Ajzen et al. Data Report, (Exhibit DOM/IDN-2), paras. 184-185.
[18] See Ajzen et al. Data Report, (Exhibit DOM/IDN-2), paras. 165-173, Appendix A, pp. 90-92.
[19] See Ajzen et al. Data Report, (Exhibit DOM/IDN-2), para. 179.
[20] See Klick Supplemental Rebuttal Report, (Exhibit HND-122), paras. 69-73, 88-100.
[21] See Klick Supplemental Rebuttal Report, (Exhibit HND-122), paras. 61-63.
[22] Professor Klick explains that unlike Wakefield et al. 2015 he omits the measure of exposure to mass media anti-smoking messages, cigarette costliness and heaviness of smoking index, because these variables are endogenous. He further explains that the inclusion of these variables does not change the results presented. See Klick Supplemental Rebuttal Report, (Exhibit HND-122), fn 35.
[23] See Klick Supplemental Rebuttal Report, (Exhibit HND-122), paras. 69-73. We note that Professor Klick considers these two questions to be relevant to the appeal of tobacco.
[24] Klick Supplemental Rebuttal Report, (Exhibit HND-122), paras. 91-100.

does not address this problem. Finally, Professor Klick argues that Durkin et al. 2015 do not explain what adjusting the models for the date of the follow-up survey means.[25]

15. Professor Chaloupka, Australia's expert, contends that Ajzen et al.'s analysis suffers from several limitations, in part resulting from the particular features of the NTPPTS data. As a result, this analysis creates more confusion than it provides help in understanding whether Australia is achieving the specific and broad objectives of the TPP measures.[26] In his view, the fact that there are inherent limitations to any data and analysis is the primary reason for considering the overall weight of the evidence based on a variety of data and methods, rather than relying on any one set of data or analysis.[27] Professor Chaloupka argues that the NTPPTS data cannot be used to assess the impact of the TPP measures on the population most likely to be influenced by TPP, namely individuals who might have taken up tobacco use in the absence of the TPP. Professor Chaloupka further submits that because of its cross-sectional nature, the NTPPTS data limit the ability of analysts to follow the impact of the TPP measures through the pathway from its impact on the most proximal outcomes, like perceived appeal and noticing of the labels, through less proximal outcomes, such as increased interest in quitting, to the most distal outcomes, such as actual tobacco use behaviour. Professor Chaloupka is of the view that a true longitudinal survey tracking the impact of the TPP measures over a number of years on the same sample of current tobacco users and recent quitters would allow for the type of sequential analysis that would be more helpful in fully understanding how the TPP measures affect the more and less proximal outcomes they are expected to influence, as well as to more fully understand their impact on actual tobacco use behaviour.[28]

16. According to Professor Chaloupka, Ajzen et al. fail to recognize that the power to detect statistically significant changes will fall for outcomes that are increasingly distal. The analysis of distal outcomes requires relatively large sample sizes to detect these increasingly smaller effects when it is based on all current tobacco users and recent quitters, and not just the subsample of users for whom more proximal outcomes were influenced by the TPP measures.[29] Professor Chaloupka further submits that given that the questions about quitting-related behaviours were only asked of smokers, any analysis that uses these as outcomes are likely to considerably understate the effects of the TPP measures on quitting-related intentions and behaviour. This is especially so given that the recent quitters whose intentions and behaviour are likely to have already been influenced by the TPP measures are not included in the analysis. Instead, analyses that employ these outcomes focus on the sample of smokers that are least likely to have been influenced by the TPP measures, making it less than surprising that the findings on the impact of the TPP measures on quitting-related behaviour variables are relatively weak.[30]

17. Overall, Professor Chaloupka contends that the pattern of results reported by Ajzen et al. is consistent with the strengths and limitations of the NTPPTS data. As expected, Ajzen et al. found consistent and statistically significant effects for the impact of the TPP measures on the most proximal outcomes. The impact is smaller, less statistically significant and less consistent as the focus shifts to less proximal outcomes, such as health knowledge and perceptions of the health risks. Likewise, the impact is smaller, less significant and less consistent as the focus shifts to more distal outcomes, including intentions and other quit-related measures.[31]

18. Ajzen et al. respond that Professor Chaloupka's assertion that the more distal outcomes will be less affected by the policy than the more proximal outcomes is unfounded. In their view, this weakening of the policy across the chain of effect does not reflect a limitation of the NTPPTS data, but the theoretically expected decline in the impact of the TPP measures across the chain of effect

[25] Klick Supplemental Rebuttal Report, (Exhibit HND-122), paras. 96-99.

[26] See Australia's response to Panel question No. 196, para. 236; and Australia's comments on the complainants' responses to Panel question No. 197, paras. 369-376.

[27] See Chaloupka Rebuttal Report, (Exhibit AUS-582), paras. 2 and 10; Chaloupka Second Rebuttal Report, (Exhibit AUS-590), para. 12; and Chaloupka Third Rebuttal Report, (Exhibit AUS-604), para. 7.

[28] See Chaloupka Rebuttal Report, (Exhibit AUS-582), paras. 2-5.

[29] See Chaloupka Rebuttal Report, (Exhibit AUS-582), para. 6; and Chaloupka Third Rebuttal Report, (Exhibit AUS-604), paras. 4-5.

[30] See Chaloupka Rebuttal Report, (Exhibit AUS-582), para. 7; Chaloupka Second Rebuttal Report, (Exhibit AUS-590), para. 20; and Chaloupka Third Rebuttal Report, (Exhibit AUS-604), para. 6.

[31] See Chaloupka Rebuttal Report, (Exhibit AUS-582), para. 9; and Chaloupka Third Rebuttal Report, (Exhibit AUS-604), paras. 2-7.

from pack appeal to behaviour.[32] Ajzen et al. further contend that Professor Chaloupka's assertion that the NTPPTS data cannot reveal longer-term effects of the TPP measures because of its cross-sectional nature is unfounded. According to them, a cross-sectional design is widely used and accepted. In addition, the results found in the NTPPTS data are confirmed by the findings based on the longitudinal International Tobacco Control (ITC) Policy Evaluation Project (ITC Project).[33] Likewise, Ajzen et al. argue that Professor Chaloupka's assertion that the NTPPTS dataset has enough statistical power to detect small changes in the most proximal outcomes but not in more distal outcomes is unfounded. Ajzen et al. claim that the statistical power of the NTPPTS data to detect small changes in proximal and distal outcomes is very similar.[34] Similarly, Ajzen et al. contend that Professor Chaloupka's assertion that the NTPPTS data underestimate changes in some measures of intention because recent quitters who gave up smoking because of the TPP measures have not been asked is unfounded because the TPP measures did not increase quitting behaviours in the short term.[35]

### 1.1.2 International Tobacco Control Policy Evaluation Project

19. As described in Appendix A, the ITC Project is a longitudinal cohort survey on the determinants of tobacco control policies in more than 20 countries, including Australia. The ITC Project covers a number of questions related to GHWs, including on consumers' thoughts, behaviours and intentions towards quitting.

20. The Australian component of the ITC Project used by Yong et al. 2015 was conducted between September 2011 and February 2012 for the pre-implementation period and between February and May 2013 for the post-implementation period.

21. Yong et al. 2015 analyse the relationship between health warning label reactions and quit intentions. The authors present the results of various GEE models controlling for age, sex, income, education, cigarettes per day, past year quit attempts, survey mode (phone vs. web) and wave of recruitment.[36]

22. Yong et al. 2015 find that the TPP measures have led to a statistically significant increase in the proportion of adult smokers thinking more about smoking health risks and quitting as a result of the GHWs. In particular, they report a statistically significant increase in the proportion of respondents who did not focus on health warning labels first, but who now focus on them first, and thought more about quitting as a result of the health warning labels. Conversely, the subgroup of respondents who chose not to focus on the health warning labels experienced less cognitive reactions and thought less about quitting. In addition, Yong et al. 2015 find that adult smokers, who had been stimulated by health warning labels to think about the harm caused by smoking and about quitting, were more motivated to quit smoking in the future. In that context, the authors argue that the cognitive reactions serve as an important pathway through which the effects of more upstream reactions to the health warning labels exert their influence on quit intentions. Yong et al. 2015 explain, however, that given that the prediction of quit intentions was based on cross-sectional data, caution needs to be exercised in interpreting the finding as causal. The authors note that while intentions are logically subsequent to reported past reactions, it is possible that the person's intentional state may affect their recall of past reactions.[37]

23. Ajzen et al., in their expert report submitted by the Dominican Republic and Indonesia, review Yong et al.'s analysis and submit that given that they only had access to a small subset of the ITC dataset, they were unable to present a comprehensive assessment of data, including correcting for the possibility of false positive findings due to multiple hypothesis testing (i.e. statistically significant results might have occurred by chance), controlling for anti-smoking advertising in mass media and testing for wear-out effects.[38] Ajzen et al. further contend that

[32] See Ajzen et al. Data Rebuttal Report, (Exhibit DOM/IDN-6), paras. 39-45.
[33] See Ajzen et al. Data Rebuttal Report, (Exhibit DOM/IDN-6), paras. 46-49, Appendix I.
[34] See Ajzen et al. Data Rebuttal Report, (Exhibit DOM/IDN-6), paras. 50-54, Appendix II.
[35] See Ajzen et al. Data Rebuttal Report, (Exhibit DOM/IDN-6), paras. 55-62, Appendix III.
[36] See Yong et al. 2015, (Exhibit DOM-382).
[37] See Yong et al. 2015, (Exhibit DOM-382).
[38] See Dominican Republic's comments on Australia's response to Panel question No. 196, paras. 409-414. See also Ajzen et al. Second Data Report, (Exhibit DOM/IDN-4), paras. 12-19.

Yong et al. 2015 failed to report effect sizes, explore wear-out effects, correct for multiple hypothesis testing and justify their selection of outcome variables available in the ITC dataset. In particular, they criticise Yong et al. 2015 for not reporting the statistically significant or non-significant impact of the TPP measures on certain downstream variables.[39]

24. Ajzen et al. find that although the TPP measures have had a small positive and statistically significant impact on thinking about quitting as a result of the GHWs, there was also a small but statistically significant decrease in the proportion of adult smokers that reported interest in quitting and intentions to quit. According to Ajzen et al., Yong et al. 2015 did not report these two last results, which are in the "wrong direction" for the TPP measures.[40]

25. Australia contends that Ajzen et al. correctly conclude that the TPP measures have significantly increased cognitive reactions to GHWs.[41] However, Australia is of the view that Ajzen et al., when concluding that interest in and intentions to quit decreased post-plain packaging, do not acknowledge or appear to appreciate that questions relating to quit intentions were not asked of recent quitters – the cohort whose intentions and behaviours were most likely to have been influenced by the TPP measures.[42] Australia concludes that consequently the sample Ajzen et al. used to analyse quit intentions was negatively biased. According to Ajzen et al., Australia and its experts do not contest the accuracy of their analytical approach or their results, and do not challenge their "serious criticisms" of Yong et al. 2015.[43] They submit that Australia's sole criticism regarding the ITC analysis seeks to denigrate the quality of the ITC dataset itself, claiming that it provides a "biased" sample to measure the impact on quit intentions and quit interests. According to Ajzen et al., Australia's argument assumes that the TPP measures led smokers to quit shortly after their introduction, and that these "recent quitters" were subsequently excluded from responding to questions on quit intentions and quit interests in the ITC survey. However, Ajzen et al. submit that the data show that no additional quit attempts were made post-implementation, which contradict Australia's argument. Further, Ajzen et al. consider it surprising to hear (for the first time) from Australia that the ITC dataset is "biased" in this regard because it was developed by the tobacco control community to measure the real-world impact of tobacco control measures, such as plain packaging, on smoking-related outcomes, including the impact of the TPP measures on quit intentions.[44] Ajzen et al. argue that it is only after having demonstrated that quit intentions have not increased post-implementation that Australia claims that the ITC sample is biased in relation to this particular measure.[45]

### 1.1.3 Australian Secondary Students Alcohol Smoking and Drug Survey

26. As described in Appendix A, the 2013 ASSAD survey extension is a survey of students in 82 secondary schools in Victoria and Queensland tracking their beliefs and attitudes about cigarette packaging, ratings of popular cigarette brands, noticing health warnings on cigarette packs, awareness of the specific harms of tobacco use, and perceptions of the prevalence of smoking and intention to smoke.[46]

27. White et al. 2015b employ the ASSAD survey data to examine the impact of the TPP measures on students aged 12-17 years. The authors estimate a logistic regression model of youth smokers' experience with quit intentions controlling for age, sex, school type, state and smoking status.[47]

28. White et al. 2015b find that the TPP measures have had no impact in the proportion of students who, having seen a cigarette pack in the previous six months, thought about quitting smoking. The authors conclude that the data suggest that the introduction of the TPP measures

[39] See Ajzen et al. Second Data Report, (Exhibit DOM/IDN-4), paras. 30-37 and 53-58.
[40] See Ajzen et al. Second Data Report, (Exhibit DOM/IDN-4), paras. 24 and 28.
[41] See Australia's response to Panel question No. 196, paras. 225 and 237.
[42] Australia's response to Panel question No. 196, para. 237.
[43] See Dominican Republic's comments on Australia's response to Panel question No. 196, paras. 409-414. See also Ajzen et al. Second Data Rebuttal Report, (Exhibit DOM/IDN-8), paras. 13-16.
[44] See Ajzen et al. Second Data Rebuttal Report, (Exhibit DOM/IDN-8), paras. 15-16.
[45] See Ajzen et al. Second Data Rebuttal Report, (Exhibit DOM/IDN-8), para. 14.
[46] See Dominican Republic's second written submission, paras. 443-456.
[47] See White et al. 2015b, (Exhibits HND-135, DOM-236, DOM-288).

did not induce adolescent to attend to and process warnings on cigarette packs to a greater extent than when GHWs covered 30% of the front of a fully branded pack.[48]

29. According to the Dominican Republic, the results published in White et al. 2015b may provide an unduly positive and inaccurate impression of the full dataset, but without full access to the ASSAD survey data, it cannot make an objective assessment of White et al.'s findings.[49] The Dominican Republic further contends that although in another study White et al. 2015a conclude that there is a change in the visual appeal of packs among adolescents[50], that reduced appeal does not have any meaningful impact on quit intentions.[51]

### 1.1.4 Cigar and cigarillo smokers surveys

30. Miller et al. 2015 conducted, in February and March 2014, individual interviews with regular premium cigar smokers, as well as two focus groups with premium cigarillo smokers and occasional premium cigar smokers, four focus groups with non-premium cigarillo smokers, and an online survey of current cigar and cigarillo smokers.[52] The interviews took place 15 months after the TPP measures became mandatory. The authors present descriptive statistics of the responses of these interviews, focus groups and the online survey.

31. Miller et al. 2015 find that non-premium cigarillo smokers reported high plain packaging exposure and more contemplation of quitting.

32. Ajzen et al., in their expert report submitted by the Dominican Republic and Indonesia, argue that the conclusions that can be drawn from the results reported by Miller et al. 2015 are limited by several methodological shortcomings, such as the non-representativeness of survey participants, the unsuitability of focus groups to draw causal inferences, failure to control for exposure to anti-smoking campaigns and changes in tobacco prices, and absence of "baseline" information collected before the implementation of the TPP measures.[53] They further claim that Miller et al. 2015 did not mention that the cigar and cigarillo smokers participating in the online survey reported no change in thinking about quitting. According to Ajzen et al., the TPP measures have had no impact on reported changes in thoughts about quitting among cigar and cigarillo smokers, because the TPP measures did not affect their ratings of the appeal, taste or quality of the products they currently smoked, their enjoyment of smoking or its perceived health risks. More generally, Ajzen et al. contend that smoking cigarettes and smoking cigars (or cigarillos) are different behaviours, with potentially different determinants, which implies that the findings about the TPP measures on cigarette-related cognitions and behaviours cannot be generalized to cigars.[54]

### 1.1.5 Commissioned Roy Morgan Research Survey (Australia and New Zealand)

33. Roy Morgan Research dataset, commissioned by Professor Klick, is a survey of current or former (in the past 12 months) smokers in both Australia and New Zealand. The first wave of the survey was completed prior to the implementation of Australia's TPP measures between 2 November 2012 and 26 November 2012 in Australia and between 8 November 2012 and 1 December 2012 in New Zealand. Subsequent waves were carried out at three-month intervals up until February 2014.[55] Professor Klick employs the Roy Morgan Research data to estimate a difference-in-difference logit model that explains the respondents' experience with their intention to quit in Australia and New Zealand, controlling for the TPP measures, an Australian baseline variable, a common baseline variable and a post-TPP implementation time-period.

---

[48] See White et al. 2015b, (Exhibits HND-135, DOM-236, DOM-288), p. ii56.

[49] See Dominican Republic's second written submission, para. 445.

[50] See White et al. 2015a, (Exhibits AUS-186, DOM-235).

[51] See Dominican Republic's second written submission, paras. 443-456; and Dominican Republic's comments on Australia's response to Panel question No. 196, para. 390.

[52] See Miller et al. 2015, (Exhibits AUS-102, DOM-315).

[53] See Dominican Republic's second written submission, paras. 457-462. See also Ajzen et al. Data Report, (Exhibit DOM/IDN-2), paras. 251-262.

[54] See Ajzen et al. Data Report, (Exhibit DOM/IDN-2), paras. 251-262.

[55] See Klick Report, (Exhibit UKR-5), pp. 6-8. See section 1.6.6 in the main body of these Reports for a description of Ukraine's participation in these proceedings.

34. Professor Klick finds that the TPP measures have had no statistically significant effect on smokers' quit intentions and expectations about future smoking in Australia relative to New Zealand. In particular, he reports no statistically significant impact of the TPP measures in Australia relative to New Zealand on the likelihood that the respondents seriously considered quitting smoking, had the desire, even slightly, to give up smoking, or had the intention to try to quit smoking in the next six months.[56] Similarly, Professor Klick's finds no statistically significant impact of the TPP measures on the likelihood that the respondents were thinking about increasing future smoking, smoking just as much, trying and easing up future smoking, changing to a low tar brand or making a definitive attempt to quit.[57]

35. As explained in Appendix A, Australia's expert, Dr Chipty, does not refer directly to Professor Klick's analysis of quitting-related cognitions outcomes based on the Roy Morgan Research data. However, a number of general criticisms formulated by Dr Chipty regarding Professor Klick's difference-in-difference analysis of smoking incidence are also relevant to the analysis of quitting-related cognitions outcomes. Dr Chipty submits that Professor Klick's analysis of the Roy Morgan Research data is invalid because the commissioned Roy Morgan survey does not include a pre-period and does not allow for distinguishing which respondents had noticed in plain packaging. Dr Chipty contends also that New Zealand is an invalid counterfactual for the purposes of studying the effects of plain packaging, because New Zealand increased the excise tax on cigarettes in January 2013 one month after the introduction of the TPP measures.[58]

### 1.2 Analysis by the Panel

36. We note that only four peer-reviewed studies papers assess the impact of Australia's TPP measures and enlarged GHWs on quitting-related cognitions among adult cigarette smokers, cigar and cigarillo smokers and adolescents.

37. A careful review of the Durkin et al. 2015 and Yong et al. 2015 papers and the econometric evidence submitted by the Dominican Republic, Honduras and Indonesia suggests that the impact of the TPP measures on quitting intentions among adult smokers is limited and mixed.

38. We note that Durkin et al. 2015 only find some evidence that the TPP measures have had a positive and statistically significant impact on the intention to quit during the late transition period of the TPP. They report no statistically significant impact of the TPP measures on daily thoughts about quitting, intention to quit and setting a firm date to quit among adult smokers in the first year of the TPP.[59] We note that Ajzen et al., applying an alternative estimation methodology, which consists of re-estimating the model without the daily trend, confirm most of Durkin et al.'s results. We note, however, that the results of the logistic model reported by Ajzen et al. suggest that the TPP measures have had a positive and statistically significant (at the 10% level) impact on the likelihood of adult smokers reporting having set a firm date to quit in the next month. The results of the linear regression also suggest that the TPP measures have had a negative and statistically significant (at the 10% level) impact on quitting importance.[60] More generally, we note that Australia did not rebut Ajzen et al.'s and Professor Klick's alternative estimation methodology and associated results. We are, however, not persuaded by Professor Klick's claim that unadjusted models, which do not control for individual characteristics, addiction level and other tobacco control policies (exposure to anti-smoking television advertising and cigarette costliness), should be considered at least equally valid compared with adjusted models that do control for these variables. We note that Ajzen et al. focused their analysis on the adjusted models. We further note that the econometric results for quitting-related variables (i.e. quitting importance, thoughts about quitting) reported by Professor Klick are also based on adjusted models that control for the respondent's individual characteristics. Professor Klick further acknowledges that the results he reported do not change when the variables for addiction level, exposure to anti-smoking television advertising and cigarette costliness are included in the model specification, which would suggest that the risk of endogenous bias might be limited.

[56] See Klick Report, (Exhibit UKR-5), pp. 18 and 37-44.
[57] See Klick Report, (Exhibit UKR-5), pp. 18 and 37-47.
[58] See Chipty Rebuttal Report, (Exhibit AUS-535) (SCI), paras. 54-71.
[59] See Durkin et al. 2015, (Exhibits AUS-215 (revised), DOM-305).
[60] See Ajzen et al. Data Report, (Exhibit DOM/IDN-2), paras. 165-173, Appendix A, pp. 90-92.

39. Unlike Durkin et al. 2015, Yong et al. 2015 find relatively more evidence of a positive impact of the TPP measures on quitting-related cognitions. In particular, Yong et al. 2015 report a statistically significant impact of the TPP measures on thinking more about quitting as a result of health warnings among adult smokers who switched from initially focusing away to focusing on health warning labels. They further find that adult smokers, having been stimulated by health warning labels to think about the harms of smoking and about quitting, reported being more motivated to quit smoking in the future.[61] We note that Ajzen et al. confirm that the TPP measures have had a positive and statistically significant impact on thinking about quitting but qualify it as small.[62] Ajzen et al. further report a small and statistically significant decrease in the proportion of adult smokers reporting their interest in quitting and intention to quit.[63] We note that Ajzen et al. do not offer an explanation as to why the TPP measures would have *decreased* smokers' interest in quitting and intention to quit. We note that it is conceivable that these findings could partly result from the fact that, as observed by Australia, questions on quitting intention were not asked of recent quitters, contrary to Ajzen et al.'s claim that the question was asked to both smokers and recent quitters, although it is not clear, in the absence of specific evidence relating to the number of recent quitters, to what extent this circumstance may account for the results.

40. We observe that the conclusion that the impact of the TPP measures on quitting intentions is limited and mixed is confirmed by the two other peer-reviewed studies on adolescent and cigar and cigarillo smokers. White et al. 2015b suggest that the TPP measures have had no impact on thinking about quitting smoking among adolescents, which is consistent with the findings reported in Durkin et al. 2015. Conversely, the descriptive statistics analysis of Miller et al. 2015 suggests that the share of non-premium cigarillo smokers contemplating quitting increased.[64] As explained in our review of the post-implementation studies before us on proximal outcomes[65], we see no basis to reject Miller et al.'s findings on the basis of Ajzen et al.'s criticism, noting that the result on quitting intentions reported in Miller et al. 2015 is in line with the results reported in Durkin et al. 2015. We note, however, that Miller et al. 2015 provide no empirical evidence on quitting intentions among (premium or non-premium) cigar smokers.

41. Finally, as explained in our review of Professor Klick's analysis of GHW effectiveness and smoking incidence based on the commissioned Roy Morgan Research data, we question the robustness of his results. We question the validity of the pre-period of the commissioned data, because it corresponds to the TPP measures transition period, during which plain packs could already be sold on the Australian market. In addition, Professor Klick's analysis does not control for New Zealand's excise tax increase one month after the introduction of the TPP measures, which leads us to question the use of New Zealand as a counterfactual.[66]

## 2 EVIDENCE RELATING TO PACK CONCEALMENT AND MICRO-INDICATORS OF CONCERN SINCE THE ENTRY INTO FORCE OF THE TPP MEASURES

42. Six peer-reviewed papers analyse empirically the impact of Australia's TPP measures on distal outcomes related to quitting: (i) Durkin et al. 2015; (ii) Yong et al. 2015; (iii) Zacher et al. 2014; (iv) Zacher et al. 2015; (v) White et al. 2015b; and (vi) Miller et al. 2015.[67] Professor Klick, in his expert report submitted by Ukraine, also assesses the impact of the TPP measures on pack concealment and related variables.[68]

[61] See Yong et al. 2015, (Exhibit DOM-382).

[62] We note that Ajzen et al. only replicate one part of the analysis of Yong et al. 2015 and do not analyse the relationship of changes in attentional orientation response pattern with changes in GHW reactions and the association of GHWs with quit intentions.

[63] See Ajzen et al. Second Data Report, (Exhibit DOM/IDN-4), paras. 24 and 28.

[64] See Miller et al. 2015, (Exhibits AUS-102, DOM-315).

[65] See Appendix A, paras. 32 and 71 above.

[66] See Appendix A, para. 72 above.

[67] See Durkin et al. 2015, (Exhibits AUS-215 (revised), DOM-305); Yong et al. 2015, (Exhibit DOM-382); Zacher et al. 2014, (Exhibits AUS-222 (revised), JE-24(68), DOM-286); Zacher et al. 2015, (Exhibits AUS-223 (revised), DOM-287); White et al. 2015b, (Exhibits HND-135, DOM-236, DOM-288); and Miller et al. 2015, (Exhibits AUS-102, DOM-315).

[68] See section 1.6.6 in the main body of these Reports for a description of Ukraine's participation in these proceedings.

## 2.1 Datasets and related studies

### 2.1.1 National Tobacco Plain Packaging Tracking Survey

43. Durkin et al. 2015 use the NTPPTS data to analyse the impact of the TPP measures on pack concealment and micro-indicators of concern.[69] The NTPPTS Technical Report measured two micro-indicators of concern about smoking, namely stubbing out a tobacco product before finishing due to thoughts about the harms of smoking, and stopping oneself from smoking despite an urge to smoke.[70] The authors estimate logistic models of the respondents' experiences regarding pack concealment and other related behaviours by adjusting for baseline levels and controlling for the date of the follow-up survey, number of days between baseline and follow-up survey, anti-smoking television advertising in the three months prior to the follow-up survey, change in cigarette costliness, addiction level and other individual characteristics.[71] Durkin et al. 2015 also present the results of unadjusted models, where the baseline response is the only control variable.

44. Durkin et al. 2015 find that the TPP measures were associated with increased rates of pack concealment and micro-indicators of smoking concern among adult cigarette smokers.[72] In particular, they find that compared to the referent group of smokers having completed their follow-up survey in the pre-TPP period, smokers who were followed-up in the late transition period showed greater increases in pack concealment, amounting to several or many times in the previous month. Similarly, they find that smokers who were followed-up in the first year of the TPP measures reported greater increases in pack concealment and stubbing out their cigarettes before finishing them because they thought about the harm of smoking.[73]

45. Ajzen et al., in their expert report submitted by the Dominican Republic and Indonesia, claim that Durkin et al.'s conclusions are unwarranted and contradicted by the data, because the TPP measures have had no impact on the majority of the quitting-related measures.[74] They further contend that Durkin et al. 2015 have adopted a very different and unorthodox comparison methodology compared to Wakefield et al. 2015, despite being the same six authors using the same dataset, including analysing the same variable on pack concealment. According to Ajzen et al., Durkin et al.'s 2015 analytical approach suffers from serious methodological shortcomings, including controlling for a non-existent daily trend, failing to investigate wear-out effects and not correcting for multiple significances tests.[75] According to Ajzen et al., Durkin et al.'s 2015 assertion that analysing changes in the participants' responses from month to month minimizes the influence of sampling variation and increase the power of the statistical tests has been mathematically proven to be wrong.[76]

46. Based on the NTPPTS data, Ajzen et al. estimate a logistic model for dichotomized outcome variables, linear model for continuous outcome variables and ordered logit model for categorical outcome variables with the same specification considered by Durkin et al. 2015 but without the variable for date of the follow-up interview. They find a statistically significant increase in the proportion of adult smokers, who were followed-up in the first year of TPP, who concealed their pack several or many times. However, they report no statistically significant change in the proportion of adult smokers who were followed-up in the first year of TPP, that stubbed out and foregone cigarettes several or many times.[77] Ajzen et al. further argue that Durkin et al. 2015 failed to report the result showing that the TPP measures has had no impact on respondents' ability to limit their consumption of cigarettes.[78]

---

[69] See Appendix A, para. 8 for a description of the NTPPTS data.
[70] NTPPTS Technical Report, (Exhibits AUS-570, HND-124, DOM-307), p. 52.
[71] See Durkin et al. 2015, (Exhibits AUS-215 (revised), DOM-305).
[72] As discussed in Appendix A, Wakefield et al. 2015 used the NTPPTS data and concluded also that there was a statistically significant increase in smokers covering their packs. See Wakefield et al. 2015, (Exhibits AUS-206, DOM-306).
[73] See Durkin et al. 2015, (Exhibits AUS-215 (revised), DOM-305).
[74] See Ajzen et al. Data Report, (Exhibit DOM/IDN-2), paras. 174-179.
[75] See Ajzen et al. Data Report, (Exhibit DOM/IDN-2), paras. 188-198.
[76] See Ajzen et al. Data Report, (Exhibit DOM/IDN-2), paras. 184-185.
[77] See Ajzen et al. Data Report, (Exhibit DOM/IDN-2), paras. 165-173, Appendix A, pp. 90-92.
[78] See Ajzen et al. Data Report, (Exhibit DOM/IDN-2), para. 180.

47. Professor Klick, in his expert report submitted by Honduras, analyses the NTPPTS data on the effect of the TPP measures on quitting-related variables.[79] Professor Klick claims that Wakefield et al. 2015 did not present the results of some variables related to quitting that did not improve or might have even gotten worse with the TPP measures.[80]

48. Professor Klick presents the results of an ordered probit model controlling for the TPP measures (early TPP period and formal TPP period), gender, age, education, socio-economic status and a linear time trend.[81] He finds that the TPP measures had no statistically significant impact on stubbing out cigarettes because of thoughts about smoking harms.[82]

49. Professor Klick further contends that Durkin et al.'s 2015 results of the unadjusted models, which show that none of the variables related to pack concealment and micro-indicators of concerns are statistically distinguishable from the pre-TPP period, must be given appropriate weight and considered as at least equally valid compared with the results of the adjusted models for various reasons.[83]

50. As explained above, Australia's expert Professor Chaloupka argues that Ajzen et al.'s analysis suffer from several limitations, in part due to the fact that the NTPPTS data do not track non-smokers who might have taken up tobacco use in the absence of the TPP measures and are cross-sectional.[84] In that context, Professor Chaloupka submits that Ajzen et al. fail to recognize that given the analysis considers all current tobacco users and recent quitters, and not just the users for whom the TPP measures have influenced their more proximal outcomes, relatively large sample sizes are required to detect the increasingly smaller effects of more distal outcomes. According to Professor Chaloupka, it is therefore not surprising that by focusing on the sample of smokers least likely to have been influenced by the TPP measures, the findings on the impact of the TPP measures on quitting-related behaviour variables are relatively weak.[85]

51. Ajzen et al. contend that, contrary to what Professor Chaloupka submits, the weakening of the TPP measures across the chain of effect from pack appeal to behaviour is not explained by the NTPPTS data's limitation, but by the theoretically expected decline in the impact of the TPP measures.[86] They further argue that the findings based on the NTPPTS, whose cross-sectional nature is widely used and accepted, are supported by those based on the longitudinal ITC survey data.[87] Ajzen et al. also claim that the NTPPTS dataset has enough statistical power to detect proximal and distal outcomes.[88] Likewise, Ajzen et al. disagree with Professor Chaloupka and argue that the NTPPTS data do not underestimate changes in some measures of secondary indicators because the TPP measures have had no impact on quitting behaviours in the short term.[89]

### 2.1.2 International Tobacco Control Policy Evaluation Project

52. Yong et al. 2015 also use the ITC data to investigate the impact of the TPP measures on pack concealment and foregoing smoking behaviours. The authors present the results of a GEE model

[79] See Klick Supplemental Rebuttal Report, (Exhibit HND-122), paras. 69-73 and 88-100.
[80] See Klick Supplemental Rebuttal Report, (Exhibit HND-122), paras. 61-63.
[81] Professor Klick explains that, unlike Wakefield et al. 2015, he omits the measure of exposure to mass media anti-smoking messages, cigarette costliness and heaviness of smoking index, because these variables are endogenous. He further explains that the inclusion of these variables do not change the results he presented. See Klick Supplemental Rebuttal Report, (Exhibit HND-122), fn 35.
[82] See Klick Supplemental Rebuttal Report, (Exhibit HND-122), paras. 74-78 and 80-81. We note that in one of Professor Klick's specification, the impact of the early TPP period variable on stubbing out is negative and statistically significant.
[83] See para. 14 above for a full summary of Professor Klick's argument and Klick Supplemental Rebuttal Report, (Exhibit HND-122), paras. 91-100.
[84] See para. 16 above for a full summary of Professor Chaloupka's arguments; and Chaloupka Rebuttal Report, (Exhibit AUS-582), paras. 2-5 and 10.
[85] See Chaloupka Rebuttal Report, (Exhibit AUS-582), paras. 6-7 and 9.
[86] See Ajzen et al. Data Rebuttal Report, (Exhibit DOM/IDN-6), paras. 39-45.
[87] See Ajzen et al. Data Rebuttal Report, (Exhibit DOM/IDN-6), paras. 46-49, Appendix I.
[88] See Ajzen et al. Data Rebuttal Report, (Exhibit DOM/IDN-6), paras. 50-54, Appendix II.
[89] See Ajzen et al. Data Rebuttal Report, (Exhibit DOM/IDN-6), paras. 55-62, Appendix III.

controlling for age, sex, income, education, cigarettes per day, past year quit attempts, survey mode (phone vs. web) and wave of recruitment.[90]

53. Yong et al. 2015 find a statistically significant increase in the proportion of adult cigarette smokers that avoided health warning labels but no statistically significant change in the proportion of those foregoing cigarettes. Similarly, Yong et al. 2015 find that the subgroup of respondents that switched from initially focusing away to focusing on the health warning labels following the TPP measures' introduction avoided health warning labels, but did not forego cigarettes. Conversely, the subgroup of respondents, which chose to focus away from the health warning labels, avoided fewer health warning labels, but did not forego cigarettes.[91]

54. As explained in the review of post-implementation studies on quit intentions, Ajzen et al., in their expert report submitted by the Dominican Republic and Indonesia, argue that they could not undertake a comprehensive assessment of the ITC dataset, because they only had access to a small subset of the data.[92] They further claim that Yong et al. 2015 did not discuss effect sizes, analyse wear out effects and account for multiple hypothesis testing. They also criticise Yong et al. 2015 for failing to justify the outcome variables available in the ITC dataset they selected for their analysis and to report the results of some variables. Ajzen et al. replicate the analysis of Yong et al. 2015 by re-estimating GEE models controlling for the survey mode (phone vs. web) and wave of recruitment, as well as respondents' age, sex, income, education, cigarettes per day consumed and past year quit attempts.[93]

55. Ajzen et al. find that although the TPP measures have had a moderate positive and statistically significant impact on avoiding health warning labels, the TPP measures did not prompt smokers to forego cigarettes.[94]

56. Australia contends that Ajzen et al. correctly conclude that the TPP measures have significantly increased avoidance of GHWs.[95] However, Australia argues that when Ajzen et al., conclude that interest in and intentions to quit decreased post-plain packaging, they do not acknowledge or appear to appreciate that questions relating to quit intentions were not asked of recent quitters, which are the cohort whose intentions and behaviours were most likely to have been influenced by the TPP measures.[96] Australia concludes that consequently the sample Ajzen et al. used to analyse quit intentions was negatively biased. The Dominican Republic disagrees with Australia's argument. According to Ajzen et al., the exclusion of recent quitters could not have produced a biased sample because the ITC data show that the number of quitters was about the same before and after implementation of the TPP measures, implying that Australia's argument is contradicted by the data.[97] Further, Ajzen et al. consider it surprising to hear (for the first time) from Australia that the ITC dataset is "biased" in this regard because it was developed by the tobacco control community to measure the real-world impact of tobacco control measures, such as plain packaging, on smoking-related outcomes, including quit intentions.[98]

### 2.1.3 Personal pack display dataset

57. CCV undertook an observational survey of the prevalence of cigarette pack display and smoking in outdoor venues in Victoria and South Australia between October 2011 and April 2012, and again between October 2012 and April 2013 and between January and April 2014. The survey provides information on active smoking rates and personal display of cigarette packs on tables observed among patrons of public venues with outdoor seating (visible from the footpath) before and after the introduction of Australia TPP measures. Two peer-reviewed papers employed the personal pack display data.

---

[90] See Yong et al. 2015, (Exhibit DOM-382).
[91] See Yong et al. 2015, (Exhibit DOM-382).
[92] See Dominican Republic's comments on Australia's response to Panel question No. 196, paras. 409-414. See also Ajzen et al. Second Data Report, (Exhibit DOM/IDN-4), paras. 12-19.
[93] See Ajzen et al. Second Data Report, (Exhibit DOM/IDN-4), paras. 30-37 and 53-58.
[94] See Ajzen et al. Second Data Report, (Exhibit DOM/IDN-4), paras. 22-27.
[95] See Australia's response to Panel question No. 196, paras. 225 and 237.
[96] Australia's response to Panel question No. 196, para. 237.
[97] See Ajzen et al. Second Data Rebuttal Report, (Exhibit DOM/IDN-8), paras. 15-16.
[98] See Ajzen et al. Second Data Rebuttal Report, (Exhibit DOM/IDN-8), para. 6.

58. Zacher et al. published two papers assessing empirically the impact of the TPP measures on pack display, smoking and pack orientation. Zacher et al. 2014 use the personal pack display dataset to analyse the rates of pack display, smoking and pack orientation between October and April 2011-2012 (pre-TPP period) and 2012-2013 (post-TPP).[99] The authors report the results of a multi-level Poisson regression analysis. Zacher et al. 2014 find that the TPP measures reduced active smoking in outdoor areas of cafés, restaurants and bars, reduced personal pack display on tables, and increased steps taken by smokers to conceal packs that would otherwise be visible. In particular, the authors find that the decrease of pack display was stronger in venues with children present and limited to mid- and high- socio-economic status areas. Conversely, the decrease in packs orientated face-up was found to be stronger in low socio-economic status areas.

59. Zacher et al. 2015 re-analyse the rates of pack display, smoking and pack orientation by extending the period of analysis with a longer post-implementation period, between January and April 2012 (pre-TPP period), 2013 (early post-TPP period) and 2014 (one year post-TPP period).[100] The authors conclude that after one year Australia's TPP measures have reduced visibility of tobacco products and active smoking in public venues, particularly in the presence of children. In particular, the authors report a statistically significant decrease in pack display from pre-TPP to early post-TPP that remained low after one year following the introduction of the TPP measures. That being said, Zacher et al. 2015 find that the statistically significant change from pre-TPP to early post-TPP was not sustained one year post-TPP for packs orientated face-up, packs concealed by telephones, wallets or other items and for the use of external case.

60. Ajzen et al., in their expert report submitted by the Dominican Republic and Indonesia, reviewed the accuracy and completeness of the findings reported in Zacher et al. 2015. According to Ajzen et al., Zacher et al. 2015 failed to report or discuss effect sizes.[101] Ajzen et al. further argue that Zacher et al.'s explanation that pack changes may have made smokers less inclined to smoke in public is hard to reconcile with the finding that pack display did not decline among smokers, who continued to smoke at outdoor venues, and that smokers' pack avoidant behaviours vanished within a year. Ajzen et al. contend that Zacher et al.'s alternative explanation, according to which the small decline observed in outdoor smoking could reflect the overall declining trend in smoking in Australia, seems more consistent with the data. According to Ajzen et al., Zacher et al. 2015 acknowledge that the Personal Pack Display dataset is unable to establish whether the small decline observed in outdoor smoking was caused by the implementation of the TPP measures.[102] Ajzen et al. replicated Zacher et al's 2015 analytical approach and conclude that although there was a small decrease in packs displayed and active smoking at outdoor venues, pack-avoidant behaviours such as packs orientated face-up, packs concealed and use of external case, "had entirely vanished" one-year after the implementation of the TPP measures.[103]

61. Australia submits that there is no evidence before the Panel to support a finding that the impact of the TPP measures will wear out. Australia argues that Professor Ajzen relies on the two studies by Zacher et al. 2014, 2015 to support the claim of wear out, although Ajzen et al. concede that Zacher et al. 2014, 2015 demonstrate that there was a statistically significant decline in active smoking at outdoor venues after the introduction of tobacco plain packaging, an effect which was enhanced over time. Similarly, Australia argues that while the Dominican Republic and Indonesia contend that changes in the way packs are displayed at outdoor venues wore out over the course of the study, Ajzen et al. concede that there was a statistically significant and lasting decline in the total number of packs displayed.[104]

62. The Dominican Republic disagrees with Australia's claim that a statistically significant decline in outdoor smoking was enhanced over time and that the TPP measures have already been successful in reducing smoking prevalence. The Dominican Republic argues that outdoor smoking initially declined and then increased again, although it did not return to its earlier level. The Dominican Republic further submits that Zacher et al. 2015 themselves acknowledge that the finding on the level of outdoor smoking cannot be attributed to the TPP measures, because they

[99] See Zacher et al. 2014, (Exhibits AUS-222 (revised), JE-24(68), DOM-286).

[100] See Zacher et al. 2015, (Exhibits AUS-223 (revised), DOM-287).

[101] See Dominican Republic's second written submission, para. 487; response to Panel question No. 126, paras. 294-297; and comments on Australia's response to Panel question No. 196, paras. 416 and 419-423.

[102] See Ajzen et al. Second Data Report, (Exhibit DOM/IDN-4), paras. 42 and 45-47.

[103] See Ajzen et al. Second Data Report, (Exhibit DOM-IDN-4), paras. 38-47 and 59-60.

[104] See Australia's response to Panel question No. 196, paras. 246-247.

did not control for a host of unrelated factors that could explain the decline in outdoor smoking levels.[105]

### 2.1.4 Australian Secondary Students Alcohol Smoking and Drug Survey

63. White et al. 2015b use the ASSAD survey data to examine the impact of the TPP measures on foregoing behaviours by students aged 12-17 years. The authors estimate a logistic regression model controlling for age, sex, school type, state and smoking status.[106]

64. White et al. 2015b find that the TPP measures have had no impact on the proportion of students, who had seen a cigarette pack in the previous six months, and that did not have a cigarette because of the health warnings. The authors conclude that the data suggest that the introduction of the TPP measures did not induce adolescents to attend to and process warnings on cigarette packs to a greater extent than when GHWs covered 30% of the front of a fully branded pack.[107]

65. The Dominican Republic submits that it could not make an objective assessment of White et al.'s findings without full access to the ASSAD survey data. The Dominican Republic further argues that although in a different study White et al. 2015a conclude that the TPP measures have reduced the visual appeal of packs among adolescents[108], that reduced appeal does not have any meaningful impact on quit intentions and secondary indicators of quitting.[109]

### 2.1.5 Cigar and cigarillo smokers surveys

66. Miller et al. 2015 use the responses from a series of interviews, focus groups and an online survey of current cigar and cigarillo smokers and find that premium cigar and cigarillo smokers, who were exposed to the TPP measures, tried to avoid the health warning labels by decanting the individual cigars and cigarillos from the box to a humidor or an unbranded tin. Similarly, they report an increase in pack concealment among non-premium cigarillo smokers.[110]

67. Ajzen et al., in their expert report submitted by the Dominican Republic and Indonesia, consider that the conclusions that can be drawn from Miller et al.'s results are limited by several methodological shortcomings, such as the non-representativeness of survey participants, unsuitability to draw causal inferences from focus groups and interviews, failure to control for confounding factors, and absence of "baseline" information collected before the TPP measures' implementation. They further argue that Miller et al. 2015 failed to report that the online survey participants in the online survey reported no change in stubbing out cigars or cigarillos and in stopping smoking. More generally, Ajzen et al. submit that any findings about the TPP measures on cigarette-related cognitions and behaviours cannot be generalized to cigars because smoking cigarettes and smoking cigars are different behaviours, with potentially different determinants.[111]

### 2.1.6 Commissioned Roy Morgan Research Survey (Australia and New Zealand)

68. Professor Klick uses the Roy Morgan Research data to estimate a difference-in-difference logit model that explains the respondents' experiences concealing their packs in Australia and

[105] See Ajzen et al. Second Data Rebuttal Report, (Exhibit DOM/IDN-8), paras. 17-21; Dominican Republic's second written submission, para. 487; Dominican Republic's response to Panel question No. 126, paras. 294-297; and Dominican Republic's comments on Australia's response to Panel question No. 196, paras. 416 and 419-423.

[106] See White et al. 2015b, (Exhibits HND-135, DOM-236, DOM-288).

[107] See White et al. 2015b, (Exhibits HND-135, DOM-236, DOM-288), p. ii56.

[108] See Appendix A, paras. 20-22 and 80-81 (reviewing White et al. 2015a, (Exhibits AUS-186, DOM-235)).

[109] See Dominican Republic's second written submission, paras. 443-456; and comments on Australia's response to Panel question No. 196, para. 390.

[110] See Miller et al. 2015, (Exhibits AUS-102, DOM-315).

[111] See Dominican Republic's second written submission, paras. 457-462. See also Ajzen et al. Data Report, (Exhibit DOM/IDN-2), paras. 251-262.

New Zealand, controlling for the TPP measures, an Australian baseline variable, a common baseline variable and a post-TPP implementation time-period.[112]

69. Professor Klick concludes that the TPP measures have had no statistically significant effect on smokers' preferences or feelings about being seen with a plain pack in Australia relative to a non-plain pack in New Zealand.[113] In particular, Professor Klick finds no statistically significant impact of the TPP measures in Australia relative to "regular" packs in New Zealand on the likelihood of respondents reporting they had placed a cover over the cigarettes, placed the cigarettes in a different container or kept the pack out of sight.[114] Professor Klick further reports no statistically significant impact of the TPP measures in Australia relative to New Zealand on the likelihood the respondents successfully giving up smoking for more than a month, switching to a brand with lower tar or nicotine, reducing the number of cigarettes smoked in a day, or recently stopping smoking for 24 hours at least once during the last three months. In addition, he reports no statistically significant impact on the likelihood of the respondents increasing the number of cigarettes smoked, unsuccessfully trying to switch to a brand with lower tar or nicotine or unsuccessfully trying to reduce the number of cigarettes smoked in a day.[115]

70. Although Australia's expert Dr Chipty does not address directly Professor Klick's analysis of pack concealment behaviours based on the Roy Morgan Research data, she raises several criticisms of Professor Klick's difference-in-difference analysis of smoking incidence, which also apply to his analysis of pack concealment behaviours. Dr Chipty contends that Professor Klick's analysis of the Roy Morgan Research data is invalid because the Roy Morgan survey does not contain a pre-period and is unable to distinguish which respondents had noticed plain packaging. Dr Chipty further argues that New Zealand is an invalid counterfactual for the purposes of studying the effects of plain packaging, because New Zealand increased its excise tax in January 2013, one month after the introduction of the TPP measures.[116]

### 2.2 Analysis by the Panel

71. The parties submitted five peer-reviewed papers that investigate empirically the impact of the TPP measures and enlarged GHWs on pack concealment and micro-indicators of concern. While four of these studies use responses of survey data on adult smokers, cigar and cigarillo smokers or adolescents, the two papers by Zacher et al. 2014, 2015 analyse actual observed behaviours of pack concealment among adult smokers. In addition, an expert report prepared by Professor Klick, submitted by Ukraine, contains a study assessing the impact of the TPP measures on pack concealment.[117]

72. A careful assessment of Durkin et al. 2015, Yong et al. 2015 and Zacher et al. 2014, 2015 and the econometric evidence submitted by the Dominican Republic, Honduras and Indonesia suggests that the TPP measures have increased pack concealment among adult cigarette smokers. However, empirical evidence of the impact of the TPP measures on stubbing out cigarettes before finishing them due to thoughts about the harms caused by smoking and stopping smoking among adult cigarette smokers is much more limited and mixed.

73. We note that Durkin et al. 2015 find some evidence that the TPP measures have had a positive and statistically significant impact on pack concealment and stubbing out cigarettes but no impact on stopping smoking among adult cigarette smokers in the first year of implementation of the TPP measures.[118] Ajzen et al., applying an alternative estimation methodology and correcting for multiple testing bias, reverse the results reported in Durkin et al. 2015 and find that the TPP measures have had no impact on pack concealment, stubbing out or stopping smoking.[119]

---

[112] See Appendix A, paras. 62-63 and Appendix B, para. 33 for a description of the Roy Morgan Research data.

[113] See Klick Report, (Exhibit UKR-5), pp. 18 and 54-56.

[114] See Klick Report, (Exhibit UKR-5), pp. 18 and 54-56.

[115] See Klick Report, (Exhibit UKR-5), pp. 18 and 37-44.

[116] See Chipty Rebuttal Report, (Exhibit AUS-535) (SCI), paras. 54-71.

[117] See section 1.6.6 in the main body of these Reports for a description of Ukraine's participation in these proceedings.

[118] See Durkin et al. 2015, (Exhibits AUS-215 (revised), DOM-305).

[119] See Ajzen et al. Data Report, (Exhibit DOM/IDN-2), paras. 188-198. We note that a comparison of the results with and without the trend variable shows that the coefficient of the constant variable is very large

Similarly, Professor Klick, using a different model specification, finds that the TPP measures had no impact on stubbing out. We note that Australia did not challenge Ajzen et al.'s and Professor Klick's alternative estimation methodology and results, but generally considers that the structure of the NTPPTS is likely to be more suited to detecting changes in proximal outcomes than in more distal variables, such as quitting-related behaviours.[120] As explained in our review of the evidence on quitting-related cognition, we are not persuaded by Professor Klick's claim that unadjusted models, which do not control for individual characteristics and tobacco control policies, should be considered at least equally valid compared to adjusted models, noting that Ajzen et al. focused their analysis on adjusted models and Professor Klick presented econometric results for stubbing out based on an adjusted model that controls for the respondent's individual characteristics.[121]

74. We observe that the results presented in Yong et al. 2015 are partially in line with those in Durkin et al. 2015. Yong et al. 2015 report a positive and statistically significant impact of the TPP measures on avoiding the health warning labels and a statistically non-significant impact on foregoing among adult smokers.[122] We note that Ajzen et al. confirm that the TPP measures have had a positive and statistically significant impact on avoiding GHWs and that they qualify it as a medium effect.[123]

75. The empirical evidence regarding the positive impact of the TPP measures on smokers' avoidant responses published in Durkin et al. 2015 and Yong et al. 2015 is to some extent corroborated by Zacher et al.'s 2014, 2015 papers, who suggest that one year after the introduction of the TPP measures, the display of personal pack in outdoor areas of cafés, restaurants and bars has decreased, particularly in the presence of children.[124] Zacher et al. 2015 find, however, that the reduction in the proportion of packs oriented face-up and the increase in concealed packs by telephones, wallets or other items and in the use of external cases was not sustained one year after the introduction of the TPP measures. We note that Ajzen et al.'s analysis of the Personal Pack Display dataset confirms all the findings reported in Zacher et al. 2015 but that they qualify the impact for both packs displayed and active smoking at outdoor venues as small.[125]

76. Similarly, the limited available evidence on adolescents and cigar and cigarillo smokers corroborates part of the findings reviewed above. Although White et al. 2015b provide no empirical evidence on pack concealment among young smokers, they find that the TPP measures have had no statistically significant impact on the frequency of not having a cigarette because of the health warnings[126], which is in line with the findings on foregoing smoking reported in Durkin et al. 2015 and Yong et al. 2015. The descriptive statistics analysis by Miller et al. 2015 suggests that the share of premium cigar and cigarillo smokers and of non-premium cigarillo smokers reporting decanting the cigar from the boxed to a humidor or concealing their pack have increased[127], which is partially in line with the results reported in Durkin et al. 2015, Yong et al. 2015 and Zacher et al. 2014, 2015. As explained in more detail in Appendix A, we see no basis to reject entirely Miller et al.'s findings on the basis of Ajzen et al.'s criticism.[128]

77. Finally, as explained above and in our review of Professor Klick's analysis of GHWs effectiveness and smoking incidence based on the commissioned Roy Morgan Research data, we question the validity of the pre-period of the commissioned data and the use of New Zealand as a

when the trend variable is included in the model specification. The coefficient value of the other variables is usually marginally affected by removing the trend variable.

[120] See Australia's comments on the complainants' responses to Panel question No. 197, paras. 371-375.

[121] Professor Klick further acknowledges that the results he reported do not change when the variables for addiction level, exposure to anti-smoking television advertising and cigarette costliness are included in the model specification, which would suggest that the risk of endogenous bias might be limited.

[122] See Yong et al. 2015, (Exhibit DOM-382).

[123] See Ajzen et al. Second Data Report, (Exhibit DOM/IDN-4), paras. 22-27.

[124] See Zacher et al. 2014, (Exhibits AUS-222 (revised), JE-24(68), DOM-286); and Zacher et al. 2015, (Exhibit AUS-223 (revised), DOM-287).

[125] See Ajzen et al. Second Data Report, (Exhibit DOM/IDN-4), para. 44.

[126] See White et al. 2015b, (Exhibits HND-135, DOM-236, DOM-288).

[127] See Miller et al. 2015, (Exhibits AUS-102, DOM-315).

[128] See Appendix A, paras. 32 and 71 above.

counterfactual, and therefore the overall robustness of Professor Klick's results based on these data.[129]

## 3 EVIDENCE RELATING TO QUIT ATTEMPTS SINCE THE ENTRY INTO FORCE OF THE TPP MEASURES

78. Two peer-reviewed papers investigate empirically the impact of Australia's TPP measures on quit attempts: (i) Durkin et al. 2015 and Young et al. 2014.[130] The expert report by Professor Klick, submitted by Ukraine, also contains an analysis of the impact of the TPP measures on quitting.[131]

### 3.1 Datasets and related studies

#### 3.1.1 National Tobacco Plain Packaging Tracking Survey

79. Durkin et al. 2015 use the NTPPTS data to assess the impact of the TPP measures on short-term changes in quitting-related behaviours. The authors estimate various logistic models of respondents' experiences with quit attempts by adjusting for baseline levels and controlling for the date of the follow-up survey, number of days between baseline and follow-up survey, anti-smoking television advertising in the three months prior to the follow-up survey, change in cigarette costliness, addiction level and demographic characteristics.[132] Durkin et al. 2015 also present the results of unadjusted models, where the only control variable is the baseline response.

80. Durkin et al. 2015 find that the TPP measures were associated with increased rates of quit attempts among adult cigarette smokers. In particular, the authors report that compared to the referent group of smokers who completed their follow-up survey in the pre-TPP period, smokers who were followed-up in the early transition period showed significantly greater increases in quit attempt in the past month. Similarly, smokers who were followed-up in the first year of TPP showed greater increases in quit attempt in the past month.[133]

81. Ajzen et al., in their expert report submitted by the Dominican Republic and Indonesia, claim that Durkin et al.'s conclusions are unwarranted and contradicted by the data. According to Ajzen et al., Durkin et al. 2015 overstate their findings because they failed to report the results showing that the TPP measures did not bring about any increase in respondents' attempts to ever quit, to quit more than once, and to quit for more than one week between baseline and follow-up.[134] More generally, Ajzen et al. criticize Durkin et al. 2015 for having adopted a very different and unorthodox methodology than in Wakefield et al. 2015, despite being the same six authors using the same dataset. In their view, Durkin et al.'s analytical approach suffers from three serious methodological shortcomings, namely controlling for a non-existent daily trend, not investigating wear-out effects and not correcting for multiple significances tests.[135] Ajzen et al. further argue that, contrary to Durkin et al.'s claim, relying on changes in the responses of participants from month to month does not minimize the influence of sampling variation and increase the power of the statistical tests.[136]

82. Ajzen et al. estimate a logistic model for dichotomized outcome variables, a linear model for continuous outcome variables and an ordered logit model for categorical outcome variables, which include the same covariates considered by Durkin et al. 2015 except for the dates of the follow-up interviews. They find that the TPP measures have had virtually no significant effect on quit attempts among adult cigarette smokers. In particular, they find no statistically significant change in the proportion of adult smokers who were followed-up in the first year of TPP that attempted to

[129] See Appendix A, para. 72 above.
[130] See Durkin et al. 2015, (Exhibits AUS-215 (revised), DOM-305); Young et al. 2014, (Exhibits AUS-214, JE-24(67), DOM-285).
[131] See section 1.6.6 in the main body of these Reports for a description of Ukraine's participation in these proceedings.
[132] See Durkin et al. 2015, (Exhibits AUS-215 (revised), DOM-305).
[133] See Durkin et al. 2015, (Exhibits AUS-215 (revised), DOM-305).
[134] See Ajzen et al. Data Report, (Exhibit DOM/IDN-2), paras. 174-180.
[135] See Ajzen et al. Data Report, (Exhibit DOM/IDN-2), paras. 188-198.
[136] See Ajzen et al. Data Report, (Exhibit DOM/IDN-2), paras. 184-185.

ever quit, attempted more than once to quit, and attempted to quit for more than one week, which Durkin et al. 2015 failed to report. The only exception is a positive and statistically significant but short-lived effect of the TPP measures on calls to Quitline to help with the last quit attempt, which was also not published in Durkin et al. 2015. Ajzen et al. further find that for the quit-related variables considered in Durkin et al. 2015, the TPP measures have statistically increased the proportion of adult smokers who attempted to quit in past month in the linear and ordered logistic regressions, but in the logistic regression the TPP measures have reduced their attempts to quit in past month.[137]

83. Professor Klick submits that the results of the unadjusted models, which show that none of the variables related to quit attempts are statistically distinguishable from the pre-TPP period, must be given appropriate weight and considered as at least equally valid compared with the results of the adjusted models for various reasons.[138]

84. As mentioned in the review of quitting-related cognitions and pack concealment, Professor Chaloupka, Australia's expert, submits that Ajzen et al.'s analysis suffers from several limitations that are partly related to the structure of the NTPPTS data, which do not include non-smokers who might have taken up tobacco use in the absence of the TPP measures and are cross-sectional.[139] In that context, Professor Chaloupka argues that Ajzen et al. do not recognize that relatively large sample sizes are required to detect the increasingly smaller effects of more distal outcomes, because the sample includes all current tobacco users and recent quitters, and not just the users influenced by the TPP measures in terms of more proximal outcomes. Professor Chaloupka is of the view that it is not surprising that the findings on the impact of the TPP measures on quitting-related behaviour variables are relatively weak when one considers the sample of smokers, who are least likely to have been influenced by the TPP measures.[140]

85. Ajzen et al. disagree with Professor Chaloupka and claim that the declining impact of the TPP measures across the chain of effect is not due to the NTPPTS data's limitation but to the theoretically expected decline in the impact of the TPP measures.[141] They further contend that the findings based on the NTPPTS, whose cross-sectional nature is widely used and accepted, are corroborated by those obtained using the longitudinal ITC survey data.[142] Ajzen et al. are also of the view that the NTPPTS dataset has enough statistical power to detect proximal and distal outcomes.[143] Ajzen et al. claim that given the TPP measures have had no impact on quitting behaviours in the short term, the NTPPTS data do not underestimate changes in quit attempts.[144]

### 3.1.2 International Tobacco Control Policy Evaluation Project

86. In their review of Yong et al. 2015, who use the ITC data to analyse whether the TPP measures are associated with increased desirable reactions towards the health warning labels, Ajzen et al., in their expert report submitted by the Dominican Republic and Indonesia, criticise the authors for not reporting the results of actual smoking behaviour. They reconsider the ITC data and estimate a GEE model of the decision to attempt to quit in the last 12 months controlling for the survey mode (phone vs. web) and wave of recruitment, as well as the respondents' ages, sex, income, education, cigarettes per day consumed and past year quit attempts.[145]

[137] See Ajzen et al. Data Report, (Exhibit DOM/IDN-2), paras. 165-173, Appendix A, pp. 90-92.
[138] See para. 14 above for a full summary of Professor Klick's argument and Klick Supplemental Rebuttal Report, (Exhibit HND-122), paras. 91-100.
[139] See para. 15 above for a full summary of Professor Chaloupka's arguments; and Chaloupka Rebuttal Report, (Exhibit AUS-582), paras. 2-5 and 10.
[140] See Chaloupka Rebuttal Report, (Exhibit AUS-582), paras. 6-7 and 9; and Chaloupka Third Rebuttal Report, (Exhibit AUS-604), paras. 2-7.
[141] See Ajzen et al. Data Rebuttal Report, (Exhibit DOM/IDN-6), paras. 39-45.
[142] See Ajzen et al. Data Rebuttal Report, (Exhibit DOM/IDN-6), paras. 46-49, Appendix I.
[143] See Ajzen et al. Data Rebuttal Report, (Exhibit DOM/IDN-6), paras. 50-54, Appendix II.
[144] See Ajzen et al. Data Rebuttal Report, (Exhibit DOM/IDN-6), paras. 55-62, Appendix III.
[145] See Ajzen et al. Second Data Report, (Exhibit DOM/IDN-4), paras. 30-37 and 53-58.

87. Ajzen et al. find no statistically significant change in the proportion of adult smokers that attempted to quit in the last 12 months. They note that this result was not reported in Yong et al. 2015.[146]

88. Professor Chaloupka is of the view that Ajzen et al. do not recognize that the impact of the TPP measures should be smaller for more distal outcomes, when one looks at the impact in the overall sample of smokers and recent quitters, because a smoker, whose likelihood of noticing health warnings did not increase following the introduction of the TPP measures, would not be expected to show any change in his/her tobacco use behaviour.[147]

89. Ajzen et al. argue that Professor Chaloupka does not contest the accuracy of their analytical approach and results, and does not question their serious criticisms of Yong et al. 2015.[148] They state that Professor Chaloupka's sole criticism regarding the ITC analysis seeks to denigrate the quality of the ITC dataset itself, claiming that it provides a "biased" sample to measure the impact on quit intentions and quit interests. According to Ajzen et al., Professor Chaloupka's argument assumes that the TPP measures led smokers to quit shortly after their introduction, and that these "recent quitters" were subsequently excluded from responding to questions on quit intentions and quit interests in the ITC survey. However, they state that the question on "quit attempt in the last 12 months" was asked of smokers and recent quitters, and, moreover, the data show that no additional quit attempts were made post-implementation.[149]

### 3.1.3 Quitline calls

90. The Quitline calls dataset reports the weekly number of calls to the quit smoking hotline Quitline in New South Wales and the Australian Capital Territory between 1 March 2005 and 7 April 2013.[150] Young et al. 2014 use the Quitline calls data to investigate the impact of the introduction of the TPP measures on Quitline calls by comparing this number to the nationwide introduction of GHWs on cigarette packaging in 2006. The authors estimate an autoregressive integrated moving average (ARIMA) model controlling for seasonal variation, anti-smoking advertising activity, number of smokers in the population and cigarette costliness.[151]

91. Young et al. 2014 find a sustained increase in calls to the Quitline, which occurred four weeks after the introduction of the TPP measures and lasted for 43 weeks. The authors further report that the positive impact of the TPP measures on calls to the Quitline has continued for a longer period than for the 2006 GHWs.

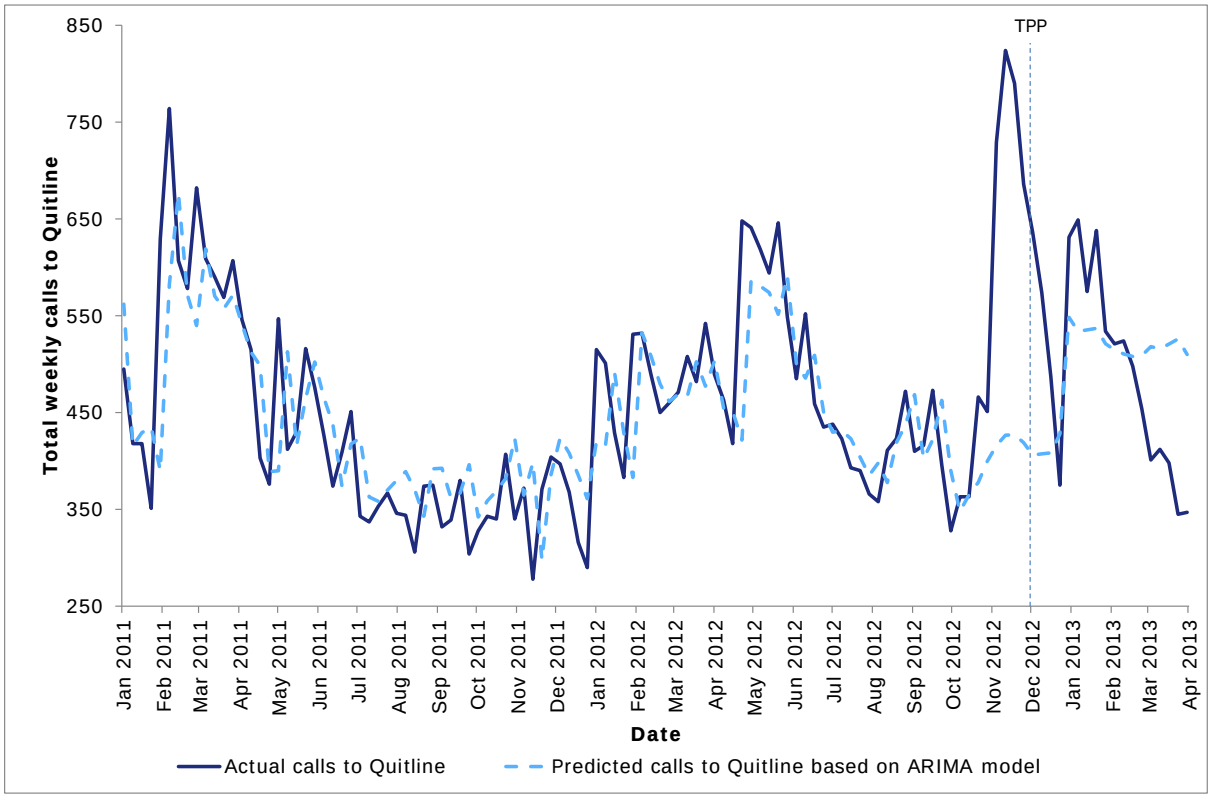
92. Ajzen et al. re-analyse the Quitline Calls data by applying a different methodology from the one used by Young et al. 2014. They present the results of a graphical analysis as well as an "event study" analysis comparing the "actual" volume of calls to the Quitline with the "forecast" volume of calls in the pre-implementation period obtained by estimating a model controlling for seasonal variation, anti-smoking advertising, number of smokers and cigarette costliness.

93. Ajzen et al. find that, as reported in **Figure B.1**, there was a statistically significant increase in the number of calls to the Quitline after the introduction of the TPP measures, which occurred approximately three weeks before the sale of plain packs with enlarged GHWs became mandatory. However, they report that the increase in the number of calls was only significant in the first 13 week period following the introduction of the TPP measures. In the second 13 week period, the level of calls dropped quickly to that of the calls level prevailing in the pre-implementation period. Ajzen et al. argue that Young et al. 2014 came to qualitatively similar conclusions using an alternative empirical strategy and controlling for the same confounding factors. They posit that, based on Young et al.'s findings, by the week of mandatory implementation (1 December 2012), the level of calls had dropped by more than 36%; by the end of 2012, the level of calls had

[146] See Ajzen et al. Second Data Report, (Exhibit DOM/IDN-4), paras. 24 and 28.
[147] See Chaloupka Rebuttal Report, (Exhibit AUS-582), para. 4.
[148] See Dominican Republic's comments on Australia's response to Panel question No. 196, paras. 409-414; and Ajzen et al. Second Data Rebuttal Report, (Exhibit DOM/IDN-8), paras. 13-16.
[149] See Ajzen et al. Second Data Rebuttal Report, (Exhibit DOM/IDN-8), paras. 15-16.
[150] See Dominican Republic's second written submission, paras. 434-442.
[151] See Young et al. 2014, (Exhibits AUS-214, JE-24(67), DOM-285).

dropped 65% compared with the peak week; and by the end of July 2013, there was no longer one extra call per week left compared to the pre-implementation period.[152]

**Figure B.1: Event Study of Calls to the Quitline[153]**




Note: The vertical dashed line indicates the introduction of TPP and enlarged GHWs.

Source: Ajzen et al. Data Report, (Exhibit DOM/IDN-2), p. 62.

94. Professor Chaloupka argues that Ajzen et al.'s conclusion that calls to the Quitline rose sharply following the implementation of the TPP measures, and then eventually returned to pre-implementation levels, does not mean that the effect of the TPP measures has "worn out". Referring to two recent econometric analyses that report strong statistically significant effects of tobacco tax increases in driving increased calls volume in the United States, Professor Chaloupka submits that the decline in call volume in the months following a tax increase does not imply that the effects of tax and price increases on tobacco use "wear out" over time. According to Professor Chaloupka, the extensive economic research that has looked at the short- and long-run effects of tax and prices on tobacco use concludes that estimates for the long-run impact of tax and price increases exceed estimates for the short-run impact, implying that the effects grow over time. Professor Chaloupka is of the view that rather than "wearing out" over time, the effects of the TPP measures are likely to grow over time as new cohorts of young people are less likely to take up tobacco use in the absence of branded packaging.[154]

95. Ajzen et al. counter that Professor Chaloupka draws a faulty analogy with the effect of a tax increase on smoking. Ajzen et al. submit that for a change in smoking behaviour to persist after Quitline calls have receded to prior levels, the intervention must reduce smoking behaviour in the first place. In the case of the TPP measures, Ajzen et al. contend that given that the TPP measures

[152] See Ajzen et al. Data Report, (Exhibit DOM/IDN-2), paras. 237-245.

[153] We note that Figure 8 in Ajzen et al. Data Report, (Exhibit DOM/IDN-2, p. 62) includes the vertical line representing the TPP measures is set to 1 October 2012.

[154] See Australia's response to Panel question No. 196, paras. 238 and 245; Australia's comments on the complainants' responses to Panel question No. 197; and Chaloupka Rebuttal Report, (Exhibit AUS-582), para. 15.

have had no impact on consumption, quitting, relapse, or prevalence, there is no change in behaviour that could persist over time, after the increase in calls to the Quitline has vanished.[155]

### 3.1.4 Commissioned Roy Morgan Research Survey (Australia and New Zealand)

96. Professor Klick uses the Roy Morgan Research data to estimate a difference-in-difference logit model of quit-related behaviours in Australia and New Zealand, controlling for plain packaging, an Australian baseline, a common baseline and a post plain packaging implementation time-period.

97. Professor Klick reports no statistically significant impact of the TPP measures in Australia relative to New Zealand on the likelihood that the respondents recently attempted to quit smoking.[156]

98. According to Dr Chipty, Professor Klick's analysis of the Roy Morgan Research data is invalid because the Roy Morgan survey does not include a pre-period and does not make it possible to distinguish which respondents noticed plain packaging. Dr Chipty further submits that New Zealand is an invalid counterfactual for the purposes of studying the effects of the TPP measures, because of New Zealand's excise tax increase implemented one month after the introduction of the TPP measures.[157]

## 3.2 Analysis by the Panel

99. We note that only three peer-reviewed studies analyse empirically the impact of the TPP measures and enlarged GHWs on quit attempts among adult smokers. In addition, the expert report prepared by Professor Klick, submitted by Ukraine, investigates the impact on quit attempts.[158] We note that none of the peer-reviewed papers that focus on adolescents or cigar and cigarillo smokers provide post-implementation evidence on quit attempts.

100. A careful review of Durkin et al. 2015, Yong et al. 2015 and Young et al. 2014 studies and the econometric evidence submitted by the Dominican Republic and Indonesia suggests that although the TPP measures have increased calls to the Quitline, empirical evidence of the impact on quit attempts is very limited and mixed.[159]

101. We note that, based on the NTPPTS data, Durkin et al. 2015 report evidence that the TPP measures have had a positive and statistically significant impact on quit attempts among adult smokers in the first year of the TPP.[160] Ajzen et al., applying an alternative estimation methodology and correcting for multiple testing bias, reverse the results reported in Durkin et al. 2015 and find that the TPP measures have had no impact on attempting to ever quit, attempting to quit more than once and attempting to quit for more than a week.[161] We note, however, that very few explanatory variables, and sometimes only two variables, are statistically significant at 5% and 10% in some of the model specifications for the variables quitting more than once and quitting for more than a week. We further note that Ajzen et al.'s own results of the linear and ordered logistic regressions confirm Durkin et al.'s finding of a positive and statistically significant impact on quit attempts. Ajzen et al. qualify this impact as small without evidence of any wear-out effect. Conversely, Ajzen et al.'s results of the linear regression suggest that the TPP measures have reduced quit attempts. We note that Ajzen et al. did not provide any rationale that would explain why the TPP measures would lead to reduced quit attempts.[162] Using the ITC dataset,

[155] See Dominican Republic's response to Panel question No. 126, para. 293; Dominican Republic's second written submission, paras. 411, 434-442 and 447; Dominican Republic's comments on Australia's response to Panel question No. 196, paras. 424-425; and Ajzen et al. Data Rebuttal Report, (Exhibit DOM/IDN-6), paras. 73-74.

[156] See Klick Report, (Exhibit UKR-5), pp. 18 and 37-44.

[157] See Chipty Rebuttal Report, (Exhibit AUS-535) (SCI), paras. 54-71.

[158] See section 1.6.6 in the main body of these Reports for a description of Ukraine's participation in these proceedings.

[159] See Ajzen et al. Second Data Report, (Exhibit DOM/IDN-4), paras. 24 and 28.

[160] See Durkin et al. 2015, (Exhibits AUS-215 (revised), DOM-305).

[161] See Ajzen et al. Data Report, (Exhibit DOM/IDN-2), paras. 165-173, Appendix A, pp. 90-92.

[162] As explained in paras. 38 and 73 we are not persuaded by Professor Klick's claim that unadjusted models, which do not control for individual characteristics and tobacco control policies, should be considered at least equally valid compared with adjusted models.

Ajzen et al. also report a non-statistically significant impact of the TPP measures on quit attempts. A review of this result shows, however, that very few explanatory variables in that specification are statistically significant at 5% and 10%.[163] As noted previously, Australia did not challenge Ajzen et al.'s alternative estimation methodology and results based on the NTPPTS and ITC data.[164] Australia generally submits, however, that the structure of both NTPPTS and ITC data are likely to be less suited to detecting changes in more distal variables, such as quitting-related behaviours.[165]

102. As explained previously in our review of Professor Klick's analysis of GHW effectiveness and smoking incidence based on the commissioned Roy Morgan Research data, it is unclear how robust Professor Klick's results are given the issues regarding the validity of the pre-period of the commissioned data and the use of New Zealand as a counterfactual.[166]

103. We note that, unlike the impact on quit attempts, the empirical evidence on the impact of the TPP measures on calls to the Quitline is unambiguous. Both Young et al. 2014 and Ajzen et al. find that there was a statistically significant increase in calls to the Quitline after the introduction of the TPP measures.[167] The only main difference between the results reported in Young et al. 2014 and Ajzen et al. is that in Young et al. 2014 the estimated TPP impact on calls to the Quitline lasted 43 weeks and was considered by "sustained" by the authors, while in Ajzen et al. the estimated TPP impact was found to last 13 weeks and qualified as "short-lived". We note that none of the explanatory variables in Ajzen et al.'s ARIMA model specification are statistically significant at 5% or 10%, except the New Year variable and the first-order autoregressive term.[168] We further note that in the pre-implementation period most of the predicted Quitline calls obtained from the ARIMA model are not close to the actual level of Quitline calls and tend to lag behind the observed level of calls. This could explain why Ajzen et al. find that the increase in Quitline calls occurred three weeks before the mandatory implementation of the TPP measures. In any event, we are not persuaded that a decline in the volume of Quitline calls following an increase in calls immediately after the introduction of the TPP measures would necessarily imply that the impact of the TPP measures on tobacco use would wear out, since such Quitline calls reflect effects of the TPP measures on *existing* smokers, and would not inform their effect on those would-be smokers who abstain from tobacco use as a result of the TPP measures.

## 4 EVIDENCE RELATING TO THE PREDICTIVE RELATIONSHIPS BETWEEN PROXIMAL OUTCOMES AND DISTAL OUTCOMES SINCE THE ENTRY INTO FORCE OF THE TPP MEASURES (NATIONAL TOBACCO PLAIN PACKAGING TRACKING SURVEY)

### 4.1 Datasets and related studies

104. Brennan et al. 2015 use the NTPPTS data to examine the predictive relationships between proximal outcomes (i.e. tobacco appeal, GHW effectiveness and pack misleading) and distal

[163] We note that in their review of Scollo et al. 2015b (Exhibit CUB-80), Ajzen et al. also present the results of several logistic, linear, and ordered logit models for various measures related to quitting and relapse. They find that there was no statistically significant change in the proportion of adult smokers that quitted for more than one month or successfully quitted between baseline and follow-up. Similarly, they find that there was no statistically significant change in the proportion of adult ex-smokers who relapsed, still abstained from smoking at follow-up or stayed quit for more than one week at follow-up. We note that in several of these results, only a few explanatory variables in the model specification are statistically significant. In some cases, such as the linear and ordered logistic models for the relapse variable, none of the explanatory variables are statistically significant, which suggests that the model might be misspecified given the low coefficient of determination. The results of the logistic model also suggest that the TPP measures have had a negative and statistically significant change at 10% in the proportion of adult smokers that quit for more than one month, did not relapse and still remained quit. See Ajzen et al. Data Report, (Exhibit DOM/IDN-2), paras. 221-224, Appendix A pp. 95-97 and backup material.

[164] A comparison of the results with and without the trend variable shows that the coefficient of the constant variable is very large when the trend variable is included in the model specification. The coefficient value of the other variables is usually marginally affected by removing the trend variable.

[165] See Australia's response to Panel question No. 196, para. 237; and Australia's comments on the complainants' responses to Panel question No. 197, paras. 371-375.

[166] See Appendix A, para. 72 above.

[167] See Young et al. 2014, (Exhibit AUS-214, JE-24(67), DOM-285); and Ajzen et al. Data Report, (Exhibit DOM/IDN-2), paras. 237-245.

[168] See Ajzen et al. Data Report, (Exhibit DOM/IDN-2), backup material.

outcomes (i.e. quitting-related thinking and behaviours).[169] The authors present the results of a logistic regression analysis testing whether baseline measures of cigarette appeal, GHW effectiveness, perceived harm and concern/enjoyment predicted each of seven follow-up measures of quitting-related cognitions and behaviours, adjusting for baseline levels of the outcome and controlling for the date of the follow-up survey, number of days between the baseline and follow-up survey, anti-smoking television advertising, change in cigarette costliness, sex, age, education, socio-economic status and addiction level.

105. Brennan et al. 2015 find that some of the appeal-related variables, namely pack dislike and lower satisfaction, are prospectively associated with thoughts about quitting. Similarly, the authors find that several indicators of GHW effectiveness, such as noticing GHWs first, believing dangers of smoking are not exaggerated, and attributing much more motivation to quit to GHWs, have positively and significantly predicted the likelihood that smokers reported thinking daily about quitting, intending to quit and setting a firm date to quite. Similarly, they report a statistically significant association between concealing packs and daily thought about quitting as well as between feeling more smoking-related concern than enjoyment and daily thoughts about quitting and intention to quit. However, the authors find no statistically significant association between lower pack appeal, lower quality and lower value for money, on the one hand, and quitting-related cognition variables, on the other hand. They also report no statistically significant association between brand perception variables and quitting-related cognition variables.

106. Turning to pack concealment and micro-indicators of concern, Brennan et al. 2015 find that among all the appeal-related variables, only lower satisfaction is statistically associated with stubbing out and stopping smoking. The authors further find that several measures of GHWs effectiveness, namely noting the GHW first, attributing much more motivation to quit to graph health warning, concealing packs and requesting different GHWs, positively and significantly have predicted the likelihood that smokers at the following-up reported stubbing out. Similarly, they find a statistically significant association between not believing dangers of smoking are exaggerated and attribution of much more motivation to quit to GHW, on the one hand, and stopping oneself from smoking on the other hand. They also report that stubbing out and stopping smoking are predicted by feeling more smoking-related concern than enjoyment. However, the authors find that none of the proximal outcomes variables, such as disliking packs, lower pack appeal, lower quality, lower value for money, believing brands do not differ in prestige, not believing dangers of smoking are exaggerated, and attributing much more motivation to quit to GHWs, have predicted pack concealment.

107. Finally, Brennan et al. 2015 report that only two measures of GHW effectiveness, namely attributing much more motivation to quit to GHWs and requesting different GHWs, have positively and significantly predicted that smokers reported having attempted to quit. The authors find that the other proximal outcomes related to appeal, perceived harm and balance between smoking enjoyment and concern have not predicted quit attempts.

108. Ajzen et al. argue that Brennan et al.'s conclusion that quitting-related cognitions and behaviours are prospectively predicted by the more proximal beliefs and perceptions and that, among adults, GHWs are likely to be particularly influential in driving quitting behaviour is unfounded, reflecting a basic misconception of the logic inherent in correlation analysis. They consider that Brennan et al.'s analysis is unable to establish that the TPP measures have changed quit intentions/secondary indicators. According to Ajzen et al., if there is no evidence that the intervention had an effect on the outcome (quit intentions/secondary indicators), a mediation analysis cannot be used to establish that the intervention did have such an effect. They contend that contrary to the requirements for a mediation analysis, Brennan et al. 2015 mostly explored correlations between assumed mediators (e.g. appeal-related variables) and outcome measures (quit intentions and secondary indicators), for which at least one, and often both of the variables involved in the correlation analysis had not been found to have changed in a statistically significant way as a result of the TPP measures. For instance, Ajzen et al. argue that not believing that dangers of smoking are exaggerated, smoking enjoyment, smoking concern and balance between smoking enjoyment and concern are mechanism (mediator) variables for which there are no statistically significant changes reported in Wakefield et al. 2015. Similarly, they claim that daily thoughts about quitting, intending to quit, setting a firm date to quit and refraining from smoking

[169] See Brennan et al. 2015, (Exhibits AUS-224, DOM-304).

are quit intention/secondary indicator variables for which there is no statistically significant change reported in Durkin et al. 2015.[170]

109. Professor Chaloupka argues that limiting the analysis of more distal measures to continuing smokers effectively forces the estimates to suggest that the effects of the TPP measures are wearing out over time, when the opposite may actually be the case. Professor Chaloupka submits that while there is relatively little evidence of wear out for the most proximal outcomes, such as appeal and noticing/avoiding GHWs, most wear out evidence is obtained for the more distal outcomes, such as quitting-related measures. According to Professor Chaloupka rather than wearing out over time, the impact of the TPP measures on overall population attitudes, beliefs, and behaviours are likely to grow over time as young people, who might otherwise have taken up tobacco use, are discouraged from doing so, while younger current users are encouraged to quit.[171]

110. Professor Chaloupka further contends that neither Brennan et al. 2015 nor Ajzen et al. use the NTPPTS data to assess directly the impact of the TPP measures on proximal intermediate and/or distal tobacco-related outcomes. Professor Chaloupka argues that several measures on appeal, not included in Brennan et al. 2015 but analysed by Ajzen et al., are questionable, at best, measures of appeal. Similarly, several quit-related measures, not analysed in Brennan et al. 2015, are questionable measures of quitting. According to Professor Chaloupka, Ajzen et al. do not asses the relationship between other proximal outcomes and quitting-related outcomes, or between intermediate outcomes, which are found to have stronger associations. Professor Chaloupka argues that Ajzen et al. do not report other quitting-related measures (e.g. having daily thoughts about quitting in the past). Professor Chaloupka further submits that Ajzen et al. are using a much more stringent criterion for defining statistical significance than that used by Brennan et al. 2015.[172]

111. Ajzen et al. respond that the NTPPTS data provide no empirical basis for the assertion that there is a strong correlation between the appeal of tobacco production and smoking behaviour. Using the NTPPTS data, they replicate Brennan et al.'s analysis for appeal-related variables and correct for multiple hypothesis testing. They find that of the 130 potential correlations between appeal variable and downstream variables relating to quit intentions, quit attempts, secondary quit indicators and smoking behaviours, 129 showed no significant correlation.[173]

112. Ajzen et al. further contend that Professor Chaloupka formulates a series of unsubstantiated and unfounded criticisms regarding the NTPPTS data and their correlational analysis. In particular, they argue that the correlational analysis measures the association among variables but does not assess the impact of the TPP measures on proximal, or distal outcomes by using the longitudinal component of the NTPPTS. They further claim that the NTPPTS technical report underscores the importance of the appeal and downstream variables, considered as inappropriate or questionable by Professor Chaloupka. They submit that Professor Chaloupka fails to understand that the correlation analysis only focuses on the appeal mechanism and not on GHW effectiveness and pack ability to deceive in order to address Professor Fong's and Australia's core argument that the appeal of tobacco products is "very highly" correlated with smoking behaviour. According to Ajzen et al., Professor Chaloupka's comment regarding the criterion for defining statistical significance shows his failure to understand the importance of correcting for multiple hypothesis testing. They also argue that Professor Chaloupka's assertion that the NTPPTS data underestimate changes in some measures of intention and secondary indicators by not asking these questions to recent quitters is unfounded because the TPP measures did not increase quitting behaviour. They further disagree with Professor Chaloupka's assertion that some appeal variables do not measure change over time, because the NTPPTS data do measure change over time for these variables. According to the authors, the NTPPTS data show that although the TPP measures did increase pack dislike there were no changes in quit-related outcomes. Finally, Ajzen et al. argue that Professor Chaloupka's assertion that the NTPPTS data do not measure the impact of initiation by

[170] See Ajzen et al. Data Report, (Exhibit DOM/IDN-2), paras. 200-213.
[171] See Chaloupka Rebuttal Report, (Exhibit AUS-582), paras. 12-14.
[172] See Chaloupka Second Rebuttal Report, (Exhibit AUS-590), paras. 16-23.
[173] See Ajzen Rebuttal Report, (Exhibit DOM/HND/IDN-5), paras. 128-130 and 207-212, p. 44.

mirror is unwarranted because Australia and CCV have been unwilling to share the results of the 2013 schools-based survey.[174]

### 4.2 Analysis by the Panel

113. We note that Brennan et al. 2015 is the only peer-reviewed study analysing the association between proximal outcomes and distal outcomes based on the NTPPTS data.

114. A careful review of Brennan et al. 2015 suggests that there is a positive and statistically significant association between several outcomes related to GHW effectiveness and quitting cognitions and behaviour among adult smokers. The empirical evidence of the association between appeal variables and quitting-related thinking and behaviour outcomes is much more mixed and limited. Similarly, Brennan et al. 2015 report no statistically significant association between the perceived harm variable and any of the quitting-related thinking and behaviour outcomes.[175]

115. We note that Ajzen et al. did not replicate completely Brennan et al.'s 2015 analysis of the predictive relationships between proximal outcomes and quitting-related cognitions and behaviours. Ajzen et al. only replicate the analysis for appeal-related variables. They find that there was no significant correlation between the appeal of tobacco products and quit intention, secondary indicators and actual quitting behaviours, except between lower values and making more than one quit attempt. More generally, Ajzen et al. submit that Brennan et al.'s (2015) analysis is unable to establish that the TPP measures have changed quit intentions and secondary indicators, because Brennan et al. 2015 mostly explore correlations between assumed mediators and outcome measures, for which at least one, and often both of the variables involved in the correlation analysis have not been found to have changed in a statistically significant way as a result of the TPP measures.[176]

116. We note that Ajzen et al.'s results confirm to a large extent the findings of Brennan et al.'s findings related to appeal-related variables. Brennan et al. find no statistically significant association between most of the appeal variable and quitting-related cognitions and behaviours, while noting that further studies would be needed to explore this relationship. However, in many estimation results obtained with the resampled data based on the multiple testing procedure, Ajzen et al. find that none or only one or two explanatory variables (besides the variable of interest) are statistically significant, which could suggest that the resample data are subject to multicollinearity. Multicollinearity arises when two (or more) explanatory variables convey the same information. As a result, the coefficient estimates may become very sensitive to minor changes in the model specification or data and their confidence interval may increase. We further note that, as mentioned above, Ajzen et al. did not replicate the analysis for the variables related to GHW effectiveness, perceived harm and balance between enjoyment and concern. Brennan et al. themselves acknowledge that further mediation analyses and controlled experimental studies are required to establish if appeal, GHW effectiveness, perceived harm and enjoyment/concern variables are causally responsible for the observed changes in quitting-related outcomes.[177]

## 5 OVERALL CONCLUSION ON POST-IMPLEMENTATION EVIDENCE ON QUITTING-RELATED OUTCOMES AND OTHER DISTAL OUTCOMES

117. As discussed above, the parties have referred to several peer-reviewed studies analysing empirically the impact of the TPP measures and enlarged GHWs on quitting-related cognitions, pack concealment and quit attempts. The Dominican Republic, Indonesia and Honduras also provided several expert reports reviewing, and in some cases replicating, the results reported in these published papers. In response, Australia discussed the features of some of the survey datasets used in these peer-reviewed papers, but did not provide its own econometric analysis, unlike in the discussion on smoking prevalence and cigarette consumption.

118. At the outset, we note that the survey data used in these studies, may, as suggested by Australia, be more suited to analysing the impact of the TPP measures and enlarged GHWs on

[174] See Ajzen et al. Second Data Rebuttal Report, (Exhibit DOM/IDN-8), paras. 22-59.
[175] See Brennan et al. 2015, (Exhibits AUS-224, DOM-304).
[176] See Ajzen et al. Data Report, (Exhibit DOM/IDN-2), paras. 200-213.
[177] Brennan et al. 2015, (Exhibits AUS-224, DOM-304).

proximal outcomes, such as appeal, GHWs and ability of the pack to mislead than more distal outcomes, such as quitting intentions and quit attempts. Questions on quit intentions and quit interests were not asked to "recent quitters". In addition, none of the survey datasets discussed above track non-smokers who might have taken up smoking in the absence of the TPP measures and enlarged GHWs.

119. We also note that the parties disagree on the extent to which the variables related to appeal, GHW effectiveness, perceived harm and enjoyment/concern variables may be considered to be predictive of smoking and quitting behaviours. The authors of the peer-reviewed study analysing the predictive relationships between proximal outcomes and quitting-related outcomes acknowledge that further empirical analyses and experimental studies are required to establish causality.

120. With this in mind, and based on the studies and expert reports before us and discussed above, the empirical evidence available to us regarding quitting-related outcomes and other distal outcomes, which is sometimes scarce, suggests that:

a. The impact of the TPP measures and enlarged GHWs on adult cigarette smokers' quitting intention and quitting-related cognition reactions is limited and mixed.

b. The TPP measures and enlarged GHWs have had a statistically significant positive impact on avoidant behaviours, such as pack concealment, among adult cigarette smokers, while their impact on stubbing out and stopping smoking is much more limited and mixed.

c. Although the TPP measures and enlarged GHWs have statistically significantly increased calls to the Quitline, the observed impact of the TPP measures and enlarged GHWs on quit attempts is very limited and mixed.

d. The empirical evidence of the impact of the TPP measures and enlarged GHWs on adolescents' quitting-related outcomes is limited. This evidence suggests that the impact of the TPP measures and enlarged GHWs on adolescents' refraining from smoking cigarettes and thoughts about quitting is statistically not significant. No empirical evidence has been submitted to us on pack concealment among adolescent smokers.

e. The empirical evidence of the impact of the TPP measures and enlarged GHWs on cigar and cigarillo smokers' quitting-related outcomes is limited. This evidence suggests that the shares of premium cigar and cigarillo smokers and of non-premium cigarillo smokers reporting having decanted the cigars and cigarillos from their boxes to a humidor or an unbranded tin or concealed their pack have increased and there has been an increase in the share of non-premium cigarillo smokers contemplating quitting.

121. No post-implementation empirical evidence has been presented to us on the impact of the TPP measures on quit attempts among adolescent smokers and cigar and cigarillo smokers.

## APPENDIX C:

## EVIDENCE ON SMOKING PREVALENCE FOLLOWING THE ENTRY INTO FORCE OF THE TPP MEASURES

1. A number of expert reports submitted by the parties are dedicated in part or in whole to an assessment of the contribution of the TPP measures to reducing smoking prevalence.[1] These expert reports rely on different databases, statistical analysis and econometric methods to determine whether TPP and enlarged GHWs have contributed to a reduction in smoking prevalence.

2. One of the only points of agreement among the parties in the discussion on the impact of the TPP measures on smoking prevalence is that the empirical econometric studies they submitted do not assess separately the impact of TPP and the impact of the enlarged GHWs, because both measures were implemented at the same time.[2] Unless specified otherwise, in this Appendix, references to the impact of the TPP measures therefore refer to the impact of the TPP measures and the enlarged GHWs implemented simultaneously.

3. The complainants argue that the overall empirical statistical and econometric studies carried out by their experts conclude that the TPP measures have failed to reduce cigarette and cigar smoking prevalence.[3] The complainants also initially suggested that the TPP measures "backfired" by increasing youth smoking prevalence[4], although they did not pursue this argument in later stages of the proceedings.

4. Notwithstanding its position that, in the early stages of introduction of the measures, the most appropriate way to discern their effects is to rely on experiments and surveys which consider drivers of choice, attitudes, and ultimately, the elicitation of behavioural intentions[5], Australia engaged in estimating econometrically the impact of the TPP measures on smoking prevalence, in response to the submissions of the Dominican Republic, Honduras and Indonesia. Australia argues that using the most appropriate dataset available and correcting for flaws in the econometric models put forward by the experts of the Dominican Republic, Honduras and Indonesia, the results show that the TPP measures have already contributed to reducing cigarette and cigar smoking prevalence.[6]

---

[1] See Chipty Report, (Exhibit AUS-17); Chipty Supplementary Report, (Exhibit AUS-511); Chipty Rebuttal Report, (Exhibit AUS-535) (SCI); Chipty Surrebuttal Report, (Exhibit AUS-586); Chipty Second Rebuttal Report, (Exhibit AUS-591); Chipty Third Rebuttal Report, (Exhibit AUS-605); Chaloupka Rebuttal Report, (Exhibit AUS-582); Chaloupka Third Rebuttal Report, (Exhibit AUS-604); List Report, (Exhibit DOM/IDN-1); List Rebuttal Report, (Exhibit DOM/IDN-3); List Second Supplemental Report, (Exhibit DOM/IDN-5); List Third Supplemental Report, (Exhibit DOM/IDN-7); List Summary Report, (Exhibit DOM/IDN-9); IPE Report, (Exhibit DOM-100); IPE Updated Report, (Exhibit DOM-303); IPE Second Updated Report, (Exhibit DOM-361); IPE Third Updated Report, (Exhibit DOM-375); IPE Summary Report, (Exhibit DOM-379); Klick Report, (Exhibit UKR-5); Klick Rebuttal Report, (Exhibit HND-118); Klick Supplemental Rebuttal Report, (Exhibit HND-122); Klick Second Supplemental Rebuttal Report, (Exhibit HND-165); Klick Third Supplemental Rebuttal Report, (Exhibit HND-166); and Klick Fourth Supplemental Rebuttal Report, (Exhibit HND-169).

[2] See Australia's first written submission, para. 518; Honduras's response to Panel question No. 8; Dominican Republic's response to Panel question No. 8, para. 61; and Indonesia's response to Panel question No. 8, para. 8.

[3] See Honduras's second written submission, paras. 56-60; Dominican Republic's comments on responses to Panel questions following the second substantive meeting, paras. 681-682; Cuba's second written submission, paras. 276-277; and Indonesia's second written submission, para. 192.

[4] See Honduras's first written submission, para. 395; Dominican Republic's first written submission, para. 523; Cuba's first written submission, para. 163; and Indonesia's first written submission, para. 412.

[5] See Australia's first written submission, paras. 147 and 670. See also the discussion reflected at section 7.2.5.3.3. in the main body of these Reports.

[6] See Australia's comments on complainants' response to Panel question No. 197, para. 214.

5. Overall, we note that the approaches proposed by the parties to analyse the trends in smoking prevalence evolved over the course of the proceedings. They address the following three main aspects, that we will review in turn:

- First, the parties have submitted economic figures and descriptive statistics analyses aimed at determining whether smoking prevalence has decreased following the implementation of the TPP measures;

- Second, Australia, the Dominican Republic, Honduras and Indonesia have submitted statistical analyses to determine whether there was a break in the trend in smoking prevalence following the implementation of the TPP measures, and in particular, whether the reduction of smoking prevalence has accelerated following the implementation of the TPP measures;

- Finally, Australia, the Dominican Republic, Honduras and Indonesia have submitted econometric analysis to determine whether the TPP measures have contributed to a reduction in smoking prevalence by isolating and quantifying the different factors that can explain the evolution of smoking prevalence.

## 1 WHETHER SMOKING PREVALENCE DECREASED FOLLOWING THE IMPLEMENTATION OF THE TPP MEASURES

6. The parties have submitted different data sources tracking smoking prevalence in Australia. Each dataset is presented separately. We consider each in turn before turning to an overall assessment.

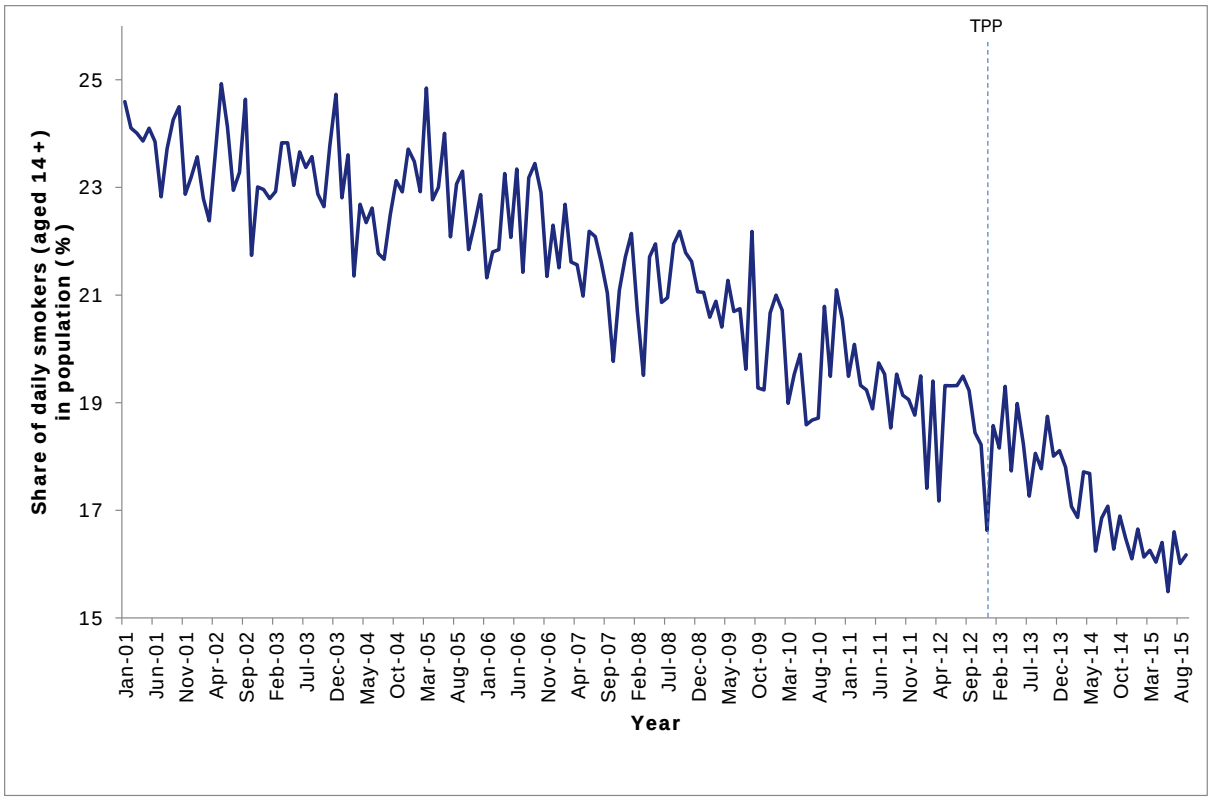
### 1.1 Datasets and related analyses

#### 1.1.1 Roy Morgan Single Source

7. The Roy Morgan Single Source (RMSS), submitted first by the Dominican Republic, is a large survey dataset based on more than 50,000 interviews conducted each year with randomly selected individuals across Australia. The dataset is based on a representative survey of the Australian population and solicits responses from respondents regarding socio-economic and demographic characteristics, as well as consumer behaviour such as smoking status and brand choices. The same individuals are however not surveyed over time. Monthly data are available for the period January 2001 through September 2015.

8. As shown in **Figure C.1**, the RMSS data reveal a downward trend in smoking prevalence that has accelerated since July 2006. In 2001, the smoking prevalence was around 24%. In 2006, smoking prevalence was slightly lower at 23%. In 2015, the level of smoking prevalence was 18%.

**Figure C.1: Smoking Prevalence Based on RMSS Data**




Note: The vertical dashed line indicates the introduction of TPP and enlarged GHWs.

Source: RMSS data (January 2001 – September 2015).

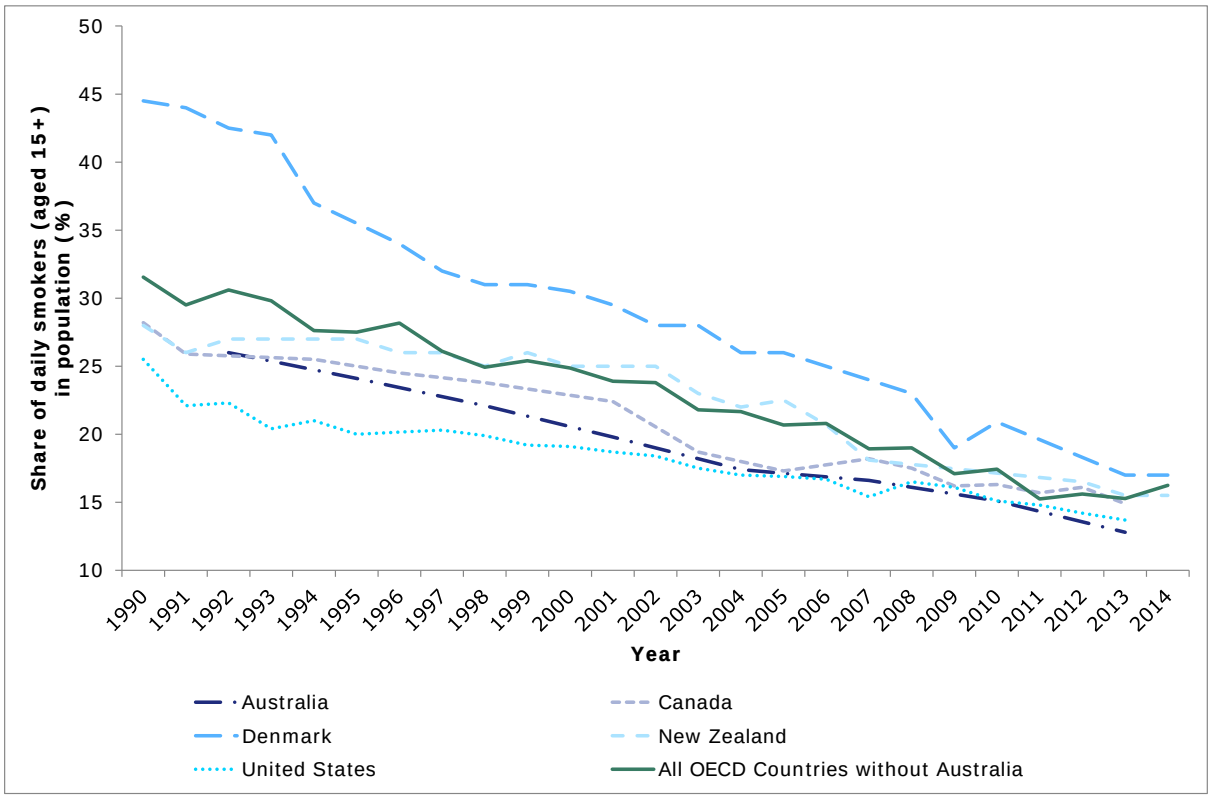
### 1.1.2 Organisation for Economic Co-operation and Development Dataset on Non-Medical Determinants

9. The Organisation for Economic Co-operation and Development (OECD) Dataset on Non-Medical Determinants is a panel dataset on various smoking behaviours, including yearly smoking prevalence rates covering all 34 OECD countries from 1960 through 2014 (2013 in the case of Australia).

10. The IPE Report submitted by the Dominican Republic contends that there is a secular downward trend in smoking prevalence in Australia and other high income countries, which are presumably, at least in part due to a combination of demographic shifts (change in the composition of population, education, etc.) as well as other factors entirely unrelated to tobacco control interventions (such as a general trend towards a healthier lifestyle and away from smoking).[7]

[7] See IPE Updated Report, (Exhibit DOM-303), paras. 43-46; and IPE Third Updated Report, (Exhibit DOM-375), paras. 193-201.

**Figure C.2: Smoking Prevalence Based on OECD Dataset on Non-Medical Determinants in Australia and Selected OECD Countries**




Source: IPE Updated Report, (Exhibit DOM-303), p. 24

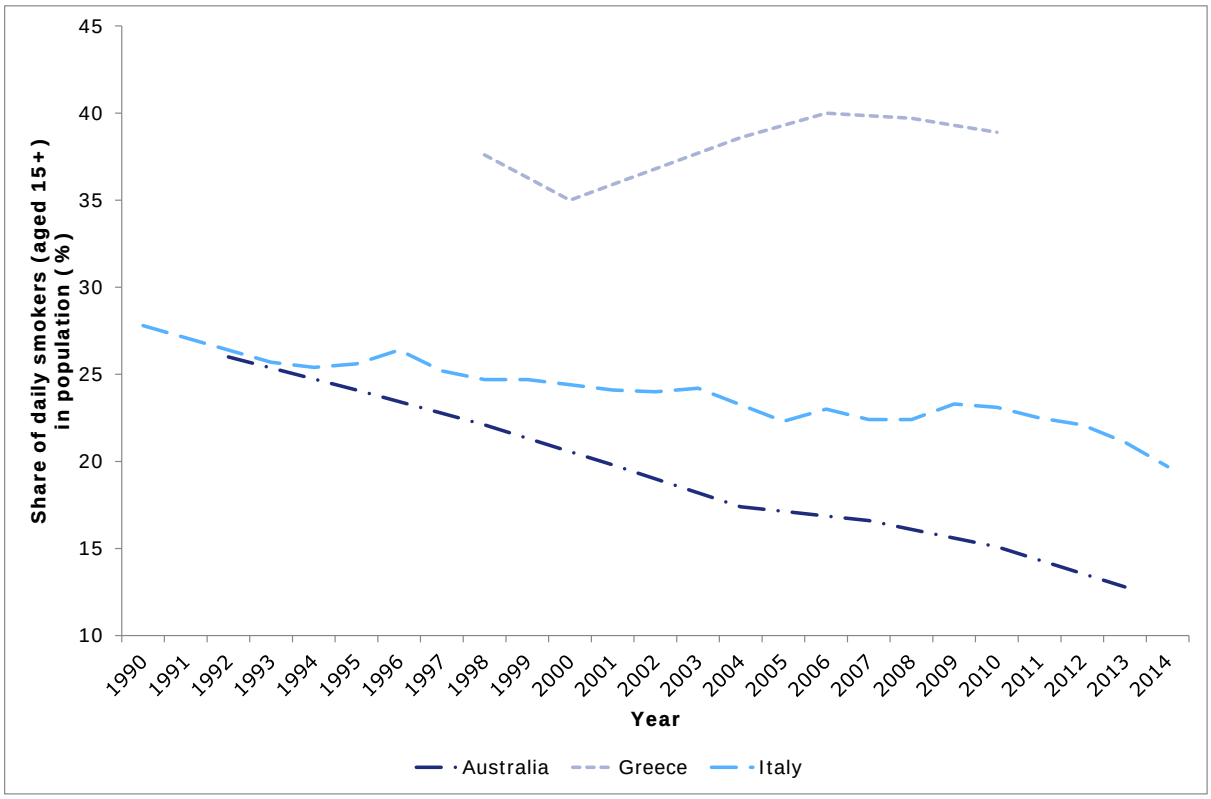
11. Professor Chaloupka, in an expert report submitted by Australia, disagrees that (1) tobacco use has been falling consistently in all OECD countries, (2) this decline has been largely linear over time and (3) these downward trends are expected to continue into the future regardless of what happens in these countries. According to Professor Chaloupka trends in tobacco use differ considerably across OECD countries and that assuming a linear downward trend over time is overly simplistic and fails to fully capture the role of tobacco control policies (or lack thereof) in accelerating (decelerating) any downward trend in tobacco use.[8] He further argues that the differences in trends in smoking prevalence between Australia and other OECD countries are more pronounced when one looks at countries, such as Germany or Italy, that have not been included in the figures shown in the IPE Report, such as in **Figure C.2**.[9] According to Australia and as shown in **Figure C.3**, the assertion that there is a secular downward trend in smoking across all OECD countries is belied by the rising trend in prevalence in Greece, which is largely attributable to the weak tobacco control policies the country has implemented (29th out of 34 European countries in the 2013 tobacco control scale).[10]

[8] See Chaloupka Rebuttal Report, (Exhibit AUS-582), paras. 37-44; and Chaloupka Third Rebuttal Report, (Exhibit AUS-604), paras. 20-28.

[9] See IPE Updated Report, (Exhibit DOM-303), p. 24.

[10] Chaloupka Rebuttal Report, (Exhibit AUS-582), para. 44.

**Figure C.3: Smoking Prevalence Based on OECD Dataset on Non-Medical Determinants in Australia and Selected European Countries**




Source: Chaloupka Rebuttal Report, (Exhibit AUS-582), pp. 37-38.

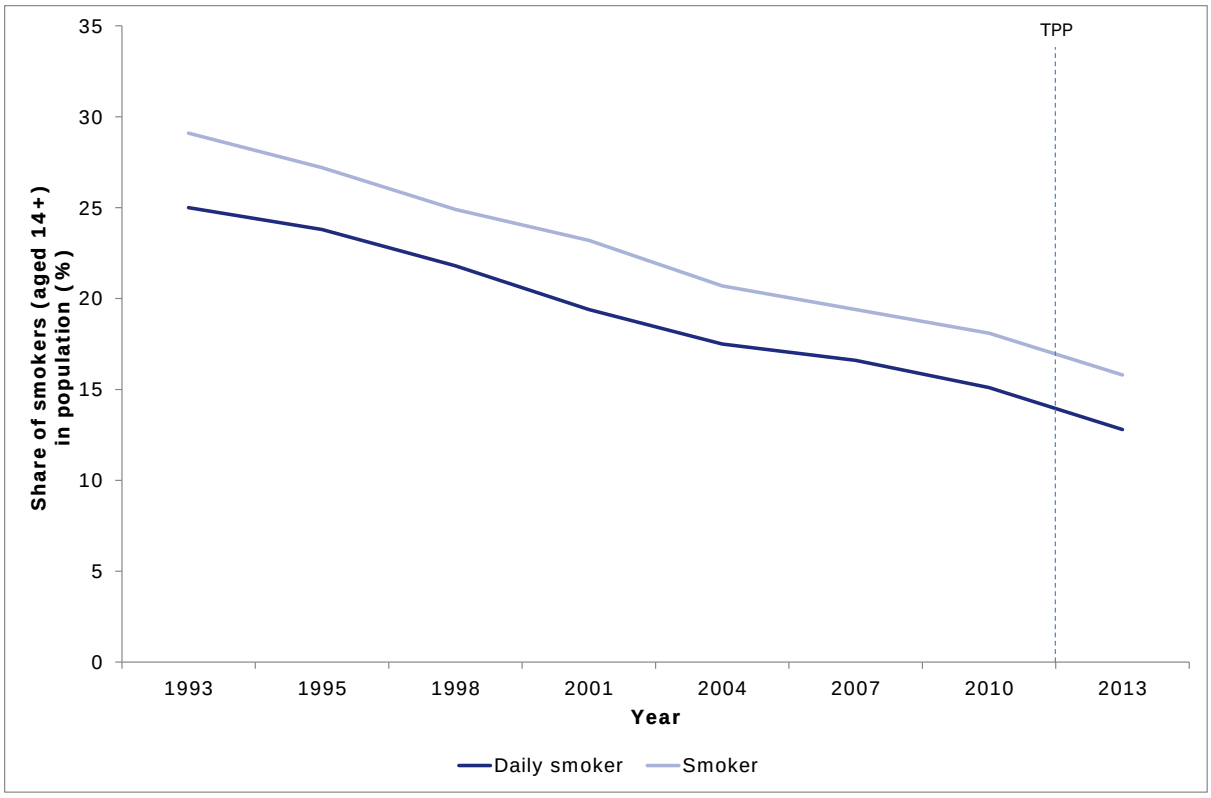
### 1.1.3 National Drug Strategy Household Survey

12. The National Drug Strategy Household Survey (NDSHS) is a nationally representative survey run by the Australian Government approximately every two or three years. At the time of these proceedings, the most recent wave of the NDSHS had been undertaken in 2013, and was the only wave of the NDSHS to have been undertaken since the introduction of the TPP measures.

13. The Dominican Republic note that the last eight NDSHS reports indicate, as shown in **Figure C.4**, that smoking prevalence (both overall smoking prevalence and daily smoking) has decreased along a roughly linear trend since 1993. The most recent NDSHS survey reveals that smoking rates in 2013, the first year after the implementation of TPP, have evolved according to this trend without "break", "shift" or "kink" in the trend line that could be attributable to the TPP measures.[11]

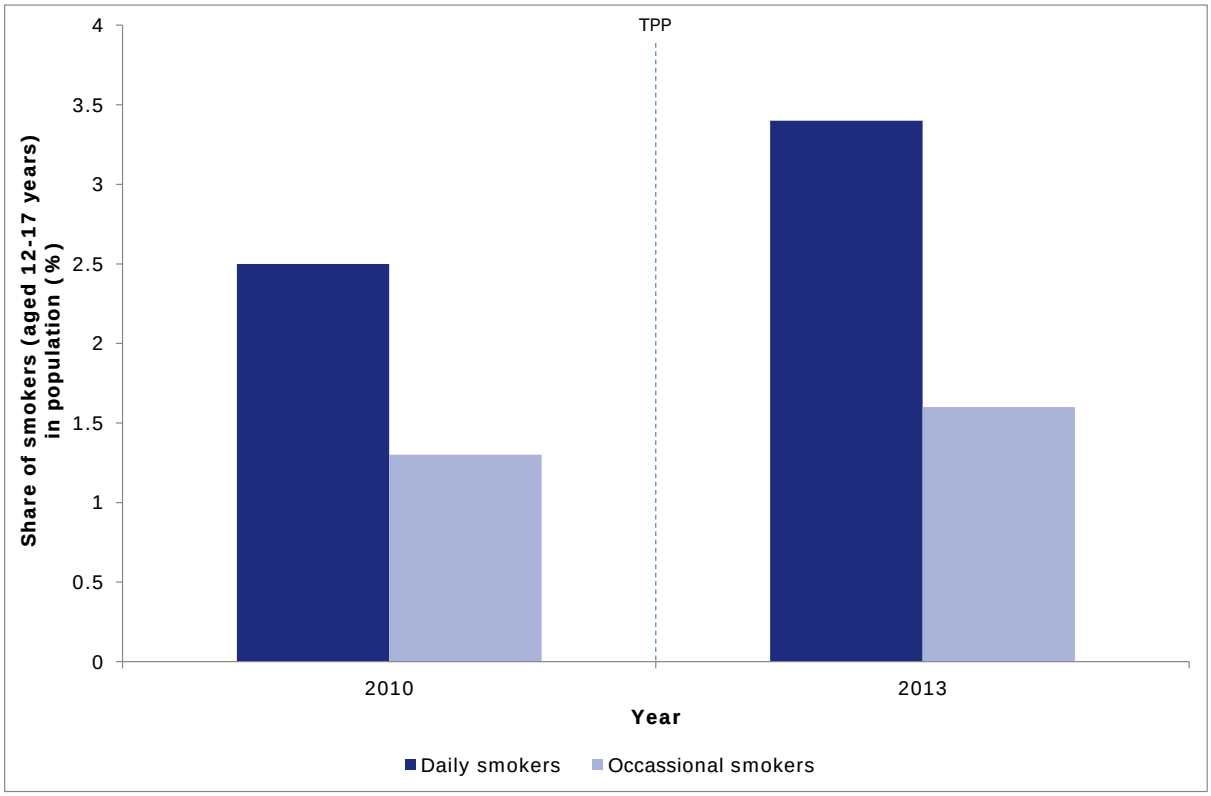
[11] Dominican Republic's first written submission, para. 523.

### Figure C.4: Smoking Prevalence Based on NDSHS Data




Note: The vertical dashed line indicates the introduction of TPP and enlarged GHWs.
Source: Dominican Republic's first written submission, p. 157.

### Figure C.5: Youth Smoking Prevalence Based on NDSHS Data



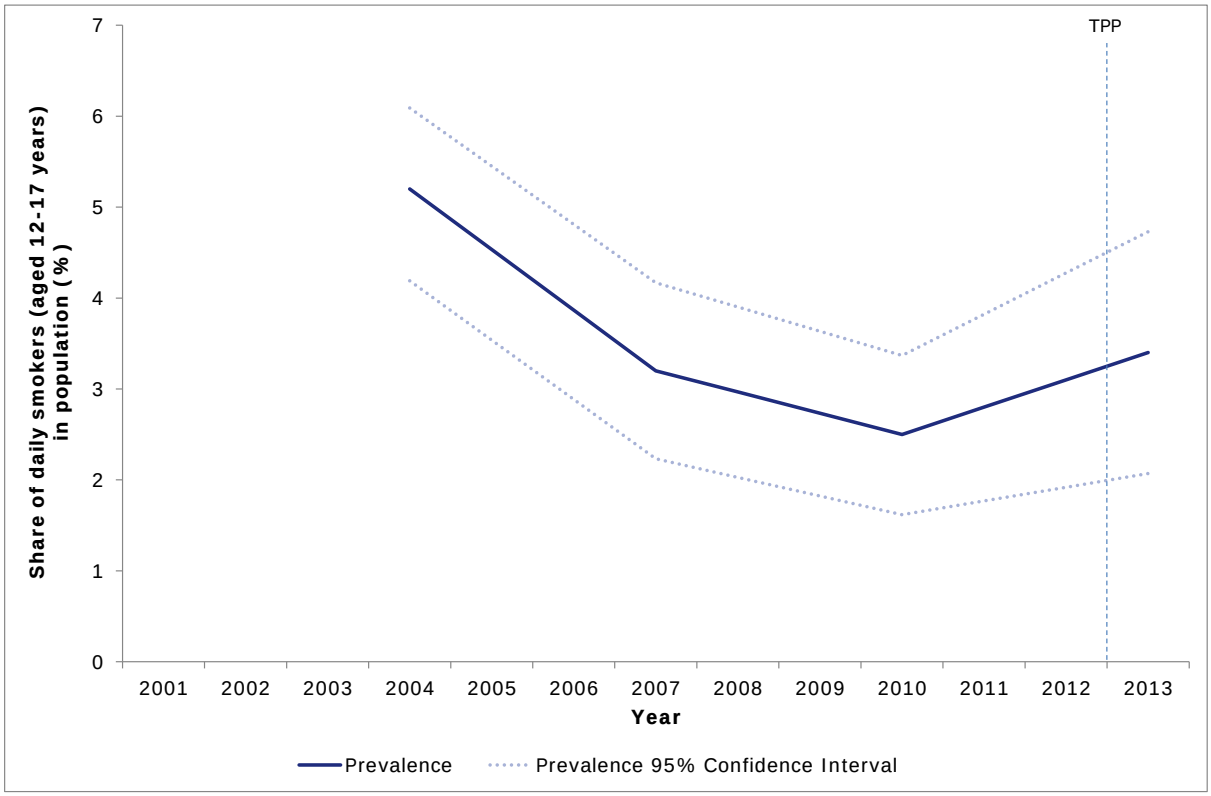

Note: The vertical dashed line indicates the introduction of TPP and enlarged GHWs.
Source: Gibson Report, (Exhibit DOM-92), p. 22.

14. According to the Dominican Republic, Honduras and Indonesia the most recent NDSHS data show an increase in smoking prevalence in certain segments of the Australian population. For instance, as shown in **Figure C.5**, prevalence of daily smoking in the 12–17 year old group has increased from 2.5% in 2010 to 3.4% in 2013, marking the first increase in prevalence rates after years of decline for this category, from 5.2% in 2004 to 3.2% in 2007 to 2.5% in 2010.[12]

15. According to Australia and its expert Dr Chipty, there were significant reductions in daily and overall smoking prevalence reported in the most recent wave of the NDSHS data. In particular, between 2010 and 2013, rates of daily smoking among people aged 18 years or older dropped from 15.9% to 13.3%. In 2013, 12.8% of people in Australia aged 14 or older were daily smokers, declining from 15.1% in 2010.[13]

16. Australia, however, notes that given the small sample sizes, particularly for certain subgroups, like adolescents or residents of specific Australian states, trend lines can be difficult to interpret. Australia argues that, for example, one cannot conclude from these data that daily smoking increased in the youth population following TPP. This is because actual youth daily smoking prevalence among the underlying population may, in reality, be flat or decreasing given the small sampling error associated with these estimates, as shown in **Figure C.6**.[14] Australia's Post-Implementation Review Tobacco Plain Packaging 2016 further explains that the NDSHS report states that this rise in the number of 12-17 year olds smoking in the 2010-2013 reporting period is not statistically significant and should be interpreted with caution.[15]

**Figure C.6: Youth Smoking Prevalence with Confidence Interval Based on NDSHS Data**




Note: The vertical dashed line indicates the introduction of TPP and enlarged GHWs.

Source: Chipty Report, (Exhibit AUS-17), p. 46.

[12] Honduras's first written submission, paras. 395-396; Dominican Republic's first written submission, para. 523; Indonesia's first written submission, para. 412; and Gibson Report, (Exhibit DOM-92), p.22.

[13] See Australia's first written submission, para. 36; and Chipty Report, (Exhibit AUS-17), paras. 77-84.

[14] See Chipty Report, (Exhibit AUS-17), paras. 77-84; and Australia's comments on the complainants' responses to Panel question No. 197, para. 387.

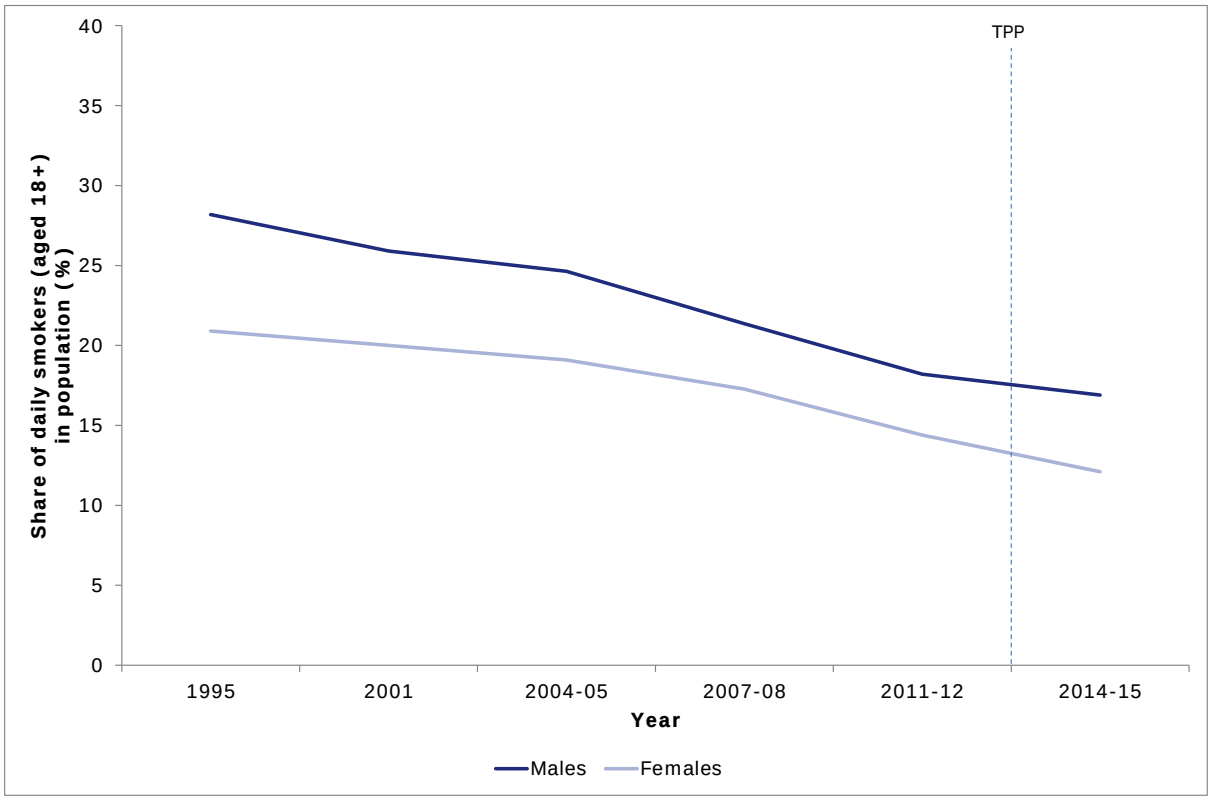
[15] See Australia's first written submission, Annexure E, para. 77; and Tobacco Plain Packaging PIR, (Exhibit AUS-624), para. 119.

### 1.1.4 National Health Survey

17. The National Health Survey (NHS) is a nationally representative survey, released by the Australian Bureau of Statistics (ABS). The latest NHS covered the period 2014-2015 and surveyed approximately 19,000 respondents. Previous surveys releases covered the period 1995, 2001, 2004-2005, 2007-2008 and 2011-2012.

18. Australia notes that according to the first results of the 2014-2015 NHS, daily smoking among Australians aged 18 and over was 14.5% in 2014-2015, down from 16.1 % in 2011-2012, as shown in **Figure C.7**.[16] Australia further recognizes that the specific results of the NDSHS and NHS are not directly comparable due to differences in methodology, age cohorts, timing and sample sizes (including across age cohorts).

**Figure C.7: Smoking Prevalence Based on NHS Data**




Note: The vertical dashed line indicates the introduction of TPP and enlarged GHWs.

Source: NHS Results, (Exhibit AUS-622), p. 30.

### 1.1.5 Australia Secondary Students' Alcohol and Drug survey

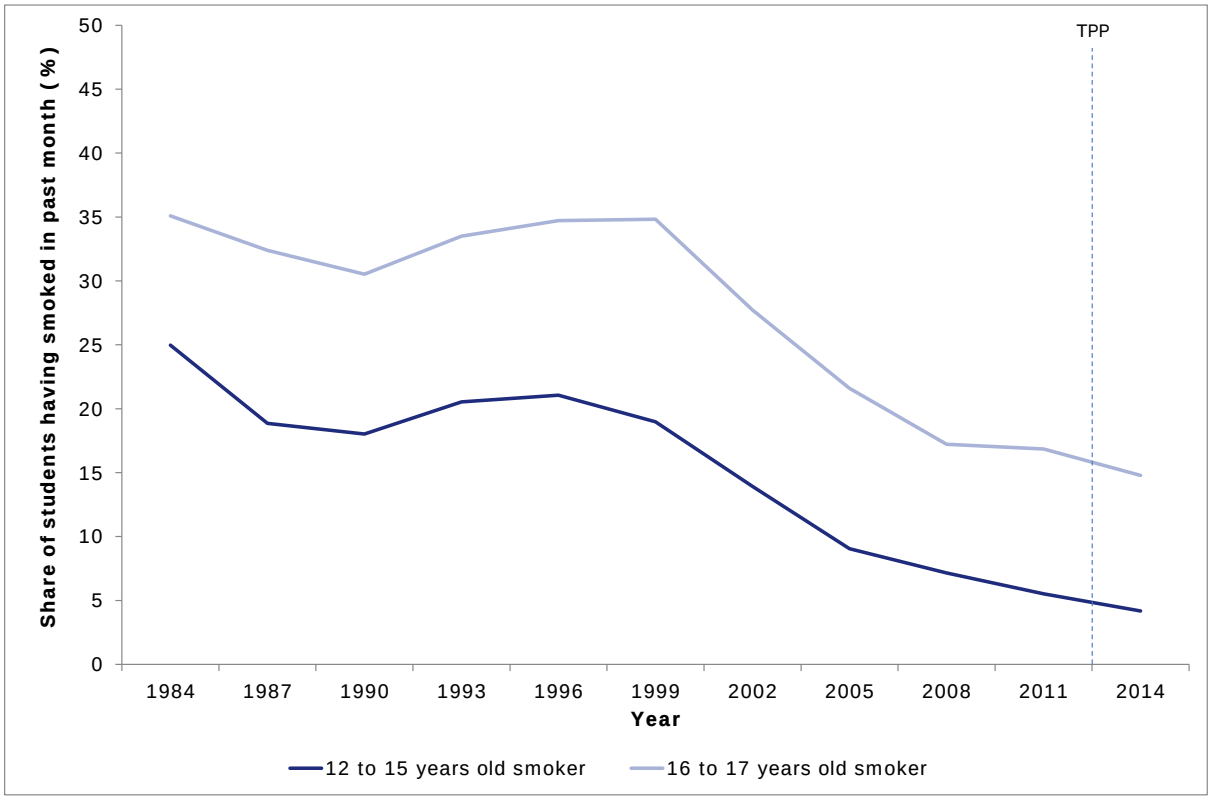
19. The Australian Secondary Students' Alcohol and Drug (ASSAD) survey is a triennial national survey of secondary school students' use of licit and illicit substances. This survey assesses, among others, their use of alcohol and tobacco and was conducted collaboratively by Cancer Councils across Australia, commencing in 1984.

20. According to Australia and in contrast to assertions of the Dominican Republic, Honduras and Indonesia with respect to the youth smoking findings in the NDSHS, more recent data from the

[16] See Australia's comments on the complainants' responses to Panel question No. 197, para. 389; NHS Results, (Exhibit AUS-622), p. 30; and Tobacco Plain Packaging PIR, (Exhibit AUS-624), para. 126.

ASSAD survey shows statistically significant declines in current smoking prevalence between 2011 and 2014 for students aged 12 to 17 years, as depicted in **Figure C.8**.[17]

**Figure C.8: Youth Smoking Prevalence Based on ASSAD Data**




Note: The vertical dashed line indicates the introduction of TPP and enlarged GHWs.

Source: CCV 2014 Survey, (Exhibit AUS-621), pp. 33 and 37.

### 1.1.6 Commissioned Roy Morgan Research Survey (Australia and New Zealand)

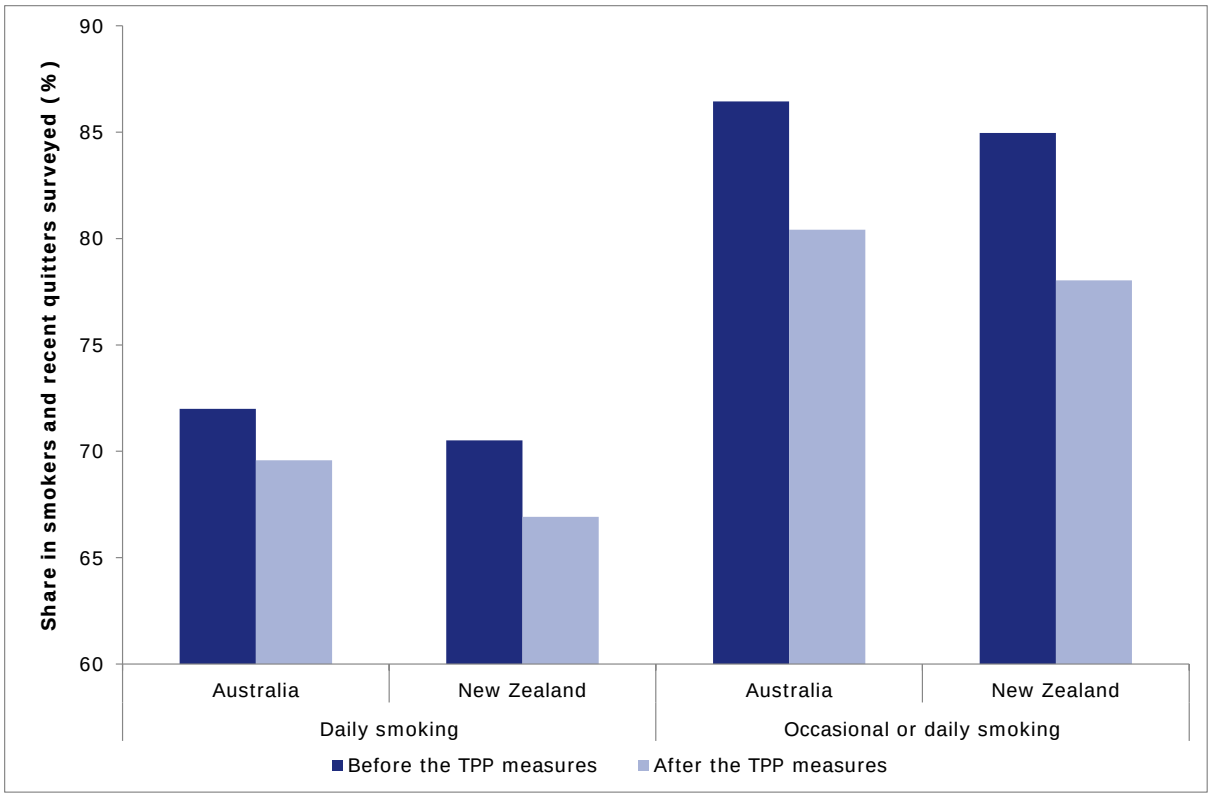
21. Roy Morgan Research data, commissioned by Honduras's expert Professor Klick, is a survey of individuals who were current or former (in the past 12 months) smokers in both Australia and New Zealand undertaken using random digit dialling sampling techniques. The first wave of the survey was completed prior to the introduction of the TPP measures in December 2012 between 2 November 2012 and 26 November 2012 in Australia and between 8 November 2012 and 1 December 2012 in New Zealand. Subsequent waves have been carried out at three-month intervals: Wave 2 between 28 February 2013 and 19 March 2013; Wave 3 between 30 May 2013 and 20 June 2013; Wave 4 between 8 August 2013 and 26 August 2013; Wave 5 between 8 November 2013 and 24 November 2013 and Wave 6 between 7 February 2014 and 28 February 2014.

22. Professor Klick notes, as shown in **Figure C.9**, that the change in Australian respondents reporting a daily smoking status from before the TPP measures (72.0%) to after the TPP measures (69.6%) was 2.4% points (averaged across the post-TPP waves), while for New Zealand, the decline was 3.6% points (before 70.5% vs. after 66.9%). Similarly, the decline in occasional smoking status in Australia was 6.2% and 7.1% point decline observed in New Zealand.[18]

[17] See Australia's comments on the complainants' responses to Panel question No. 197, para. 390; CCV 2014 Survey, (Exhibit AUS-621), Table 6.1; Tobacco Plain Packaging PIR, (Exhibit AUS-624), para. 126; and Dessaix et al. 2016, (Exhibit AUS-623), p. 1.

[18] See Klick Report, (Exhibit UKR-5), pp. 8-10.

**Figure C.9: Smoking Behaviour Based on Roy Morgan Research Data**




Source: Klick Report, (Exhibit UKR-5), pp. 8-10.

23. Australia's expert, Dr Chipty, notes that the "smoking incidence", defined by Professor Klick as the share of current smokers and individuals who have been smokers at some point during the past 12 months, is different from "smoking prevalence", which is the share of the entire population that is smoking.[19] Dr Chipty further submits that daily smoking incidence fell more in Australia after the introduction of the TPP measures (5% between Wave 1 and 6, 5% between Wave 2 and 6) than in New Zealand (3% between Wave 1 and 6, 0% between Wave 2 and 6). In addition, a higher proportion of New Zealand respondents cited costs as the primary reason for quitting as compared to Australian respondents, while a higher proportion of Australian respondents cited health as the primary reason for quitting, as compared to New Zealand respondents.[20]

### 1.1.7 State-level smoking prevalence datasets

24. In their first written submissions, the complainants presented data for four Australian States, namely New South Wales (NSW), South Australia, Queensland, and Victoria, suggesting that smoking prevalence had increased after the implementation of the TPP measures.[21] Most of these datasets were not discussed by the complainants in later submissions, except Honduras and Cuba for some datasets.[22] Each state-level dataset is reviewed next.

#### 1.1.7.1 Cancer Institute New South Wales Tobacco Tracking Survey

25. The Cancer Institute New South Wales Tobacco Tracking Survey (CITTS) is a weekly tracking telephone survey of smokers and recent quitters (in the past 12 months) based in NSW. The data

[19] See Chipty Report, (Exhibit AUS-17), fn 13.

[20] See Chipty Report, (Exhibit AUS-17), paras. 25-31, Appendix C.

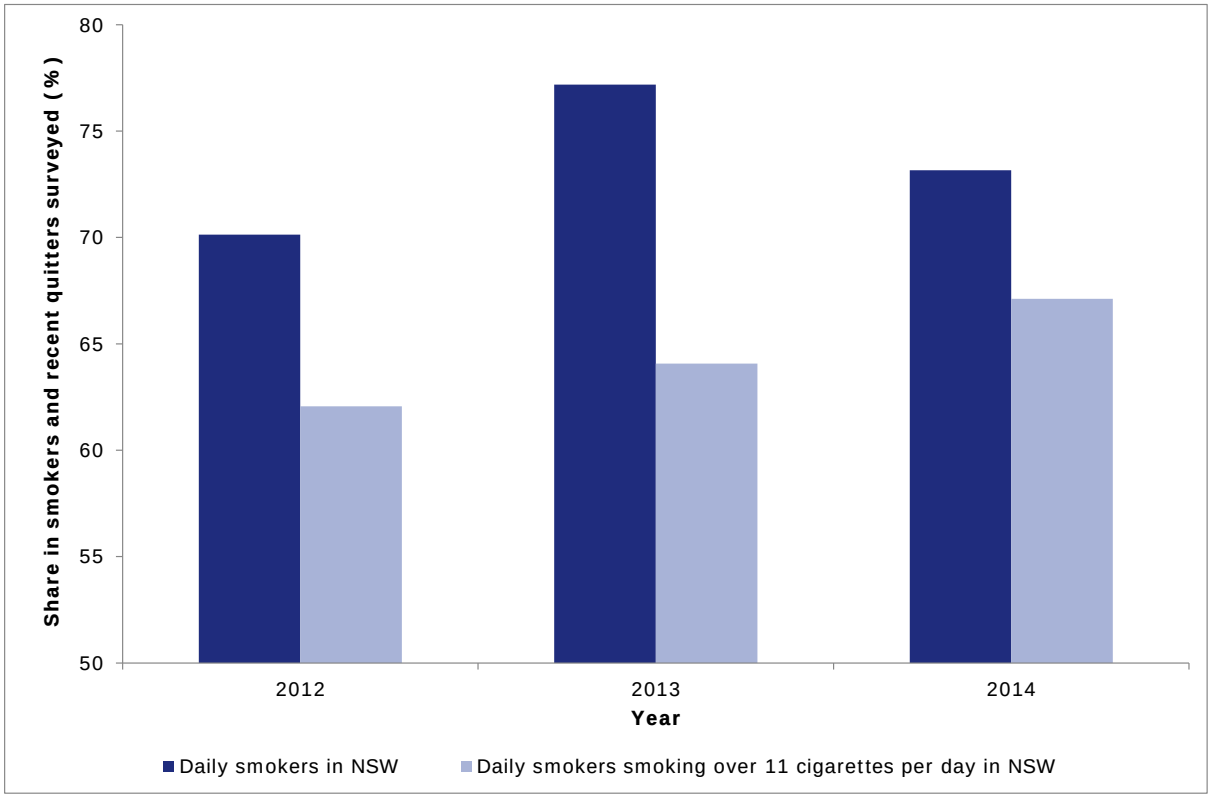
[21] See Honduras's first written submission, para. 396; Dominican Republic's first written submission, paras. 525-527; Indonesia's first written submission, para. 23; and Cuba's first written submission, paras. 164, 251.

[22] See Honduras's response to Panel question No. 199, pp. 46-47 (referring to previous analysis of the CITTS and NDSHS data); and Cuba's response to Panel question No. 199, pp. 20-21 (referring to previous analysis of the NDSHS, RMSS and CITTS data, among others).

discussed by the Dominican Republic, Honduras and Indonesia cover the years 2012, 2013 and 2014.

26. The Dominican Republic, Honduras and Indonesia refer to the Gibson Report, which was prepared at the request of British American Tobacco UK and submitted to a UK government consultation process on the introduction of the TPP measures. As shown in **Figure C.10**, the Gibson Report presents the CITTS data showing that the proportion of smokers surveyed, who smoke on a daily basis, actually increased from 70% in 2012, the year before the implementation of the TPP measures, to 77% in 2013 and remained above the 2012 levels in 2014. Similarly, the proportion of people smoking over 11 cigarettes a day increased from 62% in 2012 to 64% in 2014.[23]

**Figure C.10: New South Wales Consumer Smoking Behaviour Based on CITTS Data**




Source: Gibson Report, (Exhibit DOM-92), p. 20.

27. Australia considers that the Gibson Report's analysis is fundamentally flawed because the figures reported are incorrectly labelled as representing the "proportion of smokers" smoking on a daily basis, while they represent the proportion of the entire sample (including both smokers and recent quitters), who are, or in the case of recent quitters were, daily smokers. Accordingly, Australia submits that since the CITTS is a survey of smokers and recent quitters, and not a population survey, it is not designed to measure (and indeed is not capable of measuring) changes in smoking prevalence in the entire population.[24]

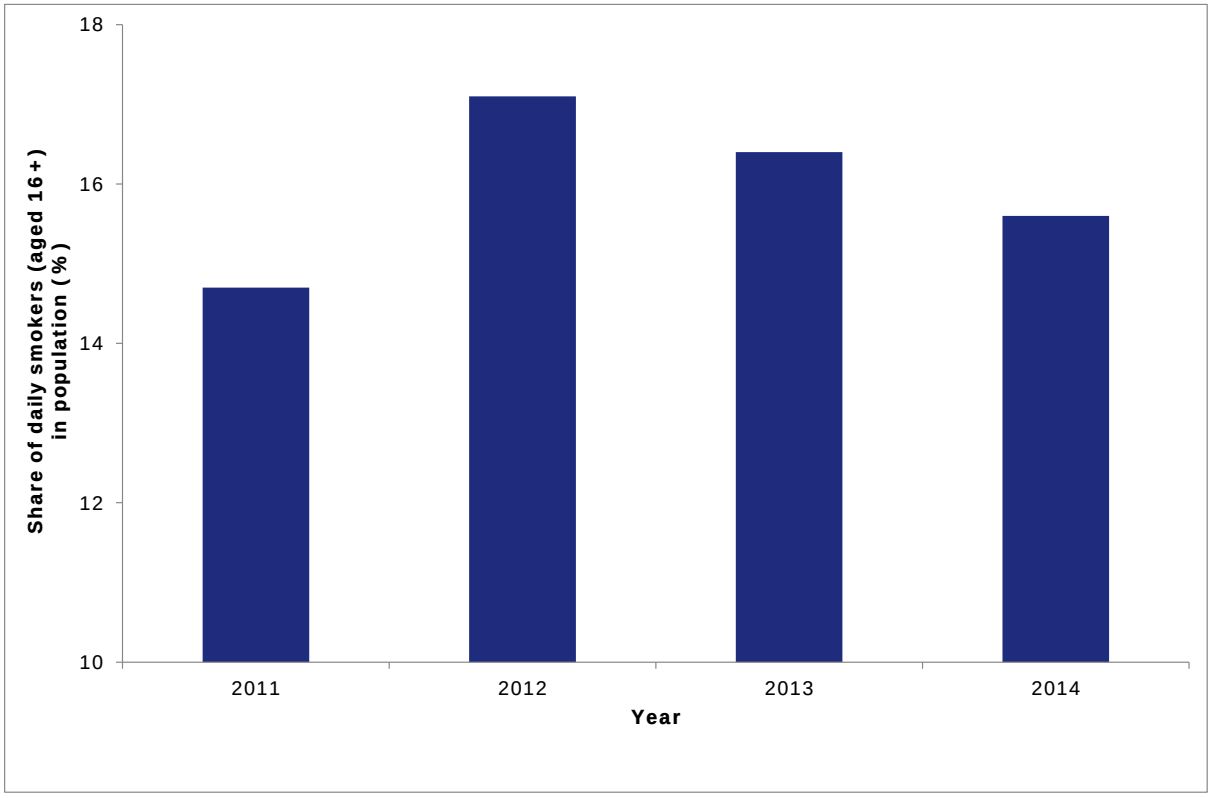
28. Australia further considers that the RMSS data, collected monthly, is better suited to comparing more accurately, including by states, smoking rates immediately prior to and following the introduction of the TPP measures. According to RMSS data, overall smoking prevalence in New South Wales has fallen significantly in the 12 months following the introduction of the

[23] Honduras's first written submission, para. 396; Dominican Republic's first written submission, para. 525; Indonesia's first written submission, para. 23; and Gibson Report, (Exhibit DOM-92), p. 22.

[24] See Australia's first written submission, Annexure E, paras. 60-61; and CINSW Rebuttal of BATA Analysis of CITTS Data, (Exhibit AUS-504).

TPP measures.[25] Australia's Post-Implementation Review Tobacco Plain Packaging 2016 also reports a reduction in daily smoking prevalence in New South Wales based on the NSW survey data (data up to 2014) for individuals aged 16 years and over, as shown in **Figure C.11**, and the NDSHS (data up to 2013) for individuals aged 14 years and over.[26]

**Figure C.11: Smoking Prevalence in New South Wales Based on CINSW Survey Data**




Source: Tobacco Plain Packaging PIR, (Exhibit AUS-624), p. 42.

### 1.1.7.2 New South Wales School Students Health Behaviours Survey

29. The New South Wales School Students Health Behaviours Survey (SSHBS) asked students aged 12–17 years a range of questions on alcohol, demographics, height and weight (including perception of body mass), injury, nutrition, physical activity, psychological distress, sedentary behaviour, substance use, sun protection (including sunburn experience and solarium use), and tobacco.

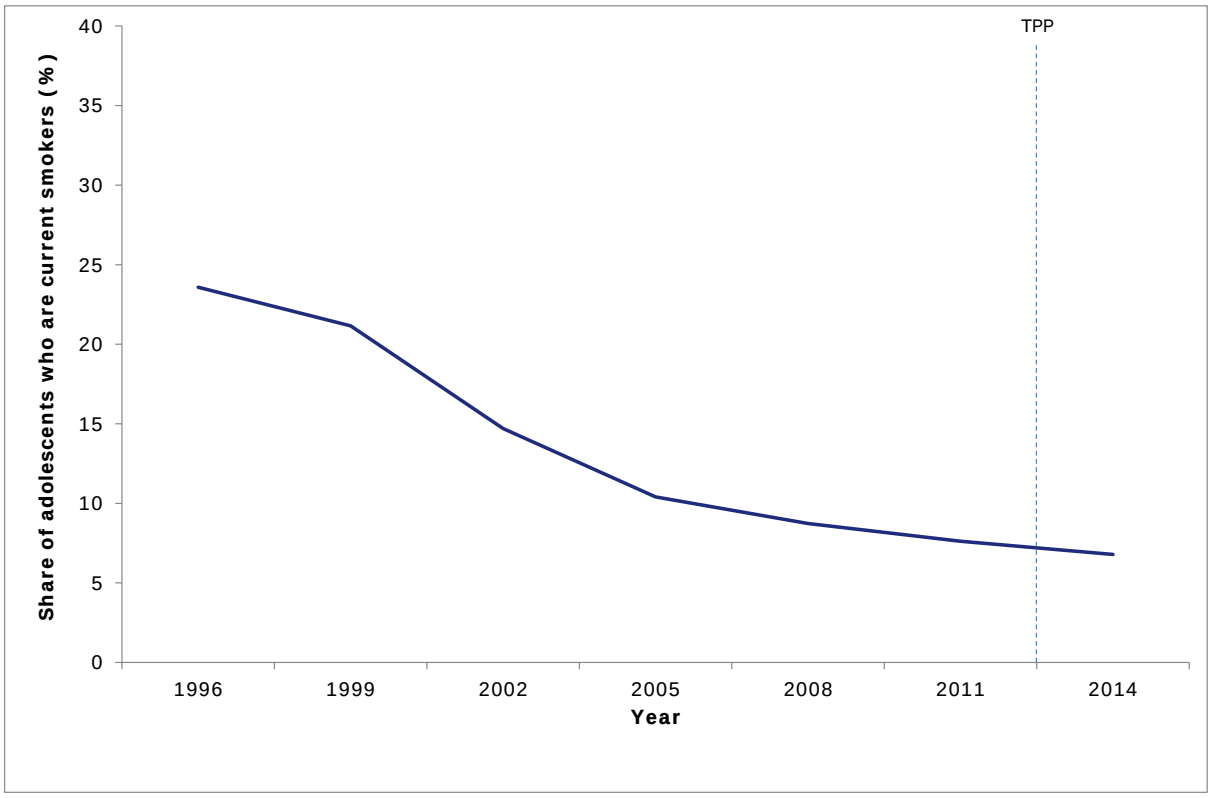
30. Australia submitted a recent study by Dessaix et al. 2016 based on SSHBS data showing, as highlighted in **Figure C.12**, the proportion of adolescents reporting current smoking as 6.7% in 2014, down from 23.5% in 1996.[27]

[25] See Australia's first written submission, Annexure E, paras. 82-86. See also Chipty Report, (Exhibit AUS-17), paras. 80-81.

[26] See Tobacco Plain Packaging PIR, (Exhibit AUS-624), para. 128.

[27] See Australia's comments on the complainants' responses to Panel question No. 197, para. 390; and Dessaix et al. 2016, (Exhibit AUS-623), p. 1.

**Figure C.12: Smoking Prevalence Among Youth in New South Wales Based on SSHBS Data**




Note: The vertical dashed line indicates the introduction of TPP and enlarged GHWs.

Source: Dessaix et al. 2016, (Exhibit AUS-623), p. 1.

### 1.1.7.3 South Australian Health Omnibus Survey

31. The South Australian Health Omnibus Survey (SAHOS) is a representative, cross-sectional survey that has been in operation since 1990. The data discussed by the Dominican Republic cover the years 2004 to 2013.

32. According to the Dominican Republic, the SAHOS data show, as depicted in **Figure C.13**, that smoking prevalence of the responding population has increased, rather than decreased, in South Australia between 2012, the year prior to the implementation of the TPP measures, and 2013.[28]

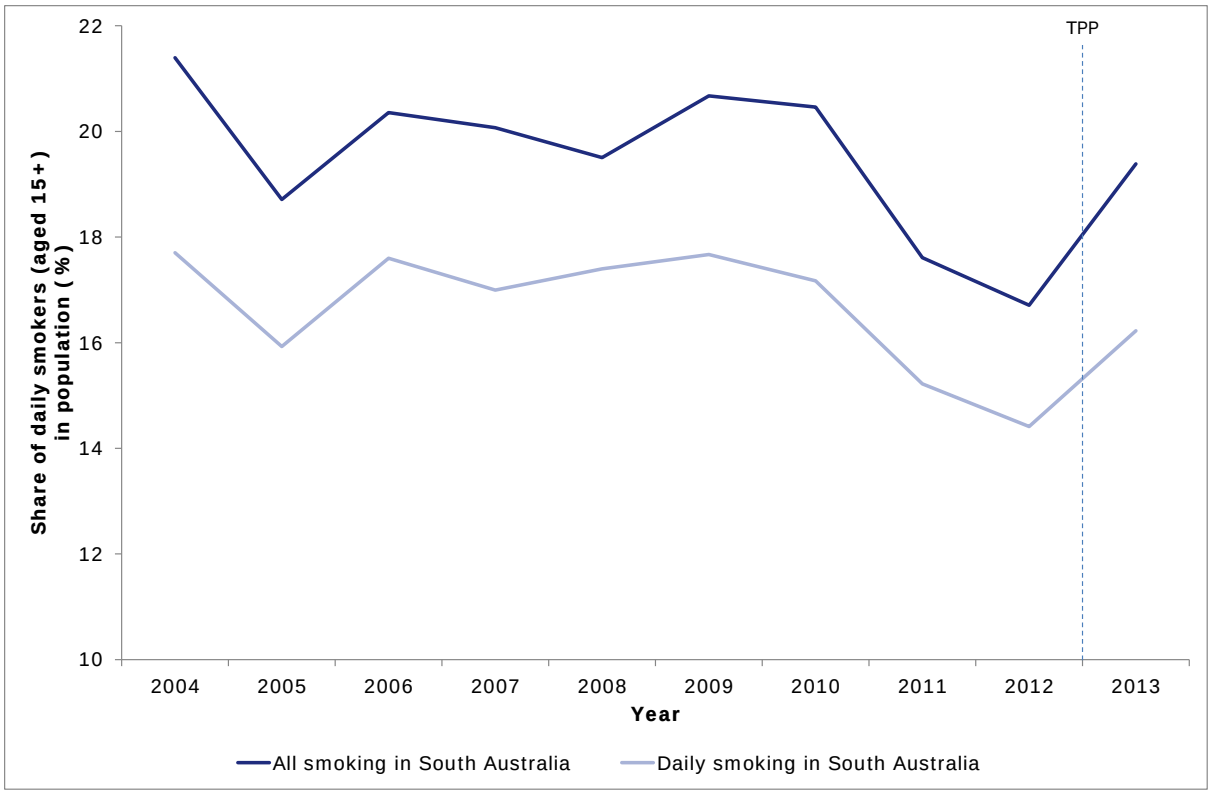
33. Australia considers that the RMSS data allow a more accurate comparison of smoking rates immediately prior to and following the introduction of the TPP measures. Based on the RMSS data, overall smoking prevalence in South Australia reduced significantly in the 12 months following the introduction of the TPP measures.[29] Australia's Post-Implementation Review Tobacco Plain Packaging 2016 also reports a reduction in daily smoking prevalence in South Australia based on SAHOS (data up to 2014) for individuals aged 15 years and over, as shown in **Figure C.14**, and the NDSHS (data up to 2013) for individuals aged 14 years and over.[30]

[28] See Dominican Republic's first written submission, para. 526.

[29] See Australia's first written submission, Annexure E, paras. 82-86; and Chipty Report, (Exhibit AUS-17), paras. 80-81.

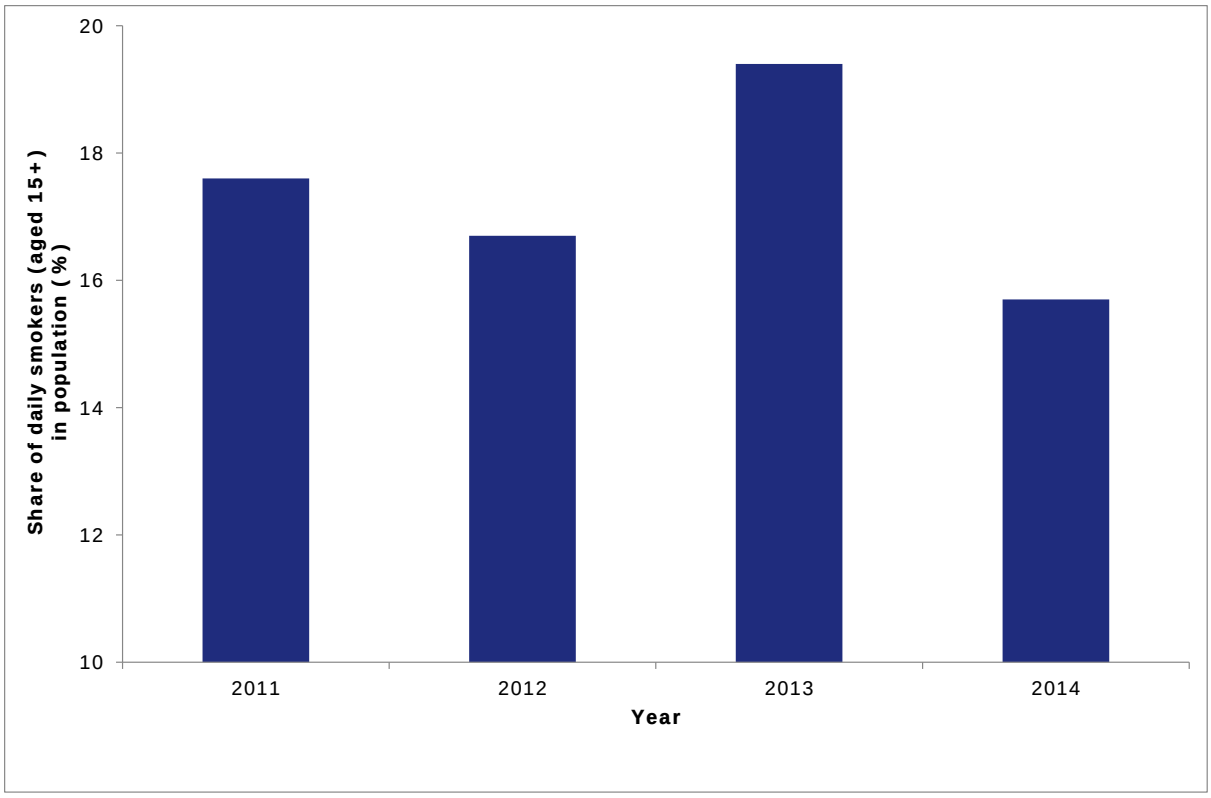
[30] See Tobacco Plain Packaging PIR, (Exhibit AUS-624), para. 128.

**Figure C.13: Smoking Prevalence in South Australia Based on SAHOS Data**




Note: The vertical dashed line indicates the introduction of TPP and enlarged GHWs.
Source: SAHOS data, (Exhibit DOM-93), p. 1.

**Figure C.14: Smoking Prevalence in South Australia Based on South Australia Survey Data**




Source: Tobacco Plain Packaging PIR, (Exhibit AUS-624), p. 42.

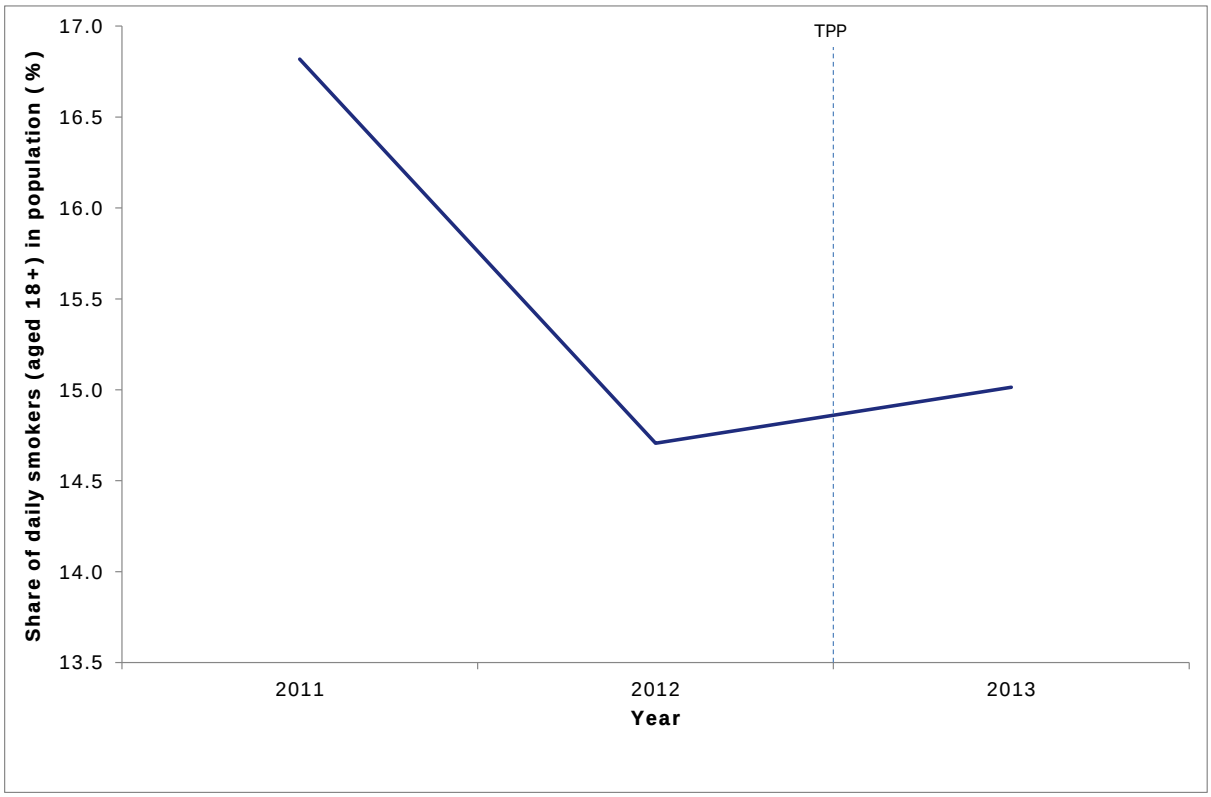
#### 1.1.7.4 Victorian Smoking and Health Survey

34. The Victorian Smoking and Health Surveys (VSHS) are cross-sectional telephone surveys undertaken with representative samples of adults aged 18 years and over and residing in the general population of the Australian state of Victoria. The surveys were undertaken from 2 November to 5 December 2011 (inclusive), from 1 November to 3 December 2012, and from 7 November to 11 December 2013.[31]

35. According to the Dominican Republic, the VSHS data demonstrate, as depicted in **Figure C.15**, that smoking prevalence of the responding population has increased, rather than decreased, in Victoria between 2012, the year prior to the implementation of the TPP measures, and subsequent periods.[32]

36. Australia claims that the VSHS data does not allow for a proper before/after analysis with respect to the introduction of the TPP measures because the 2012 survey was run from 1 November to 3 December 2012 when a majority of smokers were already using plain packaged products.[33] Australia further argues that, based on the RMSS data, overall smoking prevalence in Victoria reduced significantly in the 12 months following the introduction of the TPP measures.[34] Australia's Post-Implementation Review Tobacco Plain Packaging 2016 also reports a reduction in daily smoking prevalence in Victoria based on the NDSHS (data up to 2013) for individuals aged 14 years, as shown in **Figure C.16**.[35]

**Figure C.15: Smoking Prevalence in Victoria Based on VSHS Data**

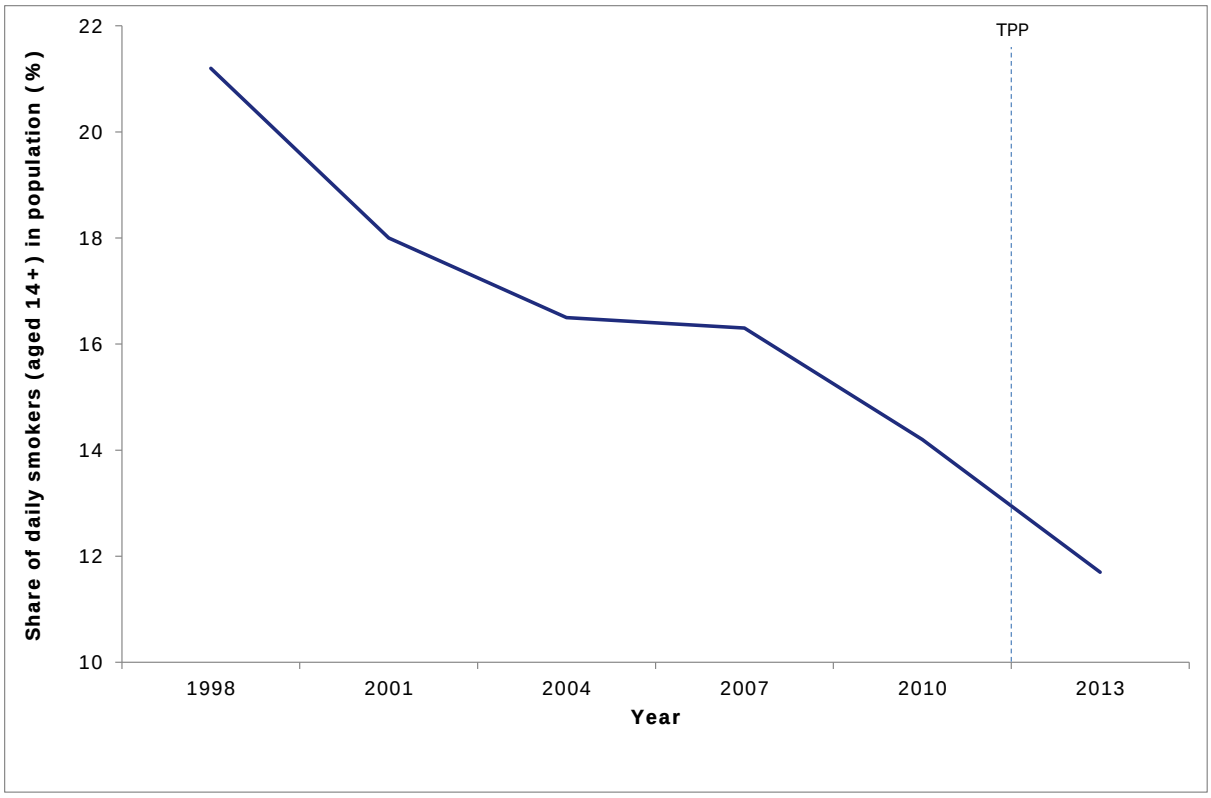



Note: The VSHS are conducted annually from November to December. The vertical dashed line indicates the introduction of TPP and enlarged GHWs.

Source: Dominican Republic's first written submission, p. 159.

[31] See Scollo et al. 2014, (Exhibits AUS-507, JE-24(57)).
[32] See Dominican Republic's first written submission, para. 526.
[33] See Australia's first written submission, Annexure E, para. 84.
[34] See Australia's first written submission, Annexure E, paras. 82-86; and Chipty Report, (Exhibit AUS-17), paras. 80-81.
[35] See Tobacco Plain Packaging PIR, (Exhibit AUS-624), para. 128.

**Figure C.16: Smoking Prevalence in Victoria Based on NDSHS Data**




Note: The vertical dashed line indicates the introduction of TPP and enlarged GHWs.

Source: Tobacco Plain Packaging PIR, (Exhibit AUS-624), p. 43.

### 1.1.7.5 Cancer Council Queensland Survey

37. Referring to a news item published in an Australian website, the Dominican Republic noted that while prevalence levels decreased for some subsets of the population, there had been a sharp increase in prevalence among Queenslanders aged between 25 and 34 after the implementation of the TPP measures.[36]

38. Australia submits that according to the RMSS data, overall smoking prevalence in Queensland decreased significantly in the 12 months following the introduction of the TPP measures.[37] More recently, Australia's Post-Implementation Review Tobacco Plain Packaging 2016 reports a reduction in daily smoking prevalence in Queensland based on Queensland survey (data up to 2014) for individuals aged 18 years and over and the NDSHS (data up to 2013) for individuals aged 14 years and over.[38]

## 1.2 Analysis by the Panel

39. We note at the outset the usefulness of relying on the most recent available (i.e. 2014 and early 2015) and comparable data to analyse trends in smoking prevalence. This is particularly important because the different datasets before us and presented above do not always cover the same period. For instance, the RMSS data continuously cover the period January 2001-September 2015, while the NDSHS covers the period 1993-2013 with data available every two or three years. We also note the importance of distinguishing between smoking prevalence, which measures the proportion of smokers in the population, and smoking incidence, which measures the proportion of smokers in a population of smokers and recent quitters. Unless

[36] See Dominican Republic's first written submission, para. 527; and ABC News 2014, (Exhibit DOM-96).

[37] See Australia's first written submission, Annexure E, paras. 82-86; and Chipty Report, (Exhibit AUS-17), paras. 80-81.

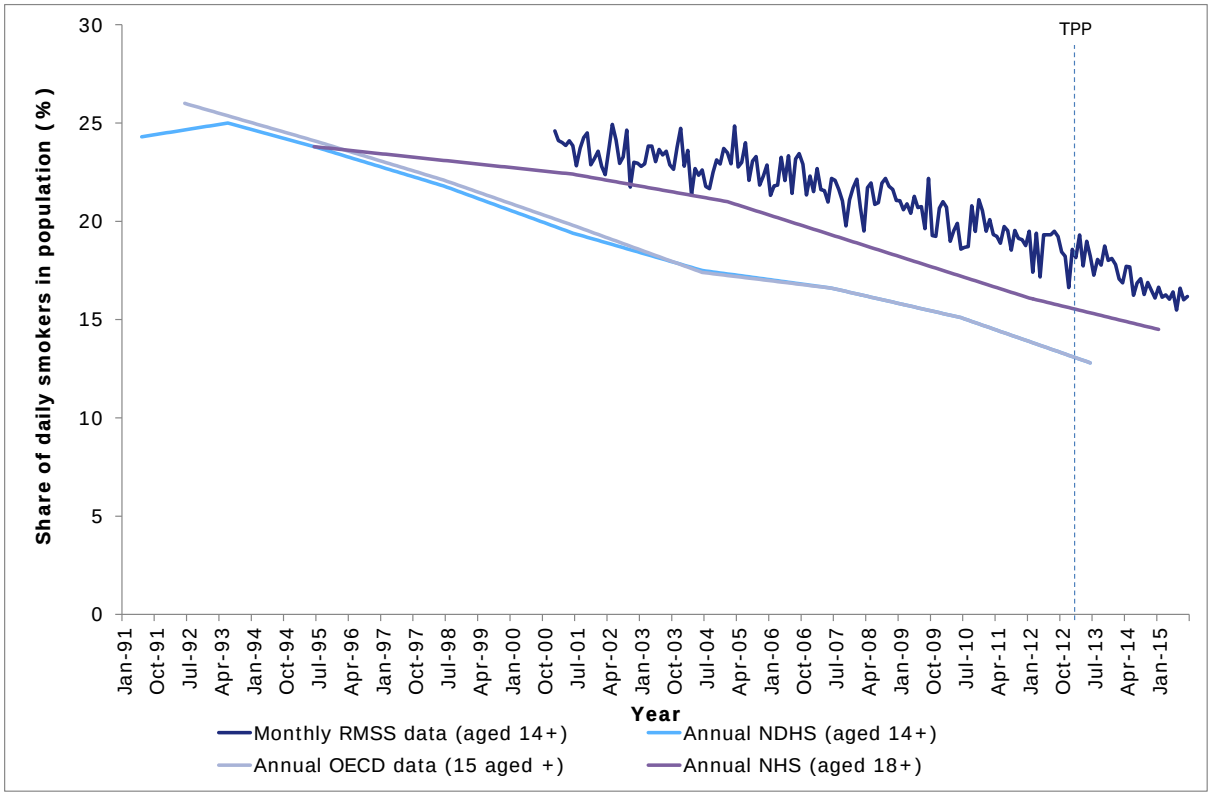
[38] See Tobacco Plain Packaging PIR, (Exhibit AUS-624), para. 128.

specified otherwise, we focus on smoking prevalence because, unlike smoking incidence, which ignores individuals who never smoked, smoking prevalence is based on the entire population (i.e. smokers, recent quitters and non-smokers).

40. After a careful review of the datasets described above, we observe that smoking prevalence fluctuates, even more when the unit of measure is disaggregated (monthly vs. yearly or state-level vs. country-level observations). We agree with Australia that in the presence of small sample sizes, in particular for subgroups such as youth and specific Australian states, it can be particularly difficult to interpret trends. That being said, we see, as depicted in **Figure C.17**, that despite different estimates and fluctuations of smoking prevalence, most of the datasets described above, including the RMSS data, OECD Dataset on Non-Medical Determinants, and NHS, show continuing declines in smoking prevalence at the national level in the period following the introduction of the TPP measures.[39]

41. While the most recent available data on smoking prevalence confirm that smoking prevalence in Australia continued to decrease following the introduction of the TPP measures, simply observing the existence of the trend does not inform, however, whether this downward trend in smoking prevalence has accelerated. This question is reviewed next.

**Figure C.17: Smoking Prevalence in Australia**




Note: The vertical dashed line indicates the introduction of TPP and enlarged GHWs.

Source: RMSS, NHS, NDHS and OECD data based on Chipty Third Rebuttal Report, (Exhibit AUS-605); Tobacco Plain Packaging PIR, (Exhibit AUS-624); NHS Results, (Exhibit AUS-622); and IPE Updated Report, (Exhibit DOM-303).

[39] We note that the smoking prevalence rates reported by the RMSS data are higher than the ones reported by the NHS and OECD data. These differences could be due to different methodology, timing and sample sizes. See Chipty Third Rebuttal Report, (Exhibit AUS-605); NHS Results, (Exhibit AUS-622); Tobacco Plain Packaging PIR, (Exhibit AUS-624); CCV 2014 Survey, (Exhibit AUS-621); and IPE Updated Report, (Exhibit DOM-303).

## 2 WHETHER THE REDUCTION OF SMOKING PREVALENCE ACCELERATED FOLLOWING THE IMPLEMENTATION OF THE TPP MEASURES (ROY MORGAN SINGLE SOURCE)

42. As discussed above, the majority of the most recent datasets presented to us show continuing declines in smoking prevalence in Australia in the period following the introduction of the TPP measures. Rather than assessing directly the TPP measures' impact on smoking prevalence, which will be discussed in detail next, the Dominican Republic's experts initially investigated whether there was a shift in smoking prevalence in the post-TPP implementation period. In other words they assessed whether the reduction of smoking prevalence accelerated or slowed down following the implementation of the TPP measures. According to them, if the reduction in smoking prevalence follows the same pre-existing downward trend after the introduction of the TPP measures, this implies that the TPP measures have not reduced smoking prevalence.

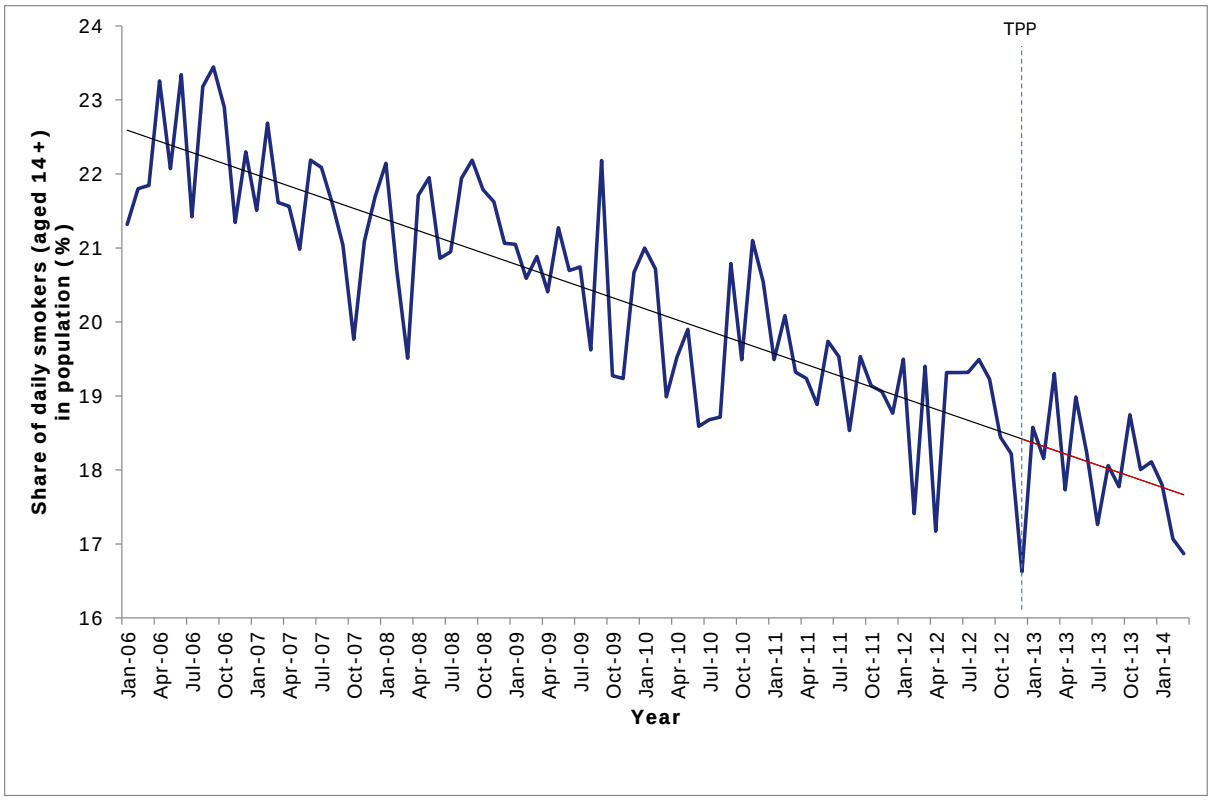
### 2.1 Datasets and related studies

43. The Dominican Republic first submitted, through IPE, a statistical trend analysis of smoking prevalence using the RMSS dataset. The trend analysis consists of (1) estimating the time trend of smoking prevalence for the pre-TPP implementation period (before December 2012); (2) predicting the prevalence rate that would have been obtain in any given month following the implementation of the TPP measures on 1 December 2012, in the absence of the TPP measures using the pre-TPP implementation trend; and (3) determining whether the difference between the observed prevalence and the estimated counterfactual prevalence is different from zero by computing confidence intervals.[40]

44. The trend analysis is undertaken by estimating either a quadratic time trend for the January 2001-March 2014 period or a linear trend model for the January 2006-March 2014 period. In both cases, IPE concludes that there is no statistical difference between observed smoking prevalence of the full population and the estimated counterfactual prevalence of the full population with the exception of the month of December 2012, implying overall that the post-implementation trend did not shift. Similar results are found when the analysis focuses only on minor population and young adult population.[41]

[40] See IPE Report, (Exhibit DOM-100), pp. 26-27.
[41] See IPE Report, (Exhibit DOM-100), pp. 28-35, 105-116, 119-156, and 181-199.

**Figure C.18: Smoking Prevalence and Linear Trend**




Note: The vertical dashed line indicates the introduction of TPP and enlarged GHWs. The black and red lines represent the linear trends.

Source: RMSS data, based on IPE Report, (Exhibit DOM-100), p. 28.

45. Australia's expert, Dr Chipty, rejects the IPE Report's results on the grounds, *inter alia*, that (1) IPE asserts without support that historical trends will continue into the future in the absence of new regulatory measures; (2) IPE does not attempt to evaluate the extent to which past policies contributed to the trend in prevalence; and (3) IPE's model design makes it less likely, and sometimes impossible, to find a policy effect.[42] Australia posits that trends in available data show that overall prevalence in Australia is declining over time, including following the TPP measures. However, given the small sample sizes for certain subgroups, like adolescents or residents of specific Australian states, trends in prevalence can be difficult to interpret.

46. Australia also submits another expert report by Professor Scharfstein, who further argues that (1) IPE's assumption that smoking prevalence would have continued to decline at the same rate after December 2012, even if the TPP measures had never been introduced, is entirely unsupported without assumptions or a valid natural experiment; (2) IPE's date restriction (i.e. January 2006) in the linear trend model is derived by simply looking at the data; (3) IPE's statistical trend analysis lacks statistical rigor by not specifying a null hypothesis to evaluate whether there is a TPP measures' effect; and most importantly (4) IPE's statistical trend analysis has low statistical power and is inadequate to detect important declines in smoking prevalence after the introduction of the TPP measures.[43]

47. Professor Scharfstein submits that IPE's methodology applied to the excise tax increase introduced by Australia in April 2010 does not identify a reduction in smoking prevalence after the tax increase, although the complainants have argued that excise taxes, in general, are an effective

[42] See Chipty Report, (Exhibit AUS-17), paras. 40-43.
[43] See Scharfstein Report, (Exhibit AUS-20), paras. 11-12, 35, and 37-64.

tool to discourage smoking. Professor Scharfstein further argues that monthly data from the RMSS data cannot reasonably rule out important declines in smoking prevalence in the post-TPP period.[44]

48. The Dominican Republic's and Indonesia's expert, Professor List, disagrees with Professor Scharfstein and argues that according to Borland (2010) the RMSS data have power to detect small effects. Professor List further argues that the justification for Scharfstein's linear and quadratic time trends starting in 2001 and 2006, respectively, are unsubstantiated. Given the nature of the data, especially the seasonality, the start date chosen will affect the estimate of the downward trend absent of the TPP measures, and therefore change the null hypothesis used. According to Professor List, Professor Scharfstein should have simultaneously considered a gradual descent and a step down of smoking prevalence. Since the only significant decrease in prevalence is December 2012 when the TPP measures were introduced, it is possible that there is a large initial deviation in smoking prevalence followed by a return to the pre-implementation trend (mean regression).[45]

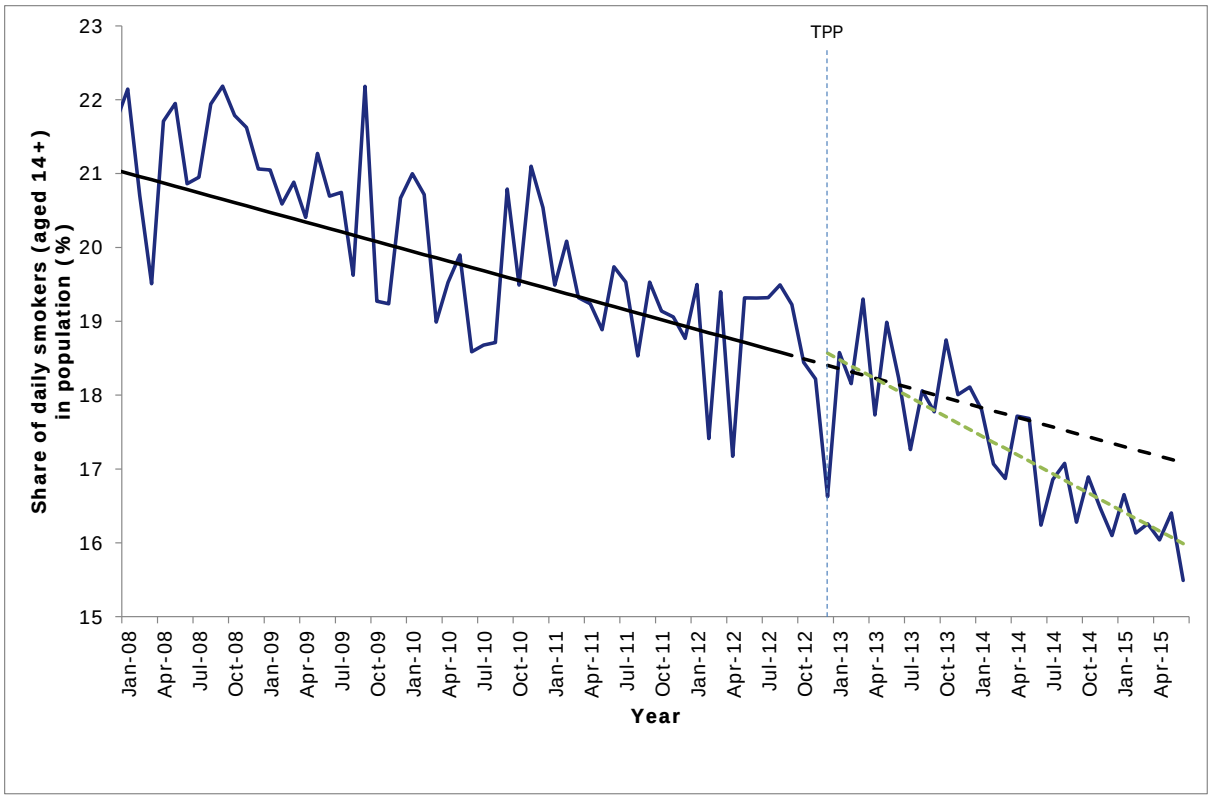
49. More generally, Professor List submits that abstracting from the general downward trend in smoking prevalence, there is no sustained change in the previously existing rate of prevalence following the vertical dashed line, depicted in a figure similar to **Figure C.18**, that denotes the start of the TPP measures in Australia.[46]

50. Australia's expert, Dr Chipty, contests Professor List's claim that the RMSS data have "power to detect small effects". In Dr Chipty's view, using the classification system in Borland (2010), IPE's methodology does not have sufficient power to detect "small" (less than 0.5 standard deviations of trend), "medium" (between 0.5 and 1), or even most "large" (greater than 1) effects.[47]

51. Dr Chipty further contends that, unlike Professor List's claim of no sustained change in the smoking prevalence trend following the introduction of the TPP measures, allowing the trend line to be different before and after the TPP measures shows, as depicted in **Figure C.19**, a break in the trend of smoking prevalence after the TPP measures' implementation, about one percentage point lower relative to where it would have been by June 2015.[48, 49]

[44] See Scharfstein Report, (Exhibit AUS-20), paras. 12, 51-55, and 65-68.
[45] See List Report, (Exhibit DOM/IDN-1), paras. 109-112.
[46] See List Report, (Exhibit DOM/IDN-1), paras. 22-23.
[47] See Chipty Rebuttal Report, (Exhibit AUS-535) (SCI), paras. 4 and 8.
[48] See Australia's comments on the complainants' responses to Panel question No. 146, para. 15; and Chipty Second Rebuttal Report, (Exhibit AUS-591), paras. 8-12. A similar graphic for the period January 2001-September 2015 is included in the Tobacco Plain Packaging PIR (Exhibit AUS-624), p. 35.
[49] See Australia's comments on the complainants' responses to Panel question No. 146, para. 15.

**Figure C.19: Smoking Prevalence and Pre- and Post-TPP Trends**




Note: The vertical dashed line indicates the introduction of TPP and enlarged GHWs. The dashed line and the dotted line denote, respectively, the pre-TPP linear trend and the post-TPP linear trend.

Source: RMSS data, based on Chipty Second Rebuttal Report, (Exhibit AUS-591), p. 10.

52. A similar figure for the period January 2001-September 2015 is included in the Post-Implementation Review Tobacco Plain Packaging 2016, with which the Dominican Republic took issue. According to the Dominican Republic, the figure is misleading, for at least three reasons: (1) it glosses over the clear break in trend that occurred in June 2006, as demonstrated by IPE[50]; (2) October, November, and December 2012 are omitted in the construction of the pre-TPP trend, thus artificially making the pre-implementation trend line less steep; (3) the post-implementation trend is inconsistent with previous visualizations adduced by Dr Chipty during the WTO proceedings. The Dominican Republic submits that given Dr Chipty's failure to provide back-up material for her analysis, this inconsistency is difficult to explain.[51]

## 2.2 Analysis by the Panel

53. We note at the outset that the parties have discussed extensively whether there is a secular and long-term downward trend in smoking prevalence in Australia and how to specify the smoking prevalence trend (e.g. linear or quadratic) in different contexts related to the contribution of the TPP measures. This will be discussed more extensively in the next subsection, which reviews the econometric studies aimed at assessing the impact of the TPP measures on smoking prevalence and smoking incidence.

54. We further note that, although more recent data were available over the course of the proceedings, IPE did not update the results of its statistical trend analysis. Instead, in reply to some of the criticism raised by Australia's experts, IPE proposed a "modified trend analysis", which acknowledges that the original statistical trend analysis does not control for other relevant

[50] See IPE Report, (Exhibit DOM-100), p. 109; and IPE Third Updated Report, (Exhibit DOM-375), paras. 32-49.

[51] See Dominican Republic's comments on Australia's Post-Implementation Review, paras. 69-70 and fn 85.

variables explaining the evolution of smoking prevalence beyond the trend. The "modified trend analysis" is based on an entirely different methodology discussed in the next subsection.

55. In this context, we consider that the results of the original trend analysis provided by the IPE Report are not informative, not only because it fails to control for other relevant variables affecting smoking prevalence but also because it is unclear if the results of the trend analysis would have changed with the more recent data made available to us.

56. Overall, based on the most recent available RMSS data, we note that smoking prevalence in Australia has not only continued to decrease following the introduction of the TPP measures, but that the downward trend in smoking prevalence has accelerated with a steeper slope of the smoking prevalence trend between December 2012 and September 2015 (latest available observations submitted by the parties) compared to the pre-TPP periods.

57. That being said, this change in the smoking prevalence trend following the introduction of the TPP measures does not necessarily imply that the TPP measures are having a statistically significant impact, to the extent that other factors could explain such evolution, including other tobacco control policies. Indeed, most of the discussion among the parties' experts focused on assessing econometrically the impact of the TPP measures on smoking prevalence by controlling for other factors. This is reviewed next.

## 3 WHETHER THE TPP MEASURES CONTRIBUTED TO THE REDUCTION IN SMOKING PREVALENCE

58. The parties disagree on whether the TPP measures contributed to reducing smoking prevalence, i.e. had a negative impact on smoking prevalence. This issue gave rise to detailed technical exchanges between IPE for the Dominican Republic, Professor List for the Dominican Republic and Indonesia, Professor Klick for Honduras, and Dr Chipty for Australia, who each proposed different econometric methods to estimate the TPP measures' impact on smoking prevalence or smoking incidence. As mentioned above, all parties, however, recognize that the empirical econometric evidence on smoking prevalence submitted does not distinguish between the impact of TPP and the impact of enlarged GHWs, because both measures were implemented at the exact same time.[52]

59. Australia has argued that the impacts of the TPP measures on prevalence may not fully manifest in short term datasets[53] and that it is inappropriate to seek to judge the efficacy of the TPP measures on the basis of limited short term datasets.[54] First, reducing smoking prevalence is a long-term objective.[55] Second, large changes in the root behaviours (i.e. initiation, cessation, and relapse) stemming from the TPP measures are likely masked in smoking prevalence because of the stock of current smokers whose behaviours may not be as affected by the TPP measures.[56] Notwithstanding this position, Australia re-estimated the impact of the TPP measures on smoking prevalence, in response to the submissions of the Dominican Republic, Honduras and Indonesia. During the course of the proceedings and as more prevalence-related data became available, Australia argued that using the most appropriate dataset available and correcting for flaws in the econometric models put forward by the experts of the Dominican Republic, Honduras, and Indonesia, the results show that the TPP measures have already contributed to the reduction of cigarette and cigar smoking prevalence.[57]

[52] See List Report, (Exhibit DOM/IDN-1), para. 113.

[53] See Australia's response to Panel questions No. 126 and 200 and Australia's first written submission, para. 670.

[54] See Australia's first written submission, para. 17.

[55] See Australia's first written submission, para. 670. Instead Australia claimed that the TPP measures' impact should be investigated through its mechanisms by looking at its impact on non-behavioural outcomes, that is, (1) reduction in the appeal of tobacco products, (2) increased effectiveness of health warnings, and (3) reduction of the ability of the pack to mislead. The evidence on these non-behavioural outcomes is considered in Appendix A.

[56] See Chipty Report, (Exhibit AUS-17), paras. 32-39; Scharfstein Report, (Exhibit AUS-20), para. 68.

[57] See Australia's opening statement at the second hearing of the Panel, paras. 110-111; and comments on the complainants' responses to Panel question No. 197, para. 214.

60. A particular feature of the exchange between the parties on this issue is that, on several occasions, the experts of the Dominican Republic, Honduras and Indonesia proposed in their rebuttal reports new models or methodologies, or both, that sometimes contradict some of the approaches taken in their earlier reports and invalidate the results reported in those reports.[58] For this reason, the description of the approaches and results below is based primarily on the most recent expert reports submitted by the parties.[59]

## 3.1 Datasets and related studies

### 3.1.1 Roy Morgan Single Source

61. The RMSS data were used by the experts of the Dominican Republic, Honduras, Indonesia, and Australia to assess whether the TPP measures have contributed to the reduction in smoking prevalence.

62. Australia and its expert Professor Scharfstein consider that the RMSS data covering 15 post-implementation months (December 2012 to March 2014) cannot rule out important declines in smoking prevalence in the post-policy period.[60] Australia's expert Dr Chipty nonetheless considers the RMSS data, which is a nationally representative dataset available from January 2001 to September 2015, to be her "preferred data source" for the analysis of smoking prevalence.[61]

#### 3.1.1.1 IPE Reports

63. The Dominican Republic submitted five reports prepared by IPE aimed at estimating econometrically the TPP measures' impact on, among other things, smoking prevalence using the RMSS data. Throughout these IPE Reports, different econometric approaches have been proposed: (1) statistical trend analysis[62]; (2) two-stage micro-econometric analysis; (3) modified trend analysis; (4) autoregressive integrated moving average with explanatory variable (ARIMAX) model; and (5) one-stage micro-econometric analysis.[63] The first two approaches were only presented in the first IPE Report.[64] The second and subsequent IPE Reports only focused on the

[58] For instance, as discussed in detail below, IPE initially proposed to control for excise tax increases by including (dummy) indicator variables for each excise tax increase (IPE Updated Report, (Exhibit DOM-303)), but subsequently contended that a more appropriate measure to capture the excise tax increases is the weighted average price per cigarette in Australia (IPE Second Updated Report, (Exhibit DOM-361)). Similarly, Professor List and IPE initially applied the STATA software command ivreg2 in order to calculate standard errors that are robust to heteroscedasticity and serial correlation using the automatic bandwidth selection procedure by Newey and West 1994. IPE Updated Report, (Exhibit DOM-303); and List Report, (Exhibit DOM/IDN-1). Subsequently, both Professor List and IPE applied an alternative way of calculating standard errors, that, according to them, is adjusted to reflect more accurately the original suggestion by Newey and West (1994). IPE Second Updated Report, (Exhibit DOM-361); and List Rebuttal Report, (Exhibit DOM/IDN-3).

[59] We note that we have nonetheless considered all the relevant evidence before us, including all the expert reports, including the methodologies and models contained therein.

[60] Australia's first written submission, Annexure E, para. 45; and Scharfstein Report, (Exhibit AUS-20), para. 68.

[61] See Chipty Third Rebuttal Report, (Exhibit AUS-605), para. 70.

[62] As described above, given that the statistical trend analysis does not specify explicitly the impact of the TPP measures, we consider it to be different from the remaining approaches, which explicitly control for the TPP measures and other relevant variables beyond the trend.

[63] See IPE Report, (Exhibit DOM-100); IPE Updated Report, (Exhibit DOM-303); IPE Second Updated Report, (Exhibit DOM-361); IPE Third Updated Report, (Exhibit DOM-375); and IPE Summary Report, (Exhibit DOM-379).

[64] The two-stage micro-econometric analysis initially proposed but then set aside by IPE was the object of several critiques raised by Australia's expert, Professor Scharfstein. Professor Scharfstein argues that the analysis misinterprets the constant terms in the first-stage model by considering the estimated likelihood of smoking for a specific nonsensical subgroup of individuals: males, age zero, with zero years of education, who are positioned at the very top of the income distribution. In Professor Scharfstein's view, this misinterpretation is important because these constant terms serve as the critical input to IPE's second-stage analysis and are the basis of IPE's conclusions about the effects of the TPP measures. Professor Scharfstein submits that the results of IPE's two-stage micro-econometric analysis would show evidence of a TPP effect if different demographic subgroups were considered by redefining ("centering") the variables in the first-stage model. See Scharfstein Report, (Exhibit AUS-20), paras. 73-84.

last three approaches, which to some extent have been adopted to address some of the criticisms raised by Australia's experts, Dr Chipty and Professor Scharfstein. As well as using different estimation techniques, these approaches differ also in terms of model specifications, that is the set of explanatory variables used to explain the evolution of smoking prevalence and that are included explicitly in the model, such as tobacco policies and demographic variables. The only variable common to all econometric approaches is a variable capturing the smoking prevalence trend, although the trend is assumed to be linear in some cases and quadratic (i.e. curved) in others depending on the sample period covered. The most recent econometric analysis presented in the IPE Report covers the period July 2006 to September 2015.

64. Overall, IPE concludes that the TPP measures had no statistically significant effect on general smoking prevalence and on cigar smoking prevalence.[65] Other factors explain the reduction in smoking prevalence, such as the overall declining trend. According to IPE, these results are robust across different specifications (e.g. different TPP measures' starting date: October, November and December; controlling for the weighted average price of cigarettes and/or tax policy change, population sample weighting changes by Roy Morgan Research, extending the sample period back to January 2001-September 2015).

65. Australia's expert, Dr Chipty, rejects the econometric results of the IPE Reports on various technical grounds. First, according to Dr Chipty, there is no credible justification for excluding data prior to July 2006 and the analysis should cover the 2001-2015 period.[66] Second, Dr Chipty rejects IPE's assertion that the STATA software command, ivreg2, used to estimate standard errors robust to "heteroscedasticity"[67] and "autocorrelation"[68] (Newey-West standard errors) is wrong, noting that IPE used the same STATA command, before Professor List claimed he found an error in the STATA programming code.[69] Third, Dr Chipty argues that controlling for prices inclusive of tax while attempting to measure the effects of a tax hike ignores the TPP measures' effect on price, leading the TPP indicator variable to capture only a partial effect of the TPP measures, and the price variable capturing the rest of its effect. Dr Chipty further notes that IPE initially controlled for excise tax increase with indicator variables in its model specification but then decided to replace it with a price variable.[70] Fourth, and similarly, Dr Chipty submits that the use of single tax level variable, as proposed by IPE, is only valid under the proportionality assumption (i.e. the effect of tax changes on prevalence is proportional to the size of the tax change), which may only be satisfied in some specifications (for instance, the model of cigar smoking prevalence does not satisfy this assumption).[71] Finally, Dr Chipty argues that the inability of IPE to measure a negative and statistically significant effect of the TPP measures on prevalence stems from challenges associated with the inclusion of a time trend: (a) the inclusion of a linear trend (for the

IPE argues in its report that Professor Scharfstein's analysis is misspecified, and that his findings are entirely driven by this misspecification, which attributes the effects of the secular downward trend to the TPP measures. See IPE Updated Report, (Exhibit DOM-303), paras. 69-74.

Professor List criticises the fact that Professor Scharfstein's analysis suffers from the same issues Professor Scharfstein argues affect the constructed constant term in the IPE Report by constructing the constant term for individuals that do not exist in the data and failing to construct the constant term for some individuals. Professor List further argues that Professor Scharfstein's reanalysis overstates the importance of the TPP measures' impact for two reasons. First, the statistical analysis does not account for the fact that this is a case of multiple hypothesis testing, which increases the probability of at least one incorrect rejection. Second, the statistical analysis does not account in the second stage analysis for all background changes not related to the policy, such as tax increases and macroeconomic factors. See List Report, (Exhibit DOM/IDN-1), paras. 113-122.

Professor Scharfstein contends that both IPE and Professor List have misinterpreted his critiques of the two-stage micro-econometric analysis, confusing his demonstration of its flaws with an affirmative opinion about the effect of the TPP measures. In his view, the two-stage micro-econometric analysis, even after the modification he made to it, remains flawed and should not be used to assess the impact of the TPP measures. See Scharfstein rebuttal report, (Exhibit AUS-587), paras. 12-19.

[65] See IPE Summary Report, (Exhibit DOM-379), paras. 11-24, Appendix 6.

[66] See Chipty Surrebuttal Report, (Exhibit AUS-586), para. 9; and Chipty Third Rebuttal Report, (Exhibit AUS-605), paras. 34-35.

[67] Heteroscedasticity arises when the regression errors variances are not constant across observations. W. H. Green, *Econometric Analysis*, 5th edn (Prentice Hall, 2002).

[68] Autocorrelation or serial correlation of the disturbances arises when the disturbances are correlated across periods. W. H. Green, *Econometric Analysis*, 5th edn (Prentice Hall, 2002).

[69] See Chipty Second Rebuttal Report, (Exhibit AUS-591), paras. 13-31.

[70] See Chipty Surrebuttal Report, (Exhibit AUS-586), para. 12.

[71] See Chipty Third Rebuttal Report, (Exhibit AUS-605), paras. 22-26 and 29-32.

July 2006-September 2015 sample); or (b) allowing the trend line to both shift and change slope at July 2006 (for the January 2001-September 2015). In each case, Dr Chipty notes that the time trend absorbs all policy effects, noting that in most specifications the impact of excise tax (or tobacco price), considered by all complainants as one of the most effective tobacco control measures, is not statistically significant.[72]

#### 3.1.1.2 Professor List's reports

66. The Dominican Republic and Indonesia's expert Professor List reconsiders the RMSS data and estimates a two-stage micro-econometric model. The first stage estimates the likelihood of a representative person being a smoker for each month controlling for demographic characteristics (e.g. age, fourth order polynomials[73] of gender, education, and income). The second stage conducts a before and after analysis on the likelihood for each month computed in the first stage by estimating a linear probability model controlling for the TPP measures, price and/or linear trend, and weighting changes by Roy Morgan Research. The most recent econometric analysis presented in Professor List's report covers the period from July 2006 to September 2015.

67. Overall, Professor List concludes that the TPP measures had no statistically significant effect on the likelihood of a representative Australian smoker or of a representative Australian minor and young adult smoking.[74] According to Professor List, these results are robust across different specifications (e.g. different TPP starting date: October, November and December). The results are also robust to an alternative way of calculating the explanatory variables' standard errors that are robust to heteroscedasticity and autocorrelation and correct compared to the STATA software package ivreg2 used by Dr Chipty. The standard error of the explanatory variables, including the TPP measures, is important to determine whether the respective variable is statistically different from zero or not. According to Professor List, his new procedure to compute standards errors follows exactly the procedure described in the seminal article of Newey and West (1994) in order to select the maximum amount of time that the data can be correlated over time (maximum lag selection). The same method was also used in the latest IPE Reports.

68. Australia's expert, Dr Chipty, submits that the econometric results on smoking prevalence of Professor List are flawed for many of the same reasons formulated with respect to IPE Reports. In particular, Dr Chipty contests restricting the sample to the period 2006-2015; using a price variable without correcting for the TPP measures' effect on price; including a trend variable that absorbs all policy effects, and the assertion that the STATA software command ivreg2, used to estimate robust standard errors, is wrong. In addition, Dr Chipty considers Professor List's attempts to correct for the sample re-weighting contained in the RMSS data to be unsound, noting that this was an entirely new concern about the RMSS data that had not been raised before in any of the expert debate.[75]

#### 3.1.1.3 Professor Klick's reports

69. Honduras's expert Professor Klick also reconsiders the RMSS data and estimates a two-stage micro-econometric model. The first stage estimates the TPP measures' impact on tobacco price controlling for excise tax indicators. The second stage estimates the TPP measures' impact on the likelihood a respondent is a smoker by estimating a linear probability model controlling for the TPP measures, the logarithm of tobacco price instrumented by excise tax indicators in the first stage, quadratic trend, 2006 GHW and demographic characteristics (e.g. sex, married status, urban status, age, income, education, territory and job category). The TPP measures' impact corresponds to the sum of the direct effect of TPP measures on the probability a respondent is a smoker and the indirect effect of the TPP measures defined as the product of the TPP measures' effect on price and the effect of price change on the probability a respondent is a smoker.[76] The most recent econometric analysis presented in Professor Klick's Report covers the period from January 2001 to June 2015.

[72] See Chipty Third Rebuttal Report, (Exhibit AUS-605), paras. 36-42.
[73] A fourth order polynomial of a given variable $x$ corresponds to $x^1$, $x^2$, $x^3$ and $x^4$.
[74] See List Third Supplemental Report, (Exhibit DOM/IDN-7), paras. 78-80.
[75] See Chipty Second Rebuttal Report, (Exhibit AUS-591), paras. 32-38.
[76] See Klick Third Supplemental Rebuttal Report, (Exhibit HND-166), para. 21.

70. Overall, Professor Klick concludes that the TPP measures had no statistically significant effect on the likelihood that a respondent is a smoker.[77] Professor Klick further re-estimates his model focusing on young people (14-24 years old). While the TPP measures' impact is not statistically significant for the age groups 14-17 and 18-19, the TPP measures' impact is positive and statistically significant for the age group 20-24.[78] According to Professor Klick, these results are robust across different specifications (e.g. different TPP starting date: October, November and December).

71. Dr Chipty rejects Professor Klick's use of a quadratic time trend variable, because it absorbs virtually all of the variation in smoking prevalence and leaves no room for any other explanatory variables, including price, to have a measurable effect on smoking prevalence.[79] In addition, Dr Chipty argues that Professor Klick's two-steps instrument variables approach does not yield reliable estimates for the standard error and confidence interval of the TPP measures' total effect.[80]

#### 3.1.1.4 Dr Chipty's reports

72. Dr Chipty considers the RMSS data to be her preferred data for assessing the TPP measures' impact on smoking prevalence because of its large sample size, sufficient pre-period, and national representativeness. However, as explained above, Dr Chipty rejects the econometric results of the reports by IPE, Professor List and Professor Klick on various technical grounds. Dr Chipty addresses these critics by re-estimating the models of smoking prevalence developed by IPE (one-stage micro-econometric analysis and modified trend analysis) by extending the period of analysis from January 2001 to September 2015, controlling for a linear trend, a 2006 GHW dummy, a set of excise tax indicators and a set of socio-demographic characteristics, and correcting for robust standard errors (when necessary). Dr Chipty also re-estimates the model of smoking prevalence developed by Professor List (two-stage micro-econometric analysis) by correcting for robust standard errors. Similarly, Dr Chipty re-estimates Professor Klick's two-stage instrument variables micro-econometric analysis by replacing the quadratic linear variable with a linear trend.

73. Overall, Dr Chipty concludes that the TPP measures had a negative and statistically significant effect on smoking prevalence.[81] Most of the results show statistically significant declines in smoking prevalence even when using Professor List's own standard error calculation procedure. The negative and statistically significant impact of the TPP measures on smoking prevalence is also robust to controlling for the reweighting of the RMSS data. Similarly, replacing the tax indicator variables with a single tax level variable produces similar point estimates of the TPP measures' effect.

74. IPE, Professor List and Professor Klick raise a number of criticisms of Dr Chipty's econometric approaches. According to IPE, Dr Chipty's model specifications covering the January 2001-September 2015 period fail to account for two different linear downward trends in smoking prevalence: one for the January 2001-June 2006 period and another for the July 2006-September 2015 period. As a result, Dr Chipty overestimates the downward trend in smoking prior to June 2006, while underestimating it afterwards.[82]

75. IPE further contends that the use of tax hike dummies, as proposed by Dr Chipty, are inferior control variables compared to cigarette prices and tax levels for three reasons: (i) consumers base their choices on what cigarettes cost; (ii) it is important to control for, by how much a certain price increase affects prevalence, and not only whether it affects prevalence; and (iii) tax hike dummies

[77] See Klick Third Supplemental Rebuttal Report, (Exhibit HND-166), paras. 11-23.
[78] See Klick Third Supplemental Rebuttal Report, (Exhibit HND-166), paras. 54-55.
[79] See Chipty Third Rebuttal Report, (Exhibit AUS-605), paras. 49-53.
[80] See Chipty Third Rebuttal Report, (Exhibit AUS-605), paras. 54-60.
[81] See Chipty Third Rebuttal Report, (Exhibit AUS-605), paras. 29-32, 52-53 and 70-73; and Chipty Second Rebuttal Report, (Exhibit AUS-591), paras. 28-38, 53-54, Appendix D.
[82] See IPE Summary Report, (Exhibit DOM-379), paras. 50-52.

(or tax levels) do not control for changes in tobacco affordability other than the price changes resulting from the tax hikes themselves.[83]

76. Professor List submits that Dr Chipty's opposition to using control variables for reweighting events of the population sample in the RMSS data may stem from the fact that an adjustment for reweighting would reverse Dr Chipty's estimates of the TPP measures' effect.[84] Failure to correct for reweighting changes may thus lead to misspecification.[85]

77. Professor List, IPE and Professor Klick argue that Dr Chipty computed incorrect standards errors by applying the STATA command ivreg2, which substantially deviates from the seminal article by Newey and West (1994). Professor List further submits that Dr Chipty's method to compute standards errors assigns statistical significance nearly 400% more frequently than it should. Applying the exact procedure proposed by Newey and West (1994), Professor List and IPE conclude that most of Dr Chipty's specifications no longer exhibit a statistically significant TPP effect.[86]

### 3.1.2 Commissioned Roy Morgan Research Survey (Australia and New Zealand)

78. Honduras's expert Professor Klick proposes comparing Australia's smoking behaviour before and after the introduction of the TPP measures with another jurisdiction that has not implemented plain packaging during the same period. Professor Klick posits that although Australia is unique in many ways, it is reasonable to use New Zealand as a reliable counterfactual jurisdiction, because (i) both countries share many similarities culturally, historically, and demographically; (ii) both countries are in the same region and share the same seasonality (useful when dealing with subannual measures); (iii) the governments of both countries themselves also recognize that they are especially comparable along dimensions such as health behaviour and socio-economic issues; (iv) tobacco prices, including taxes, are comparable between both countries; and (v) smoking rates in both countries are highly correlated (0.95 for the period 1964-2012 based on OECD data on the fraction of residents ages 15 years and older who are daily smokers).[87] Professor Klick's analysis, defined as a difference-in-difference estimation, covers the period from November 2012 to February 2014.

79. Overall, Professor Klick concludes that the TPP measures' impact on daily or occasional smoking in Australia relative to New Zealand is not statistically significant.[88] Similar findings are obtained by using different estimators (linear probability model and logit model and fixed effects), restricting the sample to only individuals who answered the survey in the 6 waves (to mitigate attrition), restricting the sample to only individuals who answered the survey at least in one post-PP waves (to mitigate attrition), restricting the sample to Wave 1 and Wave 6 (to not overweight immediate responses), restricting the sample to smokers planning to quit during Wave 1, and controlling for individual characteristics (income, unemployed, male, age).

80. Dr Chipty argues that Professor Klick's difference-in-differences analysis of smoking incidence is invalid because Professor Klick's commissioned Roy Morgan survey does not contain a pre-period and is incapable of distinguishing which respondents had noticed plain packaging. Furthermore, Dr Chipty is of the view that New Zealand is an invalid counterfactual for the purposes of studying the effects of the TPP measures, because of New Zealand's January 2013 excise tax increase, a month after the TPP measures' introduction.[89]

---

[83] See IPE Third Updated Report, (Exhibit DOM-375), para. 139; and IPE Summary Report, (Exhibit DOM-379), paras. 56-60.

[84] See List Third Supplemental Report, (Exhibit DOM/IDN-7), paras. 51-55.

[85] See IPE Summary Report, (Exhibit DOM-379), paras. 53-54.

[86] See List Third Supplemental Report, (Exhibit DOM/IDN-7), paras. 42-50; List Summary Report, (Exhibit DOM/IDN-9), paras. 83-98; IPE Summary Report, (Exhibit DOM-379), paras. 61-65; and Klick Fourth Supplemental Rebuttal Report, (Exhibit HND-169), paras. 16-20.

[87] See Klick Report, (Exhibit UKR-5), pp. 3-6.

[88] See Klick Report, (Exhibit UKR-5), pp. 23-32; and Klick Rebuttal Report, (Exhibit HND-118), paras. 19-38.

[89] See Chipty Rebuttal Report, (Exhibit AUS-535) (SCI), paras. 54-71. In its third-party submission, New Zealand similarly argues that Professor Klick's study sets up a false and virtually impossible evaluation test by looking for a marked short-term reduction in population smoking prevalence. New Zealand also submits

81. Dr Chipty submits that Professor Klick's data show that Australia has experienced both a bigger absolute decline and faster rate of decline in daily smoking incidence than New Zealand, although there is no measurable decline in overall smoking, which is consistent with the possibility that the TPP measures are having its intended effect.[90]

### 3.1.3 New South Wales Population Health Survey

82. The New South Wales Population Health Survey (NSWPHS) reports yearly observation of daily smoking prevalence and smoking in general prevalence among men and women in NSW, the most populated state in Australia.

83. Professor Klick estimates a model of (daily or overall) smoking prevalence in NSW controlling for the TPP measures and a linear time trend. The econometric analysis covers the period from 2002 to 2014. Because the NSWPHS data are only available on an annual basis, the TPP variable is coded in different ways for the year 2012: (i) using the value of 0.25 (representing three months out of the year if an October start date is assumed); (ii) using the value of 0.08 (representing a single month if a December start date is assumed), or (iii) using the value of zero (if it is assumed that the vast majority of survey respondents would have answered the survey pre-TPP).

84. Overall, Professor Klick concludes that the TPP measures either had no statistically significant effect on daily smoking prevalence in NSW or had a positive and statistically significant impact on smoking prevalence in general in NSW.[91] An analysis across age groups and/or by gender yields relatively similar results.

85. Dr Chipty considers that Professor Klick's analysis of smoking prevalence in NSW should be disregarded in its entirety because of three basic problems. First, Dr Chipty argues that Professor Klick is incorrect to assert that Australia-wide data are not granular enough to credibly examine the TPP measures' effects on smoking prevalence. Second, Dr Chipty is of the view that the NSWPHS data are incapable of providing a basis to study the TPP measures' effects in NSW, because the estimate of the TPP effect is calculated from two to three data points and cannot be distinguished from the two excise tax increases in 2013 and 2014. Finally, Dr Chipty asserts that there have been changes in the NSW survey methodology that led to a greater number of younger people and males in the survey sample; both of these groups have relatively higher smoking rates, leading to a higher overall reported rate of current smoking.[92]

### 3.1.4 Cancer Institute New South Wales Tobacco Tracking Survey

86. Professor Klick estimates a logit model of the likelihood that the respondent is a smoker, an ordered logit model of the smoking status, and a negative binomial model of the number of adults who are smokers in the household controlling for the TPP measures, annual time trend, gender, individual age fixed effects, week of survey fixed effects and location fixed effects. The econometric analysis covers the period from January 2009 to December 2014.

87. Overall, Professor Klick concludes that the TPP measures either had no statistically significant effect on the likelihood of being a smoker in NSW and on the number of adult smokers in household or had a positive and statistically significant impact on the likelihood of being a daily smoker in NSW and on the smoking status.[93]

88. Australia posits that although Honduras has presented analyses based on a range of other datasets, such as the CITTS data, these are not sources of smoking prevalence data.[94] Australia's

that the use of New Zealand as a comparison case is superficial and misleading, because Professor Klick does not separately analyse the declining trends in smoking prevalence in each country by taking into account the range and variety of tobacco control interventions taken in each country, including the 10% excise tax increase implemented in New Zealand in January 2013, which was unique to New Zealand during the relevant time period.

[90] See Chipty Report, (Exhibit AUS-17), paras. 25-31, Appendix C.
[91] See Klick Supplemental Rebuttal Report, (Exhibit HND-122), paras. 33-43.
[92] See Chipty Surrebuttal Report, (Exhibit AUS-586), paras. 78-86.
[93] See Klick Second Supplemental Rebuttal Report, (Exhibit HND-165), paras. 15-24.
[94] See Australia's response to Panel question No. 196, para. 252.

expert, Professor Chaloupka, is further of the view that using these data to assess the TPP measures' impact on the likelihood of being a smoker, as Professor Klick does, is inappropriate. Professor Chaloupka further argues that Professor Klick's conclusion of no evidence of a decline in smoking associated with the TPP measures is highly misleading, because the nature of the NSW Tracking Survey data does not allow one to use these data to assess the TPP measures' impact on adult smoking prevalence given that the sample is not a representative sample of the NSW adult population. In addition, Professor Chaloupka contends that Professor Klick failed to appropriately account for the changes in the CITTS methodology that led to a greater number of younger people and males in the survey sample; both of these groups have relatively higher smoking rates, leading almost certainly to biased estimates towards showing an increase in smoking following the change in method.[95]

### 3.1.5 National Tobacco Plain Packaging Tracking Survey

89. As described in Appendix A, Australia funded the National Tobacco Plain Packaging Tracking Survey (NTPPTS), a nationwide cross-sectional baseline tracking survey conducted by CCV, to track the effects of the TPP measures. The results from the NTPPTS were published in April 2015 in the supplement to the *Tobacco Control* journal dedicated to investigating the effects of Australia's implementation of the TPP measures.

#### 3.1.5.1 Ajzen et al.'s reports

90. Ajzen et al. re-analysed part of Scollo et al. 2015a[96], applying the approach in Wakefield et al. 2015.[97] Ajzen et al. estimated logistic, ordered logistic and linear regressions of the proportion of daily or weekly smokers, while controlling for age group, gender, education group, socio-economic status group, potential exposure to televised anti-smoking advertising campaigns in the past three months and cigarette costliness.[98]

91. Overall, Ajzen et al. conclude that the TPP measures had no statistically significant impact on the proportion of daily or weekly smokers.[99] A similar finding is found with quarterly data.[100] Ajzen et al. argue that, unlike Professor Chaloupka's assertion that the NTPPTS dataset has less power for detecting statistically significant changes in the more distal outcomes, such as actual tobacco use behaviour, the statistical power of the NTPPTS data to detect small changes in proximal and distal outcomes is very similar.[101]

92. Australia's expert, Professor Chaloupka, argues that Ajzen et al.'s analyses of the NTPPTS data fail to recognize that the TPP measures' impact should be smaller for the less proximal outcomes when one looks at the impact in the overall sample of smokers and recent quitters,

[95] See Chaloupka Second Rebuttal Report, (Exhibit AUS-590), paras. 26-27.

[96] Ajzen et al. argue that Scollo et al. 2015a's conclusion that the TPP measures had no significant effect on the average number of cigarettes consumed per day underreports additional results that were not statistically significant. According to Ajzen et al., Scollo et al. 2015a reported on only one of the seven measures on smoking behaviours and the TPP measures' impact on each of the six unreported variables, dealing with the percentage of daily or weekly smokers, quitting, and relapse was not statistically significant. See Ajzen et al. Data Report, (Exhibit DOM/IDN-2), paras. 225-227. We note, however, that Ajzen et al. do not reassess the analysis by Scollo et al. 2015a of the TPP measures' impact and tax increase on type of tobacco products and price.

[97] Wakefield et al. 2015, (Exhibits AUS-206, DOM-306).

[98] We note that Ajzen et al. also re-estimate their models by replacing the TPP measures dummies with a monthly time trend.

[99] Ajzen et al. reach the same conclusion with respect to measures related to quitting and relapse, which were unreported by Scollo et al. 2015a. In particular, on quitting, Ajzen et al. conclude that there was no statistical change in the proportion of adult smokers that quit for more than one month or had successfully quit between baseline and follow-up. On relapse, Ajzen et al. conclude that there was no statistical change in the proportion of adult ex-smokers that relapsed, still abstained from smoking at follow-up or had stayed quit for more than one week at follow-up. See Ajzen et al. Data Report, (Exhibit DOM/IDN-2), paras. 221-224, Appendix A, pp. 95-97.

[100] See Ajzen et al. Data Rebuttal Report, (Exhibit DOM/IDN-6), para. 90.

[101] See Ajzen et al. Data Rebuttal Report, (Exhibit DOM/IDN-6), paras. 50-54, Appendix II.

given that smokers and recent quitters for whom the most proximal outcomes were unaffected would not be expected to show any change in the less proximal outcomes.[102]

#### 3.1.5.2 Professor Klick's report

93. Professor Klick estimates a logistic regression of the likelihood the respondent is a daily smoker or a smoker in general or an ordered logit model of the smoking status, controlling for the TPP measures, gender, age, education, socio-economic status and a linear time trend. The econometric analysis covers the period from 9 April 2012 to 4 May 2014.[103]

94. Overall, Professor Klick concludes that the TPP measures had no statistically significant effect on the likelihood of being a smoker and on the smoking status during their first year of implementation.[104]

95. Australia posits that the NTPPTS data are not a source of smoking prevalence data.[105] Although Australia's experts do not specifically address Professor Klick's results on smoking status based on the NTPPTS data, Professor Chaloupka generally argues that the NTPPTS is particularly useful in assessing the TPP measures' impact on proximal outcomes (appeal, noticeability of health warnings and misleading) but the largely cross-sectional nature of the survey does not allow one to track the TPP measures' effects on more distal outcomes (e.g. interest in quitting) and on tobacco use behaviour (prevalence and consumption).[106] Professor Chaloupka further claims that the impact of the TPP measures on more distal outcomes should be smaller when the analysis is based on the overall sample of smokers and recent quitters, given that smokers and recent quitters for whom the most proximal outcomes were unaffected by the TPP measures would not be expected to show any change in more distal outcomes.

### 3.2 Analysis by the Panel

96. As discussed above, there is some evidence of an acceleration of the reduction in smoking prevalence since the entry into force of the TPP measures. The question before us at this stage of our analysis is whether this acceleration may, in part or in whole, be attributed to the TPP measures.

97. We note that the evidence relied on by the parties in this part of the discussion is based on an econometric analysis of the evolution of smoking prevalence or smoking incidence aimed at distinguishing and assessing the impact of the TPP measures and other determinants on the level of smoking prevalence or smoking incidence. In particular, the "dependent" variable, smoking prevalence or smoking incidence, is modelled as a function of a number of "explanatory" variables, including the TPP measures.[107] The parties use different econometric estimators (namely the ordinary least square (OLS) estimator, the maximum likelihood estimator, the linear probability

[102] See Chaloupka Rebuttal Report, (Exhibit AUS-582), para. 4.

[103] See Klick Supplemental Rebuttal Report, (Exhibit HND-122), paras. 15-32.

[104] See Klick Supplemental Rebuttal Report, (Exhibit HND-122), paras. 21-32.

[105] See Australia's response to Panel question No. 196, para. 252; and Chaloupka Second Rebuttal Report, (Exhibit AUS-590), para. 26.

[106] See Chaloupka Third Rebuttal Report, (Exhibit AUS-604), paras. 8-13.

[107] Explanatory variables ideally represent the full set of factors that have an impact on the dependent variable, and therefore contribute in "explaining" the behaviour of the dependent variable. In general, explanatory variables are assumed to be independent with respect to the dependent variables. In other words, the dependent variable is assumed to have no (direct or indirect) impact on the explanatory variables that, in turn, have an impact on the dependent variable. For this reason, explanatory variables are often referred to as independent variables. A specific parameter is attached to each explanatory variable included in the econometric model, which represents the impact that the associated explanatory variable might have on the dependent variable. Thus, when the econometric model is well specified with all the relevant explanatory variables, each parameter isolates the impact of the associated explanatory variable on the dependent variable. In addition to the explanatory variables, the econometric model includes an error term, also known as the "residual" term, to capture the facts that no matter how well the model is specified (i) it is often impossible to account for every factor that has an impact on the dependent variable, (ii) the actual relationship between the dependent variable and (some of) the explanatory variables may not be necessarily linear, (iii) data may suffer from measurement errors, and (iv) unpredicted – stochastic – effects can affect the dependent variable. Ultimately, econometric analysis consists of estimating each parameter of the explanatory variables specified in the model.

estimator, and the logistic estimator) to estimate the parameters of the explanatory variables that best fit the relevant data. Each estimated parameter is assigned a standard error, which enables the evaluation of whether this estimated value of the parameter is statistically different from zero, i.e. statistically significant, at a given level of significance typically 1%, 5% or 10%. The standard error provides information on the degree of confidence and reliability of the estimated value of each parameter considered in the model. As described above, the Dominican Republic, Honduras and Indonesia submit that the TPP measures' impact on smoking prevalence or smoking incidence is statistically not different from zero, while Australia submits that the impact of the TPP measures on smoking prevalence is negative and statistically different from zero.

98. At the outset, we note that we approach this assessment on the basis that our task is not to conduct our own econometric assessment of the impact of the TPP measures on smoking prevalence, but rather to review the robustness of the econometric evidence submitted by the parties in this respect.

99. While we acknowledge that no data are perfect, we agree with Australia that the RMSS data is the most suited available data submitted by the parties to analyse the impact of the TPP measures on smoking prevalence, for two main reasons. First, the RMSS data provide an actual measure of smoking prevalence (based on a population of smokers, recent quitters and non-smokers). Second, the data are available monthly for a long period of time before and after the introduction of the TPP measures. The parties disagree with respect to the selection sample period. We concur with Australia that a larger number of observations is likely to increase the precision of the estimates. In addition, we note that Professor List, in his report submitted by the Dominican Republic and Indonesia, suggests limiting the sample period to analyse smoking prevalence, but does not propose the same restriction in the analysis of cigarette consumption.[108]

100. The other data sources considered by Professor Klick suffer, in our view, from a number of drawbacks in comparison with the RMSS data. In particular, the commissioned Roy Morgan Research Survey data collected in Australia and New Zealand cover a short period prior to the introduction of the TPP measures in December 2012, during which plain packs were already authorized for sale in the market. In comparison, the pre-TPP period of the RMSS data are available starting January 2001. In addition, the commissioned Roy Morgan Research Survey data, just like the CITTS and NTPPTS datasets, do not actually measure smoking prevalence, because the sample is based only on smokers and recent quitters. Finally, although the NSWPHS data provide information on smoking prevalence in NSW, the fact that they are only available yearly imply that the post-TPP period used to assess the impact of TPP measures covers only two to three observations. This is extremely short in comparison with the RMSS data encompassing up to 34 post-TPP observations (December 2012-September 2015).

101. Turning to the econometric results based on the RMSS data, we note at the outset that the different conclusions reached by the parties regarding the impact of the TPP measures on smoking prevalence stem from the fact that the parties' experts use different model specification (i.e. different explanatory variables included in the model), estimation approaches and in some cases sample periods. Even among the experts of the Dominican Republic, Honduras and Indonesia, different model specifications are used.[109]

102. On a number of occasions, the rebuttal reports of the Dominican Republic's, Honduras's and Indonesia' experts proposed new model specifications or methodologies or both, to address some of Australia's criticisms but also to take into account issues that their experts themselves highlighted. For instance, IPE initially proposed (in its analysis of cigarette sales volumes) controlling for excise tax increases by including indicator variables for each excise tax increase in

[108] See List Third Supplemental Report, (Exhibit DOM/IDN-7), pp. 4-8.

[109] For instance, IPE initially (see IPE Updated Report, (Exhibit DOM-303)) and Professor Klick (see Klick Third Supplemental Rebuttal Report, (Exhibit HND-166)) control for excise tax increases by specifying dummy variables for each excise tax hike implemented during the sample period, while IPE later in the proceedings (see, e.g. IPE Second Updated Report, (Exhibit DOM-361)) and Professor List (see, e.g. List Report, (Exhibit DOM/IDN-1)) consider the variable of cigarette price to be a better proxy for excise tax increases. Similarly, Professor Klick applies the instrument variables estimator to address the potential endogeneity of the price variable, while IPE and Professor List do not. Another difference among the experts of the Dominican Republic, Honduras, and Indonesia is the fact that IPE and Professor Klick use monthly sample sizes as weights in the estimation, while Professor List does not, except in the first stage of his micro-econometric analysis.

the sample period.[110] However, it subsequently contended that a more appropriate measure to capture the excise tax increases was the average price per cigarette in Australia.[111] Although the experts of the Dominican Republic, Honduras and Indonesia ultimately reach the same conclusion when the changes they proposed are taken into account, they explicitly or in some cases implicitly suggest ignoring or giving less weight to their previous results affected by the issues their experts themselves recognized or highlighted. As explained above, it is not our task to present a unified econometric analysis but rather assess the robustness of each report.

### 3.2.1 IPE's and Professor List's econometric results

103. After a careful review of the econometric reports on smoking prevalence based on the RMSS data submitted by the Dominican Republic's and Indonesia's experts[112], we are not persuaded that these econometric results can be taken at face value, mainly because most of their model specifications are unable to detect the impact of tobacco costliness (including excise tax increases) on smoking prevalence. Yet, all parties consider tobacco excise tax to be one of the most effective tobacco control policies.[113] To some extent, the Dominican Republic, Honduras and Indonesia are asking the Panel to conclude that the TPP measures had no impact on smoking prevalence, because its effect is statistically not significant, but to disregard the fact that the same econometric results suggest that excise tax or price increase have also had no impact on smoking prevalence.

104. The manner in which the smoking prevalence trend is modelled with respect to the sample period considered (i.e. January 2001-September 2015 or July 2006-September 2015) has an important consequence on whether the econometric analysis is able to identify the impact of other variables. These variables can contribute, along with demographic shifts and other factors unrelated to tobacco control policies, to creating the smoking prevalence trend.[114] This problem is defined as overfitting. For instance, the issue of overfitting associated with the trend variable is so severe in the ARIMAX models reported in the IPE Reports that even the lagged dependent variable is not statistically significant, suggesting that the level of smoking prevalence does not depend on the level of smoking prevalence in the previous month, which is in complete contradiction with the fact that smoking prevalence follows a downward trend, as agreed by all parties.[115] Similarly, the results of Professor List's two-stage micro-econometric shows how the inclusion of the secular (long-term) trend[116] captures most of the explaining power making the price variable no longer significant in most of the specifications, while the price variable is always statistically significant when the trend variable is not included.[117]

105. In our view, it is important that the trend variable specified in the model avoids overfitting the data, to allow an identification of the impact of other variables of interest, such as individual

[110] See IPE Report, (Exhibit DOM-100), pp. 67-69.

[111] See IPE Second Updated Report, (Exhibit DOM-361).

[112] Although the methodologies proposed by IPE and Professor List differ, they share a number of similarities, including the sample period, the choice of the variable to control for tobacco price control policies and the procedure to estimate the standard errors. That is why both experts' results are discussed together. See List Report, (Exhibit DOM/IDN-1); List Rebuttal Report, (Exhibit DOM/IDN-3); List Second Supplemental Report, (Exhibit DOM/IDN-5); List Third Supplemental Report, (Exhibit DOM/IDN-7); List Summary Report, (Exhibit DOM/IDN-9); IPE Report, (Exhibit DOM-100); IPE Updated Report, (Exhibit DOM-303); IPE Second Updated Report, (Exhibit DOM-361); IPE Third Updated Report, (Exhibit DOM-375); and IPE Summary Report, (Exhibit DOM-379).

[113] See Australia's first written submission, para. 719; Honduras's first written submission, para. 589; Dominican Republic's first written submission, paras. 758 and 1027; Cuba's first written submission, para. 276; and Indonesia's first written submission, para. 63.

[114] IPE considers a linear trend for the sample period July 2006-September 2015 and a quadratic trend for the sample period January 2001-September 2015. Similarly, Professor List specifies a linear trend for the sample period July 2006-September 2015. Professor Kick specifies a quadratic trend for the sample period January 2001-September 2015.

[115] See, e.g. IPE Third Updated Report, (Exhibit DOM-375), pp. 34 and 106-108.

[116] See para. 11 above for a discussion on the secular trend.

[117] See, e.g. List Rebuttal Report, (Exhibit DOM/IDN-3), pp. 15 and 24; List Second Supplemental Report, (Exhibit DOM/IDN-5), pp. 32-33; and List Third Supplemental Report, (Exhibit DOM/IDN-7), pp. 18-19 and 29-30. In addition, we also question the validity of some of the results obtained by Professor List in the first stage, where none of the explanatory variables are statistically significant (at 10%), not even the constant. Such results suggest that the associated model specification might be misspecified or affected by another econometric issue that Professor List failed to address.

tobacco control policies. Otherwise, one cannot rule out the possibility that the smoking prevalence trend included in the model accounts not only for the trend itself but potentially also reflects any tobacco control policies that contributed to its trend. We note that while the experts of the Dominican Republic, Honduras and Indonesia discussed extensively the importance of accounting properly for the secular downward trend in smoking prevalence, they do not address the fact that in the vast majority of their results, the price variable was not statistically significant.

106. In this context, we also consider it important to specify the tobacco price control policy in the most appropriate manner. We note that the Dominican Republic's experts' view on this issue has evolved throughout the proceedings. IPE was the first party to propose controlling for tobacco tax excise increases with indicator variables (in its analysis of cigarette sales discussed in Appendix D), but later changed its view when Professor List used a price variable in his own analysis.[118] IPE referred to the tobacco price variable as a measure for costliness of tobacco products (reflecting also the effect of tax increases). IPE also proposed the level of tax as an alternative to the price variable. In our view, the three types of variables (dummy variables, tax level variable and price variable) are in theory complementary, each with advantages and disadvantages. The dummy variables are, by construction, exogenous and specific to each excise tax increase but do not specify the actual level of the tax increase. The tax level variable is also, by definition, exogenous and accounts for the actual level of the tax increase. However, as explained by Australia, it relies on the assumption that the effect of the tax increase on prevalence is proportional to the size of the tax increase. The price variable is a broader variable and accounts implicitly for all the factors that affect tobacco price, including the excise tax increases but not only that. The TPP measures can also affect the price variable, as pointed by Dr Chipty and addressed by Professor Klick.

107. In addition, we observe after a careful review, that there is, as shown in **Figure C.20**, evidence of multicollinearity between the price variable and the linear trend variable, in particular when the sample period is restricted to July 2006 to September 2015.[119] Multicollinearity arises when two (or more) explanatory variables convey the same information. In the presence of multicollinearity, the predictive power of the model remains unchanged, but the confidence interval of the coefficient estimates may increase. Moreover, the coefficient estimates may become very sensitive to minor changes in the model specification or data. One way to mitigate multicollinearity is to increase the sample period. We note, however, that including a second linear trend specific to the July 2006-September 2015 period, as suggested by IPE, would not resolve this issue. We also note that unlike Professor Klick, IPE and Professor List does not address the fact that the TPP measures might affect the price variable. IPE and Professor List's model specifications are unable to distinguish between the impact specific to the price variable and the TPP measures. Overall, given that neither IPE nor Professor List address the issue of multicollinearity, and the potential impact of the TPP measures on prices, we call into question the econometric results based on the price variable. We also note that the expert reports submitted by the Dominican Republic, Honduras and Indonesia (and Australia) failed to mention that standard unit root tests suggest that the tax level and the price variables are not stationary.[120] Yet, econometric theory recommends not estimating a model when the dependent variable (i.e. smoking prevalence) is stationary and one of the explanatory variables (i.e. tax level or price) is not stationary in order to avoid spurious and biased results.

108. The parties' experts also disagree with the manner in which the population sampling correction is addressed in the RMSS data. We first note, as pointed out by Australia, that Professor List did not consider the sample reweighting in the RMSS data to be an issue in his first two reports. Similarly, IPE did not address the issue of reweighting in its first three reports.[121] We recognize the importance of attempting to control for sample re-weighting events in the RMSS data. We note, however, that the inclusion of the three indicator variables to control for the reweighting correction in April 2009, July 2010, and April 2014, as suggested by Professor List, increases the issue of multicollinearity, in particular when the price (or tax level) and trend

[118] See IPE Updated Report, (Exhibit DOM-303), para. 150; and IPE Second Updated Report, (Exhibit DOM-361), para. 29.

[119] Evidence of multicollinearity is confirmed by the variance inflation factors statistic.

[120] A variable is said to be stationary, when its statistical properties, such as mean and variance are all constant over time.

[121] We note that, unlike Professor List and IPE, Professor Klick does not address the sample reweighting corrections in the RMSS data in his analysis.

variables are included in the specification.[122] This problem is accentuated when a fully flexible reweighting correction is adopted. For instance, none of the explanatory variables is statistically significant at 5% when the linear trend and price variables and the fully flexible reweighting correction are included in Professor List's model specification for smoking prevalence among minor and young adult.[123] Similar findings apply to IPE's modified trend analysis of overall smoking prevalence, where the only significant variable is the dummy for the trend shift in July 2006.[124] Some results of IPE's modified trend analysis even suggest that the TPP measures have led to a statistically significant increase in cigar smoking prevalence.[125] The idea that the TPP measures "backfired" is rejected not only by Australia, but also by the Dominican Republic's and Indonesia's experts. Professor List has explicitly questioned the possibility that the TPP measures "backfired".[126] IPE explains also that it does not interpret the statistically significant and positive impact of the TPP measures on cigar smoking prevalence as evidence that the TPP measures led to an increase in cigar smoking prevalence, but rather as strong evidence to reject the claim of the intended negative TPP measures' effect on cigar smoking prevalence.[127] Yet, the Dominican Republic's and Indonesia's experts do not explain why such finding should be interpreted differently, without questioning the validity of the model specification that yields such result, especially when it relates to the main variable of interest of the econometric analysis. In fact, none of the Dominican Republic's and Indonesia's experts sought to explain why the TPP measures would lead to an increase in the number of smokers. Overall, and based on the above, we have doubts about the reliability of the results obtained when the price variable, time trend and sample reweighting dummies are included in the model specifications.

**Figure C.20: Tobacco Price and Linear Trend**

[[***]]

Note: Weighted average price per stick (in AUD) of factory-made cigarettes and roll-your-own tobacco (based on a 0.8 grams conversion rate of roll-your-own tobacco to cigarette sticks). The vertical dashed line indicates the introduction of TPP and enlarged GHWs. The dashed and dotted linear time trends correspond, respectively, to the sequence 1, 2, 3 … up to 177 or the sequence 1, 2, 3, up to 111.

Source: In-Market-Sales data and Coles and Woolworths data, based on data update and computer code and calculations supporting the IPE Updated Report, (Exhibit DOM-373) (SCI); and List Second Supplemental Report, (Exhibit DOM/IDN-5).

109. Another technical issue that has been the object of extensive discussions among the parties is the choice of the procedure used to compute the standard error of each of the estimated coefficients associated with the variables specified in the econometric model.[128] As explained earlier, statistical significance is essential because, as well as being the variable's estimated coefficient, it is also important to determine whether the coefficient's variable is statistically

[122] For instance, the results reported in columns (3) and (4) of Table 8 in List Second Supplemental Report, (Exhibit DOM/IDN-5), p. 43 suggest that the TPP measures' effect on smoking prevalence of minors and young adults is positive and statistically significant at 10%. In other more recent results on smoking prevalence among minor and young adult, presented in column 3 of Table 13 in List Third Supplemental Report, (Exhibit DOM/IDN-7), p. 32, the only variable, besides the constant, that is statistically significant is the trend when the price and time trend variables as well as reweighting correction dummies are included. We note, however, that the issue of multicollinearity is less severe when excise tax increase dummies or the real tax level variable are used.

[123] For instance, in the results on overall smoking prevalence presented in column 3 of Table 14 in List Third Supplemental Report, (Exhibit DOM/IDN-7), p. 33, the only variable, besides the constant, that is statistically significant is the trend when the price and time trend variables as well as fully flexible reweighting correction dummies are included.

[124] Similar findings also apply to IPE's micro-econometric analysis, where the only significant variable is the dummy for the trend shift in July 2006, besides the constant and the socio-demographic variables. See IPE Third Updated Report, (Exhibit DOM-375), p. 38. We also note that IPE decided not to update the results of the ARIMAX model with the reweighting correction dummies.

[125] See the IPE Third Updated Report, (Exhibit DOM-375), pp. 48-50.

[126] See List Report, (Exhibit DOM/IDN-1), para. 106.

[127] See IPE Summary Report, (Exhibit DOM-379), fn 124.

[128] Heteroscedasticity and autocorrelation consistent standard errors are computed for the following models on smoking prevalence: IPE's modified trend analysis and ARIMAX models, and Professor List two-stage micro-econometric analysis. Similarly, robust standard errors are computed for the following models on cigarette consumption: IPE's modified trend analysis and ARIMAX models, and Professor List's event study analysis.

different from zero. Yet, a standard error is required to determine the estimated coefficient's level of statistical significance. We therefore recognize, like the parties, the importance of computing standard errors that are robust to heteroscedasticity and autocorrelation.

110. We note, however, that the treatment of standard errors evolved over the course of the proceedings. Initially, Professor List and IPE chose to apply the STATA software command ivreg2 to calculate standard errors that are robust to heteroscedasticity and serial correlation using the automatic bandwidth selection procedure by Newey and West (1994).[129] Subsequently, Professor List, and later on IPE, applied an alternative way of calculating standard errors, that, according to Professor List, is adjusted to reflect more accurately the original suggestion by Newey and West (1994).[130] Technically speaking, the disagreement between Professor List (as well as IPE and Professor Klick) and Dr Chipty concerns the procedure to correct for autocorrelation, in particular the choice of the maximum amount of time, defined as the maximum lag, that the data can be correlated over time. Professor List proposes to set a smaller parameter value, resulting also in a smaller maximum lag than the one specified in the ivreg2 command. A careful review of the evidence and discussions shows that the choice of the maximum lag is not well established in the statistics and econometric literature, as pointed out by in an email exchange with STATA developers.[131] As a result, it is unclear whether the results associated with Professor List's procedure would have changed for a range of parameter values, taking into consideration the fact that the maximum lag should be able to take into account all lags until the serial correlation in the data vanishes.[132]

111. In sum, and based on the elements discussed above, we have reservations regarding IPE and Professor List's methodologies[133] and therefore question their results, based on these methodologies, that suggest that the TPP measures had no statistically significant impact on smoking prevalence.[134]

---

[129] See IPE Report (Exhibit DOM-100); IPE Updated Report, (Exhibit DOM-303); List Report (Exhibit DOM/IDN-1); and List Rebuttal Report, (Exhibit DOM/IDN-3).

[130] See Second IPE updated report, (Exhibit DOM-361); IPE Third Updated Report, (Exhibit DOM-375); and IPE Summary Report, (Exhibit DOM-379); List Second Supplemental Report, (Exhibit DOM/IDN-5); List Third Supplemental Report, (Exhibit DOM/IDN-7); and List Summary Report, (Exhibit DOM/IDN-9).

[131] See the correspondence between the parties and STATA developers in the IPE Summary Report, (Exhibit DOM-379), p. 70; and Chipty Second Rebuttal Report, (Exhibit AUS-591), pp. B1-B2. In the email exchange, the STATA software developers also explain that the automatic choice of the maximum lag in the command ivreg2 is in line with the criteria necessary for asymptotic optimality.

[132] Professor List presents the results of simulations to compare the frequency of so-called false positives using Professor List's procedure and the ivreg2's automatic selection procedure. Professor List concludes that the STATA ivreg2's automatic selection procedure leads to a wrong finding of a statistically significant result 16% of the time, instead of 5% (see List Summary Report, (Exhibit DOM/IDN-9), paras. 95-98). We note that the results are based on the sample size of 111 observations (July 2006-September 2015), while Dr Chipty considers a larger sample period of 177 observations (January 2001-September 2015). It is therefore unclear to what extent Professor List's results would change if the sample size increases, taking into account the fact that according to the STATA developers the formulae used in ivreg2 meet the criteria necessary for asymptotic optimality.

[133] We note that Gibson Report (Exhibit DOM-92) presents the main results of an econometric analysis of smoking prevalence of adolescents based on the RMSS data. The results suggest that the TPP measures and enlarged GHWs had no statistically significant impact on the smoking prevalence of 14-17 year-old smokers of manufacturing cigarettes, RYO cigarettes, pipes and cigars, respectively. We question however the validity of these results for some of the same reasons that apply to the econometric analysis of adult smoking prevalence reported by IPE, namely the use of the price variable and (quadratic) time trend.

[134] Professor List proposes a framework to synthesize the diverse analyses, submitted by all the parties, on the TPP measures' impact on smoking prevalence and cigarette consumption (see for instance List Second Supplemental Report, (Exhibit DOM/IDN-5), paras. 30-71). According to Professor List, the probability that the TPP measures are working as intended is nearly zero whether just the studies of the experts of the Dominican Republic, Honduras and Indonesia or the universe of the work of all the experts are considered. List Summary Report, (Exhibit DOM/IDN-9), paras. 167-173. Dr Chipty argues that Professor List's framework for reweighing the evidence is both misleading and misguided because it amounts to nothing but a counting exercise that does not account for the interdependence of studies or the biases they may contain. Chipty Third Rebuttal Report, (Exhibit AUS-605), paras. 6-8. We acknowledge the usefulness of having a single number (Post-Study Probability) determining the likelihood that the TPP measures have caused a reduction in smoking prevalence based on all available econometric studies. We note, however, that Professor List assigns the same weight to all the results prepared by the experts of the Dominican Republic, Honduras and Indonesia, even when Professor List, himself, questions the validity of some of these results. For instance, Professor List is not

### 3.2.2 Professor Klick's econometric results

112. Similarly, a review of the econometric results based on the RMSS data reported by Professor Klick leads us to question their robustness.[135] In particular, as demonstrated by Dr Chipty, the use of a quadratic trend to capture the downward trend in smoking prevalence leads the predicted tobacco price variable to be not significant. As explained above, specifying an excessively flexible smoking prevalence trend (i.e. quadratic trend) is likely to overfit the data on smoking prevalence and make redundant any other variables, such as individual tobacco control policies, that can potentially have also an impact on smoking prevalence. Finally, it is also unclear how the standard error and confidence interval of the total effect of the TPP measures, composed of the direct estimated impact of the TPP measures on smoking prevalence obtained in the second stage of the procedure and the indirect estimated impact of the TPP measures on cigarette price obtained in the first stage of the procedure through the impact of cigarette price on smoking prevalence, were calculated.

113. Our review of Professor Klick's econometric analyses of the TPP measures' impact on smoking prevalence and incidence based on the other datasets leads us also to question his results. With the exception of the NSWPHS data, the other datasets used smoking incidence, instead of smoking prevalence, by focusing only on smokers and recent quitters. As a result, data on smoking incidence is, by definition, unable to measure the impact of TPP measures on non-smokers. Therefore, the studies on smoking incidence are in our view less relevant than the studies focusing on smoking prevalence.

114. We question the validity of Professor Klick's difference-in-difference analysis on smoking incidence based on the commissioned Roy Morgan Research Survey, because Professor Klick is unable to accurately identify the respondents, who had noticed plain packs in the pre-period given that the question about noticing changes in the packaging was not asked of all respondents. As a result, the pre-period is, in our view, not valid. In addition, when Professor Klick attempts to address some of Dr Chipty's critics regarding the pre-period, several results find a positive and statistically significant TPP measures' effect on the likelihood an individual reports being a smoker. Professor Klick explains that he does not view these results as suggesting that smoking has surely increased under the TPP measures, because of the "multiple comparison problem", which implies that when "very many outcomes" are examined, there is a relatively high likelihood that one will find statistically significant results even by mere chance.[136] Yet, Professor Klick does not explain how ten different alternative specifications should be viewed as "very many outcomes".[137] Furthermore, Professor Klick fails to account for the excise tax increase that took place in New Zealand between Waves 1 and 2 and Waves 5 and 6 of the commissioned survey, making any inference about the reduction in smoking incidence in Australia compared to New Zealand questionable.

115. We are also not persuaded by Professor Klick's results based on an analysis of the NSWPHS data on the ground that the nature of the data (i.e. yearly observations) limits the number of observations to two post-packaging observations (2013 and 2014), which prevents distinguishing between the TPP measures and tobacco excise tax increases in 2013 and 2014.

116. We question also Professor Klick's results of the CITTS data, which, as explained above, do not analyse smoking prevalence, but smoking status. Unlike Dunlop et al. 2014, who also analyse the CITTS data but for a shorter period[138], Professor Klick does not explicitly account for other tobacco control policies (besides the TPP measures), such as the respondents' level of exposure to anti-smoking advertising prior to their interview and changes in cigarette costliness. Professor Klick includes an annual time trend, but the variable is actually never statistically significant. This result is at odds with the view shared by all the experts of the Dominican Republic,

persuaded that the TPP measures "backfired" as suggested by Professor Klick's difference-in-difference analysis. We are of the view that a methodology that is unable to make a distinction between robust results and biased results is not informative.

[135] See Klick Third Supplemental Rebuttal Report, (Exhibit HND-166), paras. 11-23.

[136] See Klick Rebuttal Report, (Exhibit HND-118), fn 24.

[137] The ten specifications for daily smoker are reported in Tables 3 to 10 in Klick Rebuttal Report, (Exhibit HND-118). The ten specifications for overall smokers are reported in Tables 3 to 9 and Table 11 in Klick Rebuttal Report, (Exhibit HND-118).

[138] Dunlop et al. 2014, (Exhibits AUS-207, HND-132, DOM-199).

Honduras and Indonesia, including Professor Klick, that smoking prevalence is characterized by a downward trend. Similarly, several results suggest that the TPP measures "backfired" and led to an increase in the likelihood of a respondent reporting smoking daily and in general. It is unclear what such results can be attributed to, and for that reason we cannot consider these results as relevant.

117. We also question the validity of Professor Klick's results on smoking incidence and self-reported frequency of smoking based on the NTPPTS data. Professor Klick chose not to control for the exposure to tobacco-related media activity and tobacco costliness, on the grounds that the results do not change and both omitted variables are likely to be endogenous.[139] We note that, unlike in his analysis of the RMSS data, Professor Klick did not attempt to address the potential endogeneity of cigarette costliness. Yet, failing to account for other tobacco control policies, in particular the December 2013 excise tax hike, might lead to biased results. We further note that although Professor Klick refers to Durkin et al. 2015 and other articles in *Tobacco Control*, the explanatory variables considered by Professor Klick are different to those in Durkin et al. 2015. For instance, Professor Klick uses a single variable for age, while Durkin et al. 2015 considers a dummy variable for each age group. The same difference applies to the variable education. In addition, unlike the vast majority of the micro-econometric results submitted by the parties, Professor Klick's econometric results suggest that being a male has no statistically significant effect on the probability an individual reported being a (daily or in general) smoker and on the self-reported frequency of smoking. This finding contradicts econometric evidence submitted by the parties, including Professor Klick, according to which men are more likely to smoke (daily) than women.[140] This puzzling result raises questions about the validity of Professor Klick's model specification.[141]

118. In sum, and based on the elements discussed above, we have reservations regarding Professor Klick's methodologies and therefore question his contradictory results suggesting that the TPP measures either had no statistically significant impact on smoking prevalence and/or smoking incidence or a positive and statistically significant impact on smoking prevalence and/or smoking incidence.

### 3.2.3 Ajzen et al.'s econometric results

119. A careful review of Ajzen et al.'s analysis of the proportion of daily or weekly smokers reported in the NTPPTS data leads us to conclude that Ajzen et al.'s finding is in line with Scollo et al. 2015a, who find that daily cigarette consumption did not change during the first year of implementation of the TPP measures.[142] Like the results submitted by Professor Klick, we note, however, that Ajzen et al. surprisingly find that gender has no statistically significant effect on the probability that an individual reported being a (daily or weekly) smoker.

### 3.2.4 Dr Chipty's econometric results

120. Turning to the econometric results analysing smoking prevalence based on the RMSS data submitted by Australia's expert, Dr Chipty, we note that a number of concerns that we raised while reviewing the complainants' approaches and results have been addressed by Dr Chipty. In particular, Dr Chipty acknowledges and addresses the issue of overfitting associated with a too flexible trend. Dr Chipty's model specification also includes the excise tax increases dummy variables and thus avoids the problems of multicollinearity and endogeneity associated with the inclusion of the price variable (in combination with a quadratic trend variable). In addition, the use of the tax dummies avoids the issue of non-stationarity of the price or tax level variables.

---

[139] See Klick Supplemental Rebuttal Report, (Exhibit HND-122), fn 10.

[140] See NHS Results, (Exhibit AUS-622), p. 30; and Klick Supplemental Rebuttal Report, (Exhibit HND-122).

[141] We also note that only three out of the 54 variables are statistically significant in the model for the number of adults who are smokers in household. See Klick Supplemental Rebuttal Report, (Exhibit HND-122), backup material.

[142] We note that the model specification considered by Ajzen et al. is different from the one developed by Professor Klick. For instance, Ajzen et al. include dummies variables for two age groups, while Professor Klick includes a single age variable. Ajzen et al. include also dummies variables for different groups of education and socio-economic status, respectively, while Professor Klick includes a single variable for education and another variable for socio-economic status.

121. A careful review of Dr Chipty's econometric results further shows that the negative and statistically significant impact of the TPP measures on overall smoking prevalence is robust to alternative specifications, including different TPP starting date (October, November and December 2012), the use of an excise tax level variable (instead of the excise tax increase dummies) and sample reweighting dummies.[143] In addition, the impact of the TPP measures on overall smoking prevalence remains negative and statistically significant in most specifications when Professor List's procedure to compute standard errors is implemented.[144]

122. Overall, based on the most recent econometric evidence submitted by Australia, there is econometric evidence suggesting that TPP and enlarged GHWs contributed to the reduction in overall smoking prevalence in Australia.[145] A similar conclusion applies also to cigar smoking prevalence in Australia.

## 4 OVERALL CONCLUSION ON POST-IMPLEMENTATION EVIDENCE ON SMOKING PREVALENCE

123. Overall, based on the most recent data available and econometric evidence submitted by the parties, we find that:

a. There is evidence that overall smoking prevalence in Australia continued to decrease following the introduction of the TPP measures.

b. The downward trend in overall smoking prevalence in Australia appears to have accelerated in the post-TPP period.

c. Although it is impossible to distinguish between the impact of TPP and the impact of enlarged GHWs, there is some econometric evidence suggesting that the TPP measures, together with the enlarged GHWs implemented at the same time, contributed to the reduction in overall smoking prevalence as well as in cigar smoking prevalence observed after their entry into force.

124. We note, however, that no post-implementation empirical evidence has been presented to us on the impact of the TPP measures on cigarillos smoking prevalence.

[143] See Chipty Third Rebuttal Report, (Exhibit AUS-605), p. 22; and Chipty Second Rebuttal Report, (Exhibit AUS-591), p. D2.

[144] See Chipty Second Rebuttal Report, (Exhibit AUS-591), p. 33.

[145] See Chipty Third Rebuttal Report, (Exhibit AUS-605), pp. 22 and 51; and Chipty Second Rebuttal Report, (Exhibit AUS-591), pp. 33, D1 and D2.

## APPENDIX D:

## EVIDENCE RELATING TO TOBACCO PRODUCT SALES AND CONSUMPTION FOLLOWING THE ENTRY INTO FORCE OF THE TPP MEASURES

1. Similar to smoking prevalence, the contribution of the TPP measures to reducing tobacco consumption has been the object of extensive discussion between the parties. A number of expert reports are dedicated in part or in whole to this discussion.[1] Different databases, statistical analysis and econometric methods have been proposed to determine whether TPP and enlarged GHWs have contributed to the reduction in cigarette consumption.

2. Most studies use cigarette sales volumes as a measure of cigarette consumption, with the exception of a few other studies[2], which use reported cigarettes consumption. Cigarette sales volumes are taken from market data, while data on cigarette consumption are reported in various surveys. The underlying assumption is that cigarette consumption (i.e. cigarette demand) can be proxied by cigarette sales (i.e. supply of cigarettes). We note that none of the parties questioned this assumption. Where relevant, this Appendix will distinguish between cigarette consumption and cigarette sales.

3. As in respect of the analysis of smoking prevalence, one of the only points of agreement among the parties is that the empirical econometric evidence on cigarette consumption submitted does not distinguish between the impact of TPP and the impact of enlarged GHWs on cigarette sales or consumption, because both measures were implemented at the exact same time.[3] Unless specified otherwise, the impact of TPP in this Appendix therefore refers to the impact of TPP <u>and</u> enlarged GHWs.

4. The Dominican Republic, Honduras and Indonesia argue that all their experts' empirical statistical and econometric studies show that the TPP measures failed to reduce cigarette sales volumes or consumption.[4] The complainants even suggested initially that TPP measures "backfired" by increasing tobacco sales.[5] This argument was however not developed later in the proceedings.

5. Notwithstanding its contention that it is too early to investigate the impact of TPP on tobacco consumption[6], Australia submitted expert reports estimating the TPP measures' impact on

[1] See Chipty Report, (Exhibit AUS-17); Chipty Supplementary Report, (Exhibit AUS-511); Chipty Rebuttal Report, (Exhibit AUS-535) (SCI); Chipty Surrebuttal Report, (Exhibit AUS-586); Chipty Second Rebuttal Report, (Exhibit AUS-591); Chipty Third Rebuttal Report, (Exhibit AUS-605); List Report, (Exhibit DOM/IDN-1); List Rebuttal Report, (Exhibit DOM/IDN-3); List Second Supplemental Report, (Exhibit DOM/IDN-5); List Third Supplemental Report, (Exhibit DOM/IDN-7); List Summary Report, (Exhibit DOM/IDN-9); IPE Report, (Exhibit DOM-100); IPE Updated Report, (Exhibit DOM-303); IPE Second Updated Report, (Exhibit DOM-361); IPE Third Updated Report, (Exhibit DOM-375); IPE Summary Report, (Exhibit DOM-379); Klick Report, (Exhibit UKR-5); Klick Rebuttal Report, (Exhibit HND-118); Klick Supplemental Rebuttal Report, (Exhibit HND-122); Klick Second Supplemental Rebuttal Report, (Exhibit HND-165); Klick Third Supplemental Rebuttal Report, (Exhibit HND-166); Klick Fourth Supplemental Rebuttal Report, (Exhibit HND-169); Ajzen et al. Data Report, (Exhibit DOM/IDN-2); and Ajzen et al. Second Data Report, (Exhibit DOM/IDN-4).

[2] See Klick Report, (Exhibit UKR-5); Klick Supplemental Rebuttal Report, (Exhibit HND-122); Klick Second Supplemental Rebuttal Report, (Exhibit HND-165), Ajzen et al. Data Report, (Exhibit DOM/IDN-2); and Ajzen et al. Second Data Report, (Exhibit DOM/IDN-4).

[3] See Australia's first written submission, para. 518; Dominican Republic's response to Panel question No. 8, para. 61; Honduras's responses to Panel question No. 8; and Indonesia's response to Panel question No. 8, para. 8.

[4] See Dominican Republic's comments on Australia's response to Panel question No. 196, paras. 681-682; Honduras's second written submission, paras. 56-60; Indonesia's second written submission, para. 192; and Cuba's second written submission, paras. 276-277.

[5] See Indonesia's first written submission, para. 412; Cuba's first written submission, para. 163; Dominican Republic's first written submission, para. 523; and Honduras's first written submission, para. 395.

[6] See Australia's first written submission, para. 670. Australia considers that, in the early stages of introduction of the measures, the most appropriate way to discern their effects was to rely on experiments and surveys which concern drivers of choice, attitudes and, ultimately, the elicitation of behavioural intentions. Australia's first written submission, para. 147.

cigarette sales, in response to the submissions of the Dominican Republic, Honduras and Indonesia. Australia contends that once the most appropriate dataset available (i.e. In-Market-Sales data) are used and the flaws of the econometric models put forward by the experts of the Dominican Republic, Honduras and Indonesia are corrected, the econometric results show that TPP measures have already contributed to their objectives by reducing cigarette sales volumes.[7]

6. Similarly to the discussion on smoking prevalence, we note that the approaches presented by the parties to analyse cigarette sales volumes and consumption evolved over the course of the proceedings. Overall, these address three main aspects, reviewed below in sections 1 to 3:

- First, the parties have submitted economic figures and descriptive statistics analyses aimed at determining whether cigarette sales or consumption have decreased following the implementation of the TPP measures;[8]

- Second, Australia, the Dominican Republic and Indonesia have submitted statistical analyses to determine whether there was a break in the trend in cigarette sales following the implementation of the TPP measures, and in particular, whether the reduction of cigarette volumes sales has accelerated in the post-TPP period;

- Finally, Australia, the Dominican Republic, Honduras and Indonesia have submitted econometric analysis to determine whether the TPP measures have contributed to a reduction of cigarette sales or consumption by isolating and quantifying the different factors that can explain the evolution of cigarette sales or consumption.

7. The parties also presented and discussed a graphical and descriptive analysis of cigar trade data. This is addressed below in Section 4.

## 1 WHETHER CIGARETTE SALES VOLUMES DECREASED FOLLOWING THE IMPLEMENTATION OF THE TPP MEASURES

8. Different market data sources tracking cigarette sales volumes in Australia have been submitted by Australia, the Dominican Republic, Honduras and Indonesia. We first present these datasets, before providing an overall analysis on the basis of these data.

### 1.1 Main datasets and arguments

#### 1.1.1 In-Market-Sales/Exchange of Sales

9. Managed by InfoView Technologies, In-Market-Sales (IMS), also known as Exchange of Sales (EOS), is a dataset comprising monthly sales volume data at the Stock-Keeping Unit-level for factory-made cigarettes (FMC) and fine-cut tobacco covering the period from January 2000 through September 2015. The data comprise monthly sales from manufacturers to wholesalers and retailers, as reported by the three largest tobacco companies in Australia - British American Tobacco Australia (BATA), Philip Morris Limited, and Imperial Tobacco Australia Limited (ITA) - which collectively account for 99% of sales in the Australian market. The total tobacco sales volumes include FMC and fine-cut tobacco converted to cigarette stick equivalents (CSE).

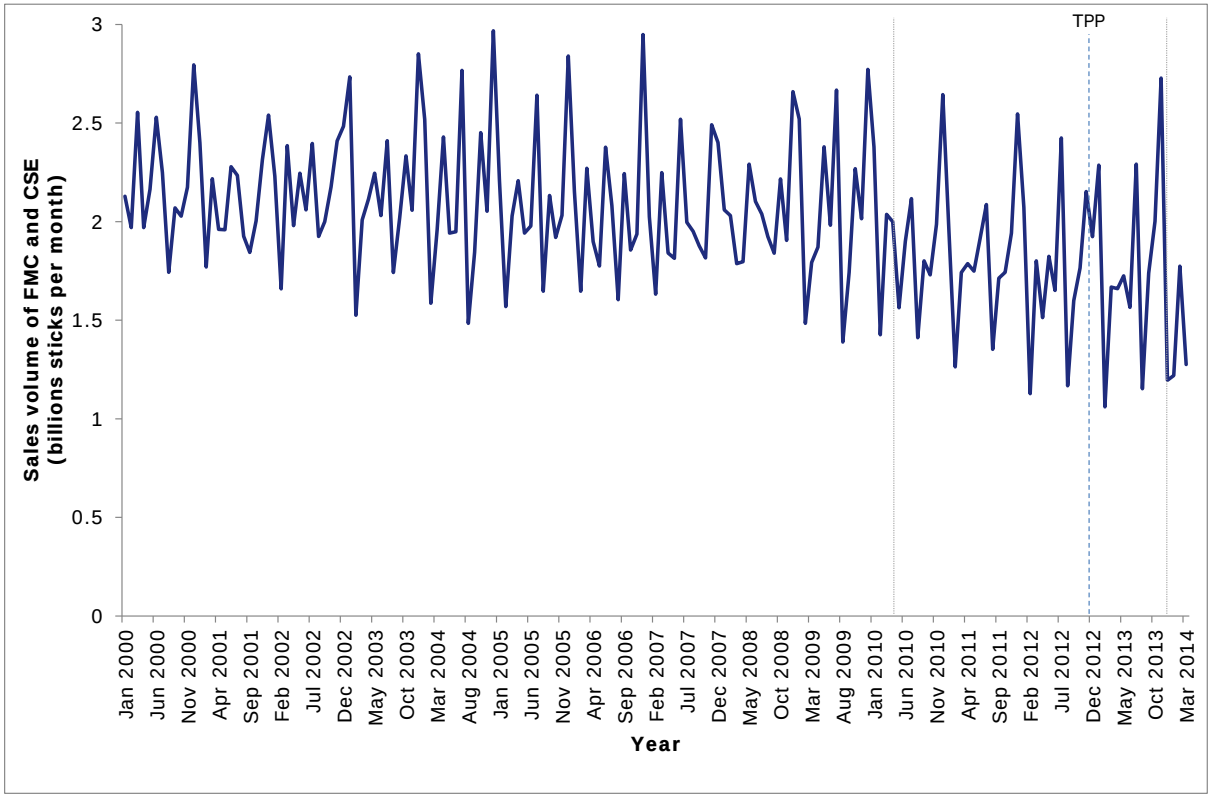
10. IPE notes, as depicted in **Figure D.1**, that cigarette sales volumes based on the IMS/EOS data, as a proxy for cigarette consumption, are marked by strong seasonal patterns, with sales spikes at the end of each quarter (especially in December and June), followed by a trough immediately thereafter. IPE further argues that there is an observable overall downward trend in the total volume of cigarette sticks (FMC and CSE) sold in Australia over the period 2000-2013, as

[7] See Australia's response to Panel question No. 196, para. 214.

[8] We note that the parties also submitted data on the value and volume of Australian imports of tobacco, including cigars and cigarillos. See HoustonKemp Report, (Exhibit AUS-19) (SCI); Dominican Republic's response to Panel question No. 5.; Honduras's response to Panel question No. 5; Cuba's response to Panel question No. 5; and Indonesia's response to Panel question No. 5. These data are not reviewed here because they were not used in any of the econometric reports submitted by the parties.

depicted in **Figure D.2**.[9] IPE contends that the downward trend accelerated after the introduction of the 2010 excise tax increase. IPE further submits that the data reveal an increase in the total volume of cigarette sticks sold in 2013.

**Figure D.1: Monthly Cigarette Sales Volumes Based on IMS/EOS Data**

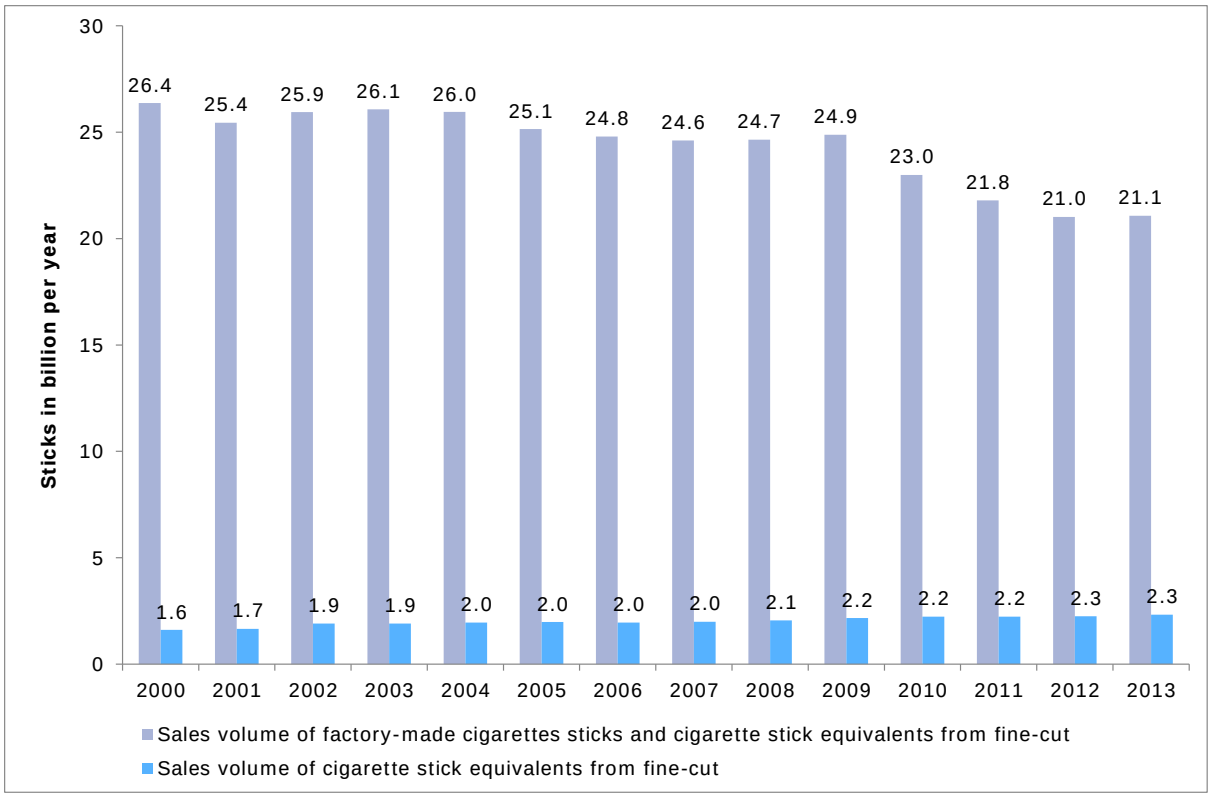



Note: Conversion rate of 0.8 grams of fine-cut tobacco per stick. The vertical dotted lines indicate excise tax increases. The vertical dashed line indicates the introduction of TPP and enlarged GHWs.

Source: IPE Report, (Exhibit DOM-100), p. 67.

[9] See IPE Report, (Exhibit DOM-100), pp. 65-68.

**Figure D.2: Annual Cigarette Sales Volumes Based on IMS/EOS Data**



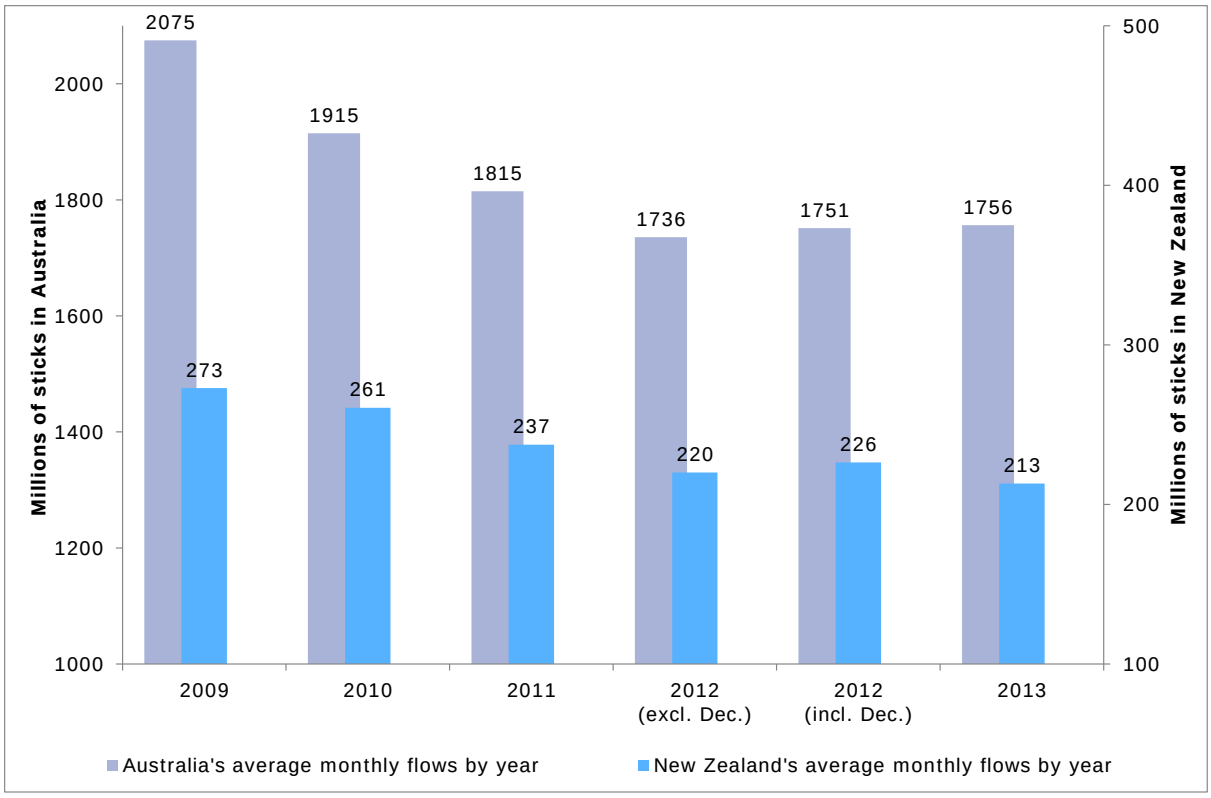

Note: Conversion rate of 0.8 grams of fine-cut tobacco per stick.

Source: IPE Report, (Exhibit DOM-100), p. 66.

11. Professor Klick argues that while sales data are not equivalent to smoking incidence, market data are valuable because they do not suffer from the potential inaccuracies associated with self-reported survey responses, and captures even those smokers who would not be included in the survey sample such as underage smokers.[10] Professor Klick further submits that there is some doubt regarding whether the TPP measures are associated with a decline in cigarette flows given that despite the steady decline in cigarette sales flows observed in the pre-TPP period, cigarette flows increased by 0.3% between 2012 and 2013, as depicted in **Figure D.3**. In comparison, Professor Klick argues that although the pre-December 2012 decline in average monthly cigarette sales was just as striking in New Zealand as in Australia, the decline in cigarette sales continued into 2013 in New Zealand, unlike in Australia.[11] In addition, Professor Klick contends that, unlike Dr Chipty's claim of positive strategic inventory effects on sales in anticipation of the December 2013 tax increase in Australia , the decline in cigarette wholesale sales in Australia appears to have diminished with the introduction of the TPP measures unlike in New Zealand.[12]

[10] See Klick Report, (Exhibit UKR-5), p. 12.
[11] See Klick Report, (Exhibit UKR-5), p. 16.
[12] See Klick Rebuttal Report, (Exhibit HND-118), paras. 66-68.

**Figure D.3: Average Monthly Cigarette Sales Volumes Based on IMS/EOS Data**



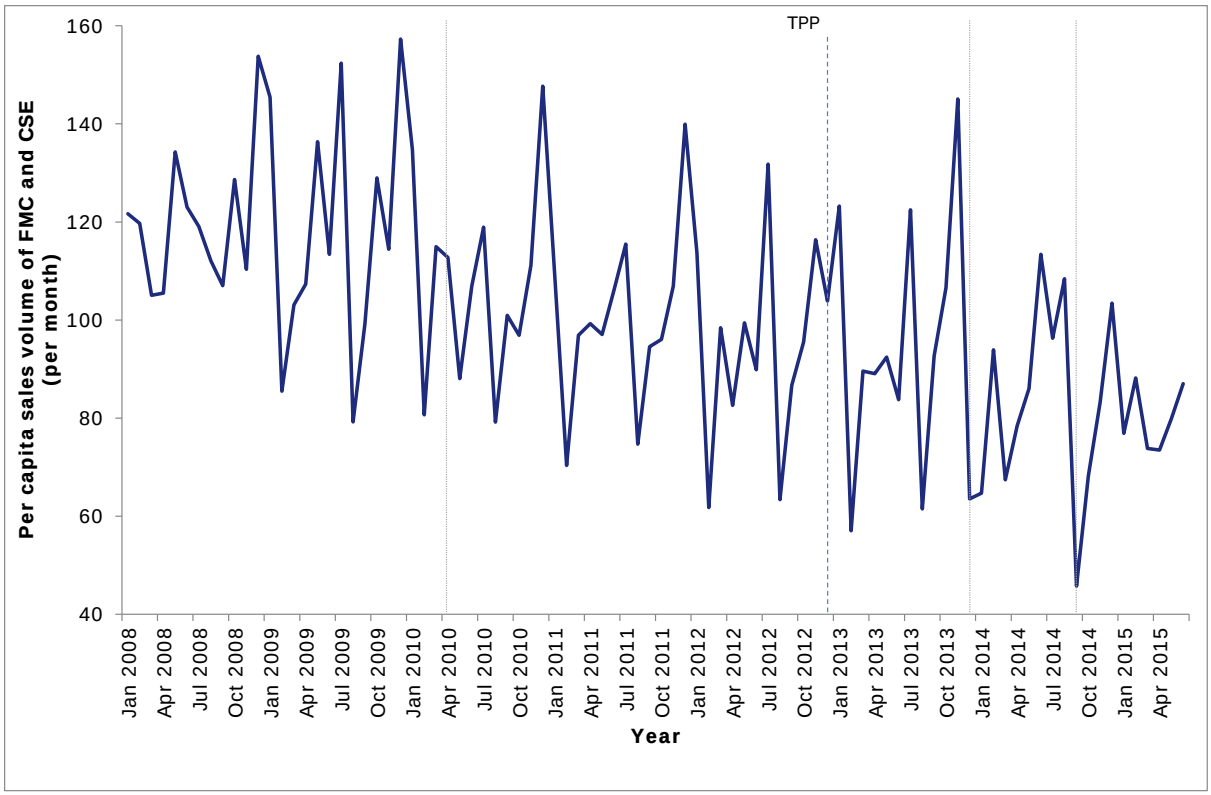

Note: Conversion rate of 0.8 grams of fine-cut tobacco per stick.

Source: Klick Report, (Exhibit UKR-5), pp. 17-18.

12. Unlike IPE and Professor Klick, Professor List considers the evolution of cigarette sales per capita, as shown in **Figure D.4**, and argues that smoking consumption is characterized by a general downward trend in smoking consumption.[13]

[13] See List Report, (Exhibit DOM/IDN-1), p. 27; List Rebuttal Report, (Exhibit DOM/IDN-3), p. 12; List Second Supplemental Report, (Exhibit DOM/IDN-5), p. 8; and List Third Supplemental Report, (Exhibit DOM/IDN-7), p. 35. We note that IPE presents also a graphical analysis of cigarette sales per capita in its second report, without updating it in its subsequent reports. IPE Updated Report, (Exhibit DOM-303), p. 57.

**Figure D.4: Per Capita Monthly Cigarette Sales Volumes Based on IMS/EOS Data**




Note: Conversion rate of 0.8 grams of fine-cut tobacco per stick. The vertical dotted lines indicate excise tax hikes. The vertical dashed line indicates the introduction of TPP and enlarged GHWs.

Source: List Rebuttal Report, (Exhibit DOM/IDN-3), p. 12; and List Third Supplemental Report, (Exhibit DOM/IDN-7), p. 35.

13. Australia's expert Dr Chipty argues that IPE's and Professor Klick's graphical analysis is misleading and fails to account for strategic inventory management that likely took place on the eve of the December 2013 tax increase in the post-TPP period. Dr Chipty is of the view that the presence of the anticipated tax response at the end of December 2013 coupled with the fact that the TPP measures went into effect in October 2012 makes the comparison of sales volume between 2012 and 2013 not meaningful. Dr Chipty contends that a comparison of the year beginning October before and after TPP is analytically sounder and shows a reduction in cigarettes sales volumes.[14]

### 1.1.2 Nielsen Retail Sales (Australia and New Zealand)

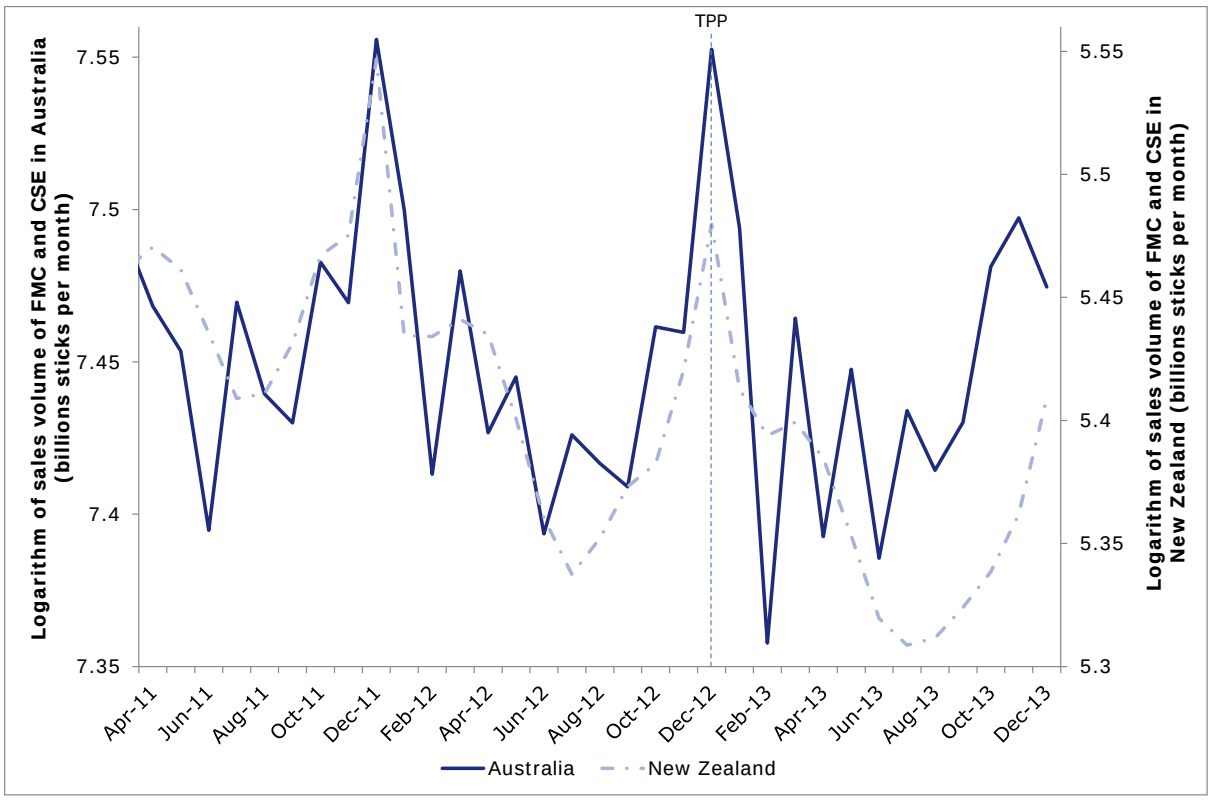
14. The Nielsen data reports actual sales of cigarettes, including "roll your own" equivalents, through all retailers in Australia and New Zealand's markets except specialty tobacconist shops in Australia from February 2011 through December 2013. Professor Klick also included in his analysis Nielsen processed data on the specialty tobacconist channel, which is collected by BATA and not available from any other source. According to Professor Klick, the inclusion of this channel is preferred as the specialty tobacconist channel in Australia accounts for about 25% of the Australian market. The Nielsen data for Australia are available at monthly intervals, while they are available for New Zealand on a four week rolling basis.[15]

[14] See Chipty Report, (Exhibit AUS-17), paras. 62-66.

[15] The four-week interval of the Nielsen data for New Zealand does not align with months. Professor Klick notes that: "[t]o align the periods between the two countries, I allocated the New Zealand 4 week periods to the months in which they were collected, using linear interpolation to account for periods that spanned two months". See Klick Report, (Exhibit UKR-5), fn 35.

15. Professor Klick argues that there was a decline in sales post-TPP in Australia. Specifically, average monthly sales in Australia before TPP (1,720 million) fell to 1,719 million for a reduction of 0.07%. However this reduction in cigarette sales is lower after the introduction of the TPP measures than the drop of almost 6% in New Zealand during the same period and without any TPP, as shown in **Figure D.5**.[16]

**Figure D.5: Cigarette Sales Volumes Based on Nielsen Data**




Note: Conversion rate of 0.8 grams of fine-cut tobacco per stick. The vertical dashed line indicates the introduction of TPP and enlarged GHWs.

Source: Klick Report, (Exhibit UKR-5), pp. 6 and 15.

16. Australia's expert Dr Chipty is of the view that Professor Klick's analysis is uninformative. She argues that even in the pre-TPP period, cigarette sales volume is trending down faster than in Australia. In her view, New Zealand appears to be trending differently from Australia before and after the introduction of the TPP measures.[17]

### 1.1.3 Aztec Scanner Retail Sales

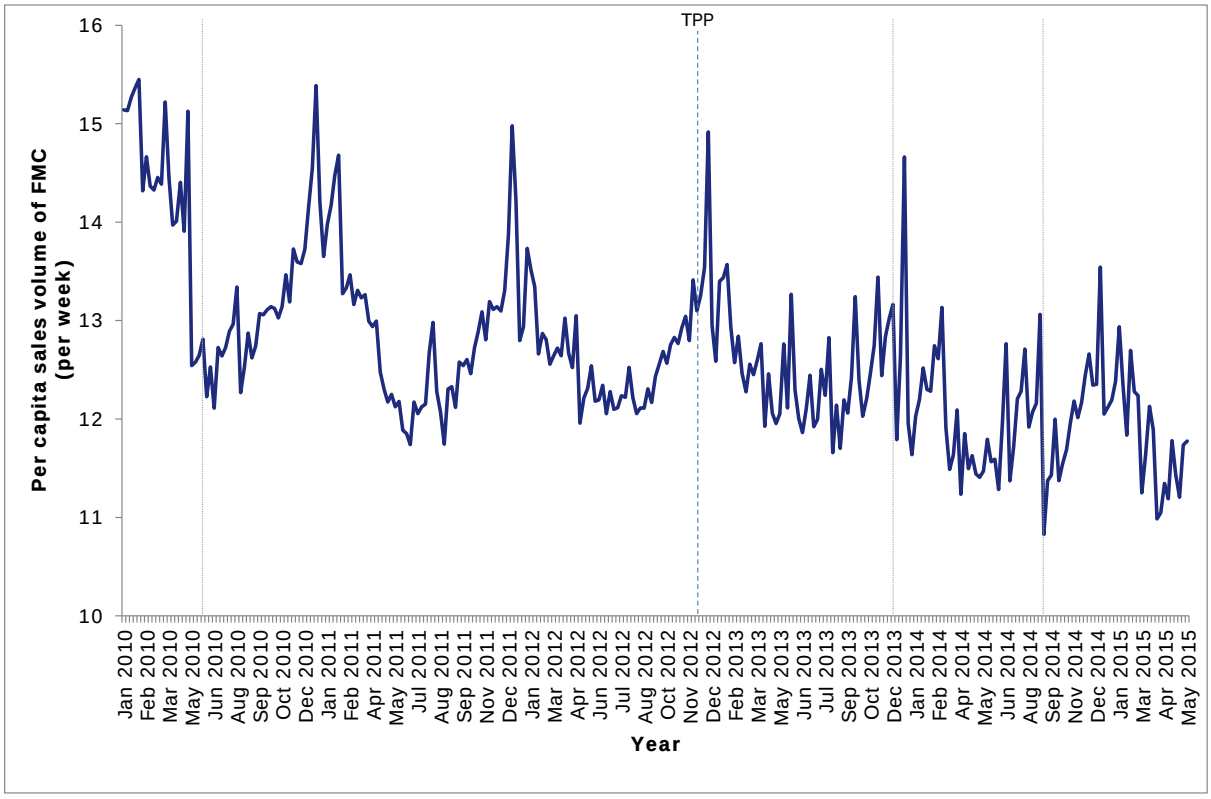
17. The Aztec scanner data tracks sales on a weekly basis at the store- and stock-keeping unit-level, recording detailed product information on brand and variant, quantity, and price per package for factory-made cigarettes and fine-cut tobacco sold across Australia. The Aztec scanner dataset covers the period 27 July 2008 through 27 September 2015.[18]

18. The experts of the Dominican Republic, Honduras and Indonesia do not present a graphical analysis using the Aztec data. Instead, they use the Aztec scanner data to estimate econometrically the TPP measures' impact on (per capita) cigarette sales volumes, which will be

[16] Klick Report, (Exhibit UKR-5), p.15.
[17] See Chipty Report, (Exhibit AUS-17), paras. 45-48.
[18] As discussed below, Professor Klick combines the Nielsen and Aztec data without the Convenience Independent sample when Nielsen data are missing.

reviewed in detail below. In that context, Professor List reports the evolution of per capita sales based on the Aztec data, as depicted in **Figure D.6**.[19]

**Figure D.6: Per Capita Monthly Cigarette Sales Volumes Based on Aztec Scanner Data**




Note: Conversion rate of 0.8 grams of fine-cut tobacco per stick. The vertical dotted lines indicate excise tax hikes. The vertical dashed line indicates the introduction of TPP and enlarged GHWs.

Source: List Rebuttal Report, (Exhibit DOM/IDN-3), p. 18.

19. Australia's expert, Dr Chipty, notes that the Aztec scanner data covers only a portion of total retail sales in Australia, as evident in a side-by-side comparison of the Aztec, Nielsen and IMS/EOS data, as depicted in **Figure D.7**. Dr Chipty further argues that the (econometric) analysis of the Aztec scanner data is misleading because the share of total cigarette sales covered by the Aztec data has grown over time. Dr Chipty submits that these changes in the underlying data interfere with the ability to use them meaningfully to study changes in cigarette sales before and after the introduction of the TPP measures.[20]

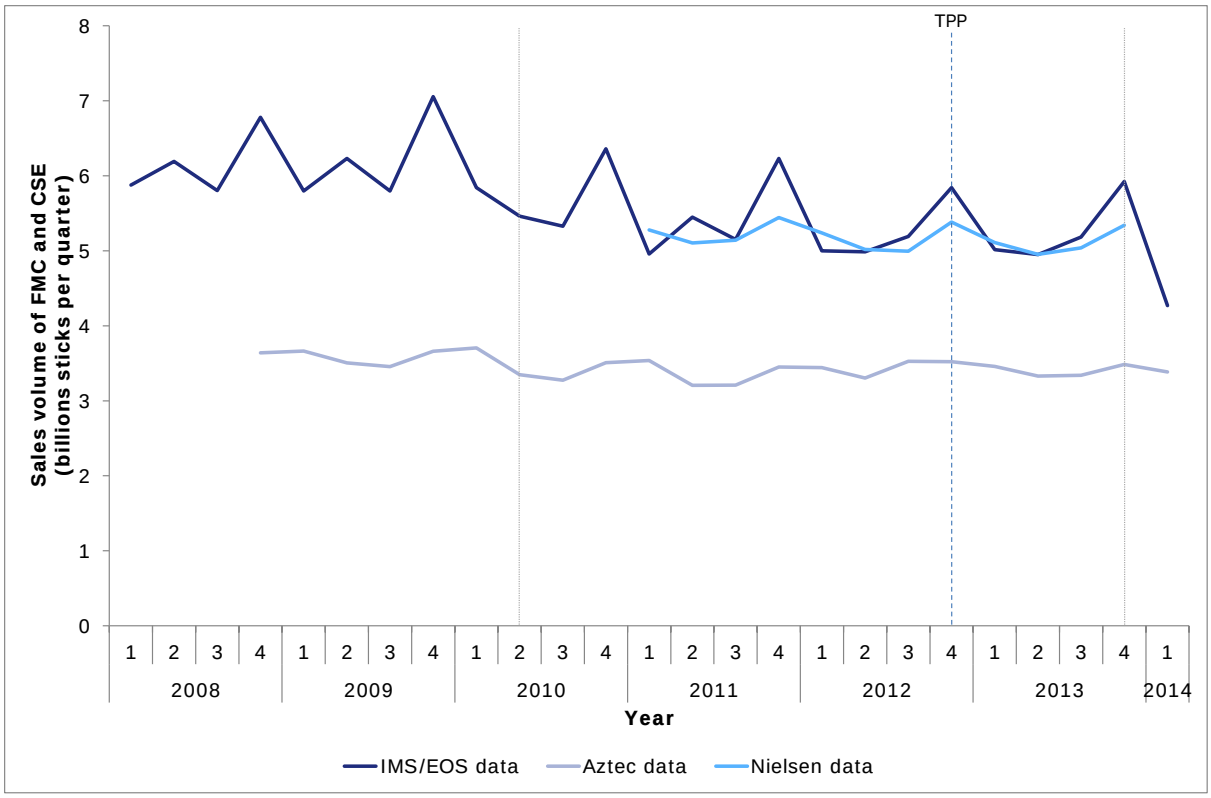
20. Professor Klick disagrees with Dr Chipty and submits that the Aztec dataset coverage of the retail market has stayed substantially the same over the sample period, while noting that there are changes as some retail outlets close and others open (including within the convenience independent market segment). Professor Klick argues that Dr Chipty did not present direct evidence demonstrating that the coverage of the Aztec sales data relative to total sales is growing over time. Professor Klick further rejects Dr Chipty's indirect evidence of such a change, which can, according to Professor Klick, be explained by the strategic inventory management that Dr Chipty identified.[21]

[19] See List Rebuttal Report, (Exhibit DOM/IDN-3), p. 18. We note, however, that unlike IPE, Professor List does not include the sales of RYO tobacco in his analysis of the Aztec data.

[20] See Chipty Report, (Exhibit AUS-17), paras. 55-61.

[21] See Klick Third Supplemental Rebuttal Report, (Exhibit HND-166), paras. 33-34; and Klick Fourth Supplemental Rebuttal Report, (Exhibit HND-169), paras. 6-8 fn 4.

**Figure D.7: Quarterly Cigarette Sales Volume based on IMS/EOS, Nielsen and Aztec Data**




Note: Conversion rate of 0.8 grams of fine-cut tobacco per stick. The vertical dotted lines indicate excise tax hikes. The vertical dashed line indicates the introduction of TPP and enlarged GHWs.

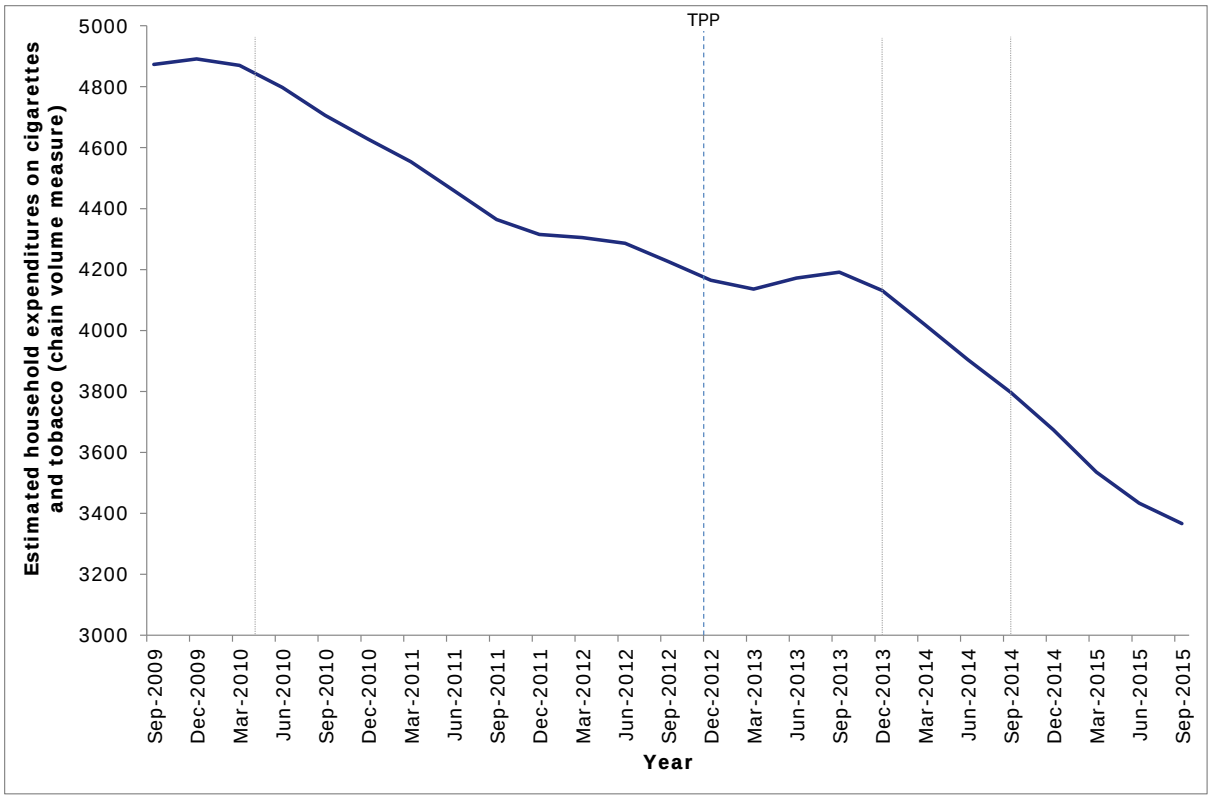
Source: Chipty Report, (Exhibit AUS-17), p. 33.

### 1.1.4 Australian Bureau of Statistics Household Expenditure

21. Australia's Post-Implementation Review Tobacco Plain Packaging 2016 reports the evolution of estimated household expenditure on tobacco and cigarettes released by the Australian Bureau of Statistics (ABS), as shown in **Figure D.8**. According to the ABS, although there was a rise in estimated consumption expenditure in the June 2013 and September 2013 quarters compared with the previous quarters, estimated consumption expenditure on tobacco and cigarettes has been declining in the March 2013 quarter and in all other quarters since implementation.[22]

[22] See Tobacco Plain Packaging PIR, (Exhibit AUS-624), paras. 132-134. The Tobacco Plain Packaging PIR also refers to a 2014 Euromonitor Report on Tobacco, which shows a continued decline in tobacco sales, although limited information about the sources of the data and methodological processes used is provided in the report. Ibid. para. 135. Similarly, the Tobacco Plain Packaging PIR refers to the Australian Taxation Office and Customs clearance data showing a reduction in tobacco clearances in stick equivalent terms (including excise and customs duty) between calendar year 2012 and calendar year 2014.

**Figure D.8: Household Expenditure on Tobacco and Cigarettes based on ABS Data**




Note: The vertical dotted lines indicate excise tax hikes. The vertical dashed line indicates the introduction of TPP and enlarged GHWs.

Source: Tobacco Plain Packaging PIR, (Exhibit AUS-624), p. 45.

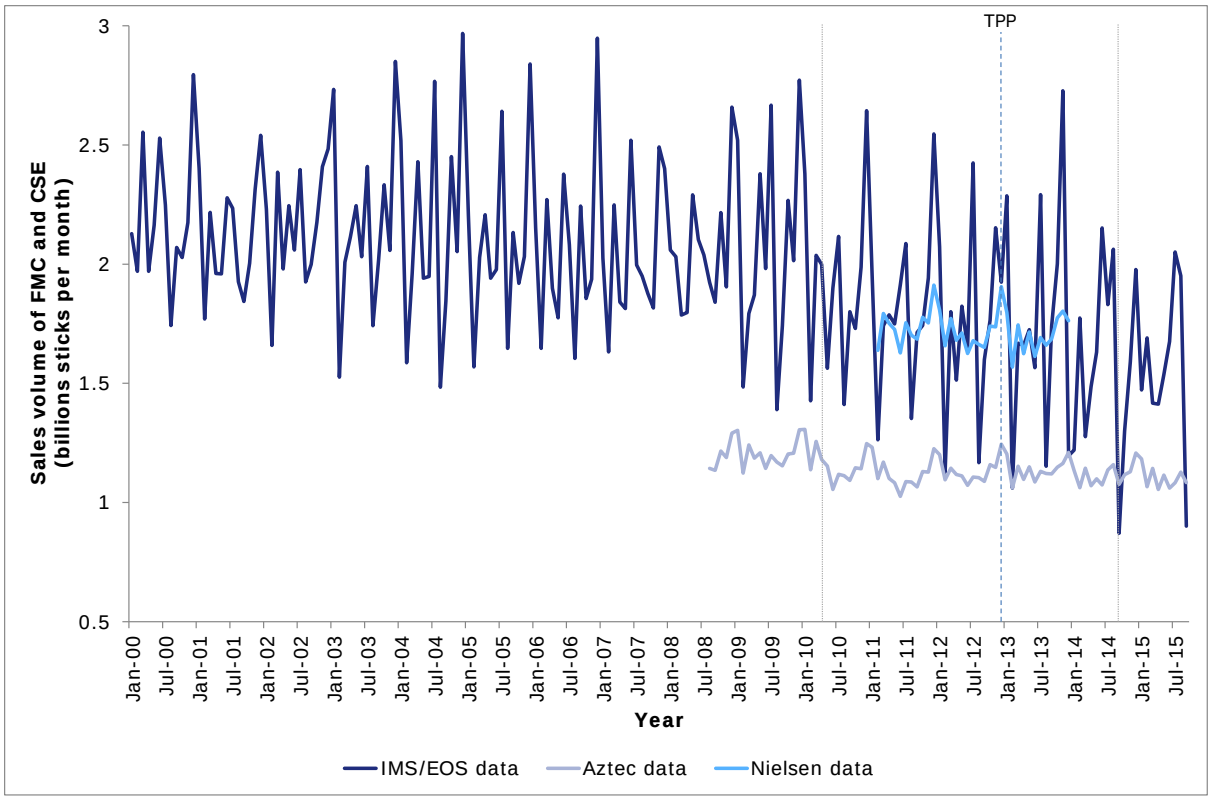
## 1.2 Analysis by the Panel

22. As with smoking prevalence, we acknowledge the importance of analysing the trends in cigarette sales with the most recent available data. We further recognize, as pointed out by IPE, that there is no perfect dataset for a sales analysis in terms of market coverage, frequency (weekly, monthly or annually) and period covered.[23]

23. The IMS/EOS data cover sales from manufacturers to wholesalers and retailers, while both Aztec and Nielsen data cover sales from retailers to consumers. Similarly, both Aztec and Nielsen data are only available for February 2011 to December 2013, and 27 July 2008 to 27 September 2015, respectively, while the IMS/EOS data cover the larger sample period, from January 2000 to September 2015. The Nielsen data and the IMS/EOS data cover, respectively, almost 100% and 99% of the Australian market, while the Aztec data cover 67% of the Australian market.[24] We also note that the experts of the Dominican Republic, Honduras and Indonesia used different definitions of cigarette sales, namely level of cigarette sales and per capita cigarette sales. We further note that survey datasets on cigarette consumption have been used by some of the experts of the Dominican Republic, Honduras and Indonesia to analyse the impact of the TPP measures. These will be reviewed in detail when discussing the econometric analysis.

[23] See IPE Updated Report, (Exhibit DOM-303), para. 137.
[24] See IPE Summary Report, (Exhibit DOM-379), paras. 131-133.

**Figure D.9: Monthly Cigarette Sales Volumes Based on IMS/EOS, Nielsen and Aztec Data**




Note: Conversion rate of 0.8 grams of fine-cut tobacco per stick. The vertical dotted lines indicate excise tax hikes. The vertical dashed line indicates the introduction of TPP and enlarged GHWs.

Source: IPE Summary Report, (Exhibit DOM-379), back-up material.

24. We note, at the outset, that neither IPE nor Professor Klick updated their graphical analysis of cigarette sales volume with more recent data in the course of the proceedings. The Dominican Republic, Honduras and Indonesia no longer referred to their initial assertion that cigarette sales volumes increased in the post-TPP period. After a careful review of the most recent available IMS/EOS and Aztec datasets, we note that despite fluctuating, cigarette sales volumes follow a downward trend in the period following the introduction of the TPP measures, as depicted in **Figure D.9**.[25] However, the downward trend is more pronounced in the IMS/EOS data than in the Aztec scanner data. As pointed out by Australia and acknowledged by Professor List and IPE[26], the share of total cigarette sales covered by the Aztec data has grown over time, which would explain, at least partially, why the downward trend is less pronounced in the Aztec data.[27]

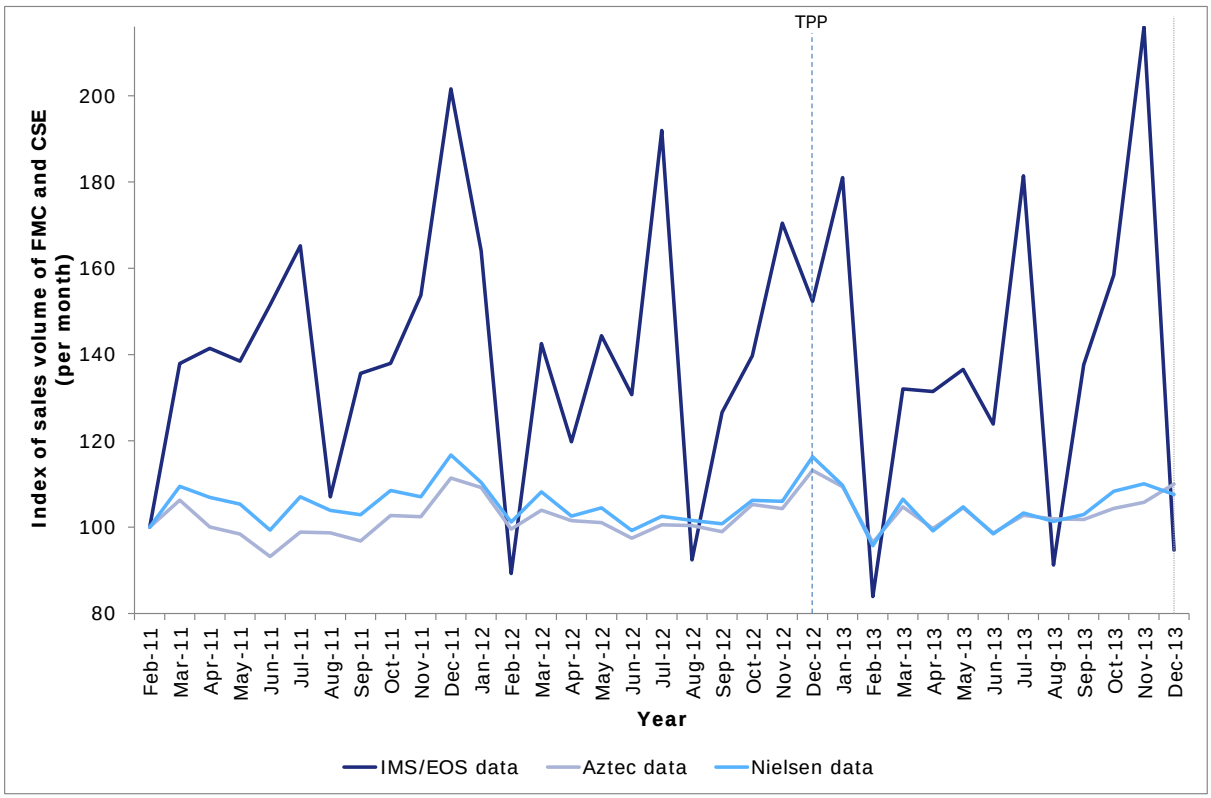
[25] The weekly Aztec data were aggregated to monthly data by assuming that each weekly cigarette quantity can be evenly split for each day of the week.

[26] See List Rebuttal Report, (Exhibit DOM/IDN-3), fn 15; and IPE Third Updated Report, (Exhibit DOM-375), para. 171.

[27] We note that Professor Klick is the only one of the complainants' experts to reject the claim that the Aztec dataset's market coverage has increased. Professor Klick argues that the upward trend of the ratio between the retail and wholesale data is the result of strategic inventory management. See Klick Fourth Supplemental Rebuttal Report, (Exhibit HND- 169), paras. 33-34. Dr Chipty disagrees with Professor Klick's claim and argues that strategic inventory management cannot explain the upward trend of the ratio beyond the 2012 and 2014 excise tax increase. See Chipty Third Rebuttal Report, (Exhibit AUS-605), fn 148. We note that Professor Klick does not provide in his reports any evidence that would suggest that the retail market coverage has stayed substantially the same over the sample period. In fact, a review of the underlying Aztec data reporting the associated store count for each sales channel shows that the number of stores covered by the data increased from 6,605 in January 2012 to 9,437 covered stores in September 2015. See Klick Fourth Supplemental Rebuttal Report, (Exhibit HND-169), back up material. We also note that a comparison of the Aztec data to tax statistics suggests that the market share covered by the Aztec has increased over time. The same conclusion applies with a comparison of the Aztec data to the IMS/EOS data. See IPE Report, (Exhibit DOM-100), pp. 101-103.

25. We will review in the next subsections all the econometric evidence submitted to us based on the IMS/EOS, Nielsen and Aztec data. However, we consider the IMS/EOS data to be the most appropriate available market data for analysing the impact of the TPP measures on cigarette sales, for a number of reasons. First, the IMS/EOS data are available for the longest period, which is more likely to yield more accurate estimates. Second, although the IMS/EOS data do not cover sales from retailers to consumers, the correlation between the IMS/EOS data and the Nielsen data, which do cover sales from retailers to consumers, is relatively high, with a correlation coefficient of 0.58, as shown in **Figure D.10**.[28] The correlation coefficient is even higher when the unit of analysis is quarterly sales, with a correlation coefficient equal to 0.85.[29]

**Figure D.10: Indexes of Cigarette Sales Volumes Based on IMS/EOS, Nielsen and Aztec Data**




Note: Conversion rate of 0.8 grams of fine-cut tobacco per stick. The vertical dotted lines indicate excise tax hikes. The vertical dashed line indicates the introduction of TPP and enlarged GHWs. Each data are normalized to its respective cigarette sales in January 2011.

Source: IPE Summary Report, (Exhibit DOM-379), back-up material.

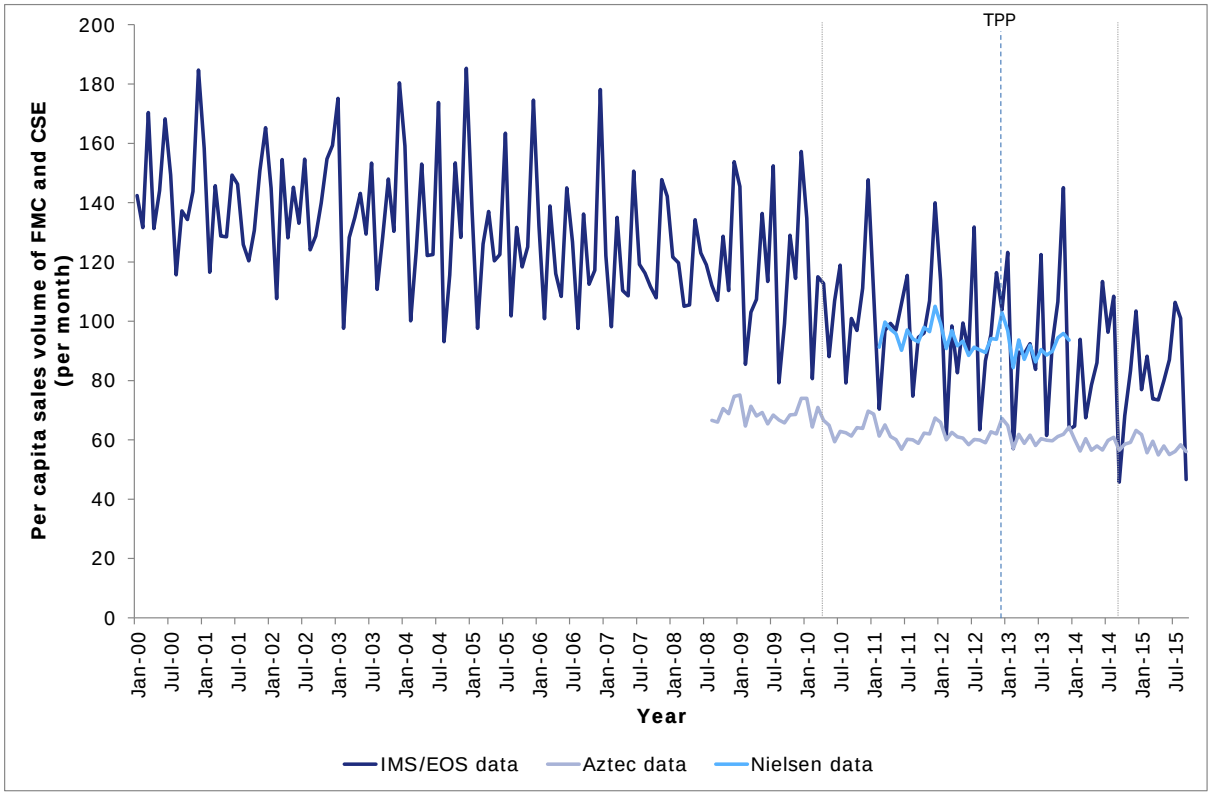
[28] According to the English Oxford Dictionary, the correlation coefficient is a number between -1 and 1 calculated so as to represent the linear interdependence of two variables. *Shorter Oxford English Dictionary*, 5th edn, W.R. Trumble and A. Stevenson (eds.) (Oxford University Press, 2002), Vol. 1, p. 524.

[29] We note that Professor Klick submits that for the March 2012-December 2013 period the correlation coefficient between the Nielsen data and the Aztec data is 0.985 if the Convenience Independent sample is included in the Nielsen data, and 0.999 if the Convenience Independent sample is omitted. See Klick Supplemental Rebuttal Report, (Exhibit HND-122), para. 123. We note, however, that the monthly Aztec data considered by Professor Klick differs significantly from the monthly cigarette sales constructed from the weekly Aztec data considered by IPE, Professor List, and Dr Chipty. This explains why the correlation coefficient between the Nielsen data (including the Convenience Independent sample) and the Aztec data we obtained for the March 2012-December 2013 period is different and equal to 0.929. In addition, we note that the correlation coefficient between the Nielsen data and the Aztec data for the February 2011-December 2013 period is smaller and equal to 0.84. In other words, the Nielsen data and the Aztec data are less correlated for the February 2011-February 2012 period than for the March 2012-December 2013 period, which is in line with the Aztec data's growing market coverage (given that the Nielsen data's market coverage is close to 100%). The correlation coefficient is also smaller (0.756) when the data are aggregated at the quarter level for the February 2011-December 2013 period.

26. Although the Nielsen data cover sales from retailers to consumers, which represent a closer measure to cigarette consumption than the IMS/EOs data, they are only available for a short period, with only 13 post-implementation observations, which could make it more difficult to estimate accurately any impact on sales volumes. Similarly, although the Aztec data have the advantage of covering sales from retailers to consumers, the share of total cigarette sales covered by the Aztec data has grown over time, unlike the general downward trend of cigarette sales in the Australian market, which could make it more difficult to identify the impact of any factors from the increasing market coverage.

27. The evolution of per capita cigarette sales volumes, as depicted in **Figure D.11**, leads us to the same conclusion, namely that the evidence before us suggests that per capita cigarette sales have, on average, continued to decrease after the introduction of the TPP measures. This finding is in line with the evolution of house expenditure on tobacco and cigarettes. Evidence before us on the extent to which the downward trend in cigarette sales has accelerated or not in the post-TPP period is reviewed next.

**Figure D.11: Per Capita Cigarette Sales Volumes Based on IMS/EOS, Nielsen and Aztec Data**




Note: Conversion rate of 0.8 grams of fine-cut tobacco per stick. The vertical dotted lines indicate excise tax hikes. The vertical dashed line indicates the introduction of TPP and enlarged GHWs.

Source: IPE Summary Report, (Exhibit DOM-379), back-up material.

## 2 WHETHER THE REDUCTION OF CIGARETTE SALES VOLUMES ACCELERATED AFTER THE ENTRY INTO FORCE OF THE TPP MEASURES

28. Having determined that the most recent market data show continuing declines in cigarette sales volumes in the period following the introduction of the TPP measures, we turn to whether there was a shift in cigarette sales volumes in the post-TPP implementation period. Instead of assessing directly the impact of the TPP measures on cigarette sales, which we review next, the Dominican Republic's and Indonesia's expert Professor List investigated whether there was a change in cigarette sales in the post-TPP implementation period. The underlying assumption is that

if cigarette sales follow the same pre-existing pattern after the introduction of the TPP measures, then it follows that the TPP measures failed to reduce cigarette sales.

29. Professor List investigates this question using two distinct datasets: the IMS/EOS data and the Aztec scanner data.[30] We describe Professor List's results and Australia's criticisms before presenting our analysis of this evidence.

## 2.1 Main datasets and arguments

### 2.1.1 In-Market-Sales/Exchange of Sales

30. Professor List presents an event study analysing whether there has been a shift in the evolution of per capita cigarette sales following the introduction of TPP and enlarged GHWs. Specifically, a seasonally adjusted Autoregressive Integrated Moving Average (ARIMA) model and a dynamic model of per capita cigarette wholesales volumes are estimated controlling for per capita cigarette wholesales during the previous month, price during the previous month, a linear time trend and the 2006 GHWs.[31] The most recent analysis covers the period from February 2002 to June 2015.

31. Overall, Professor List concludes that according to both models, there is no statistical difference between the observed per capita cigarette sales volumes and the estimated counterfactual per capita cigarette sales volumes, implying that the post-implementation downward trend in cigarette sales did not shift.[32]

32. In response, Australia's expert, Dr Chipty, argues that Professor List's event study is fundamentally flawed. Dr Chipty contends that the estimates of Professor List's first-stage consumption model are highly imprecise and many of the explanatory variables are statistically insignificant, sometimes with the wrong estimated coefficient sign, making the post-TPP projections unreliable. In her view, the estimated moving average parameters of the seasonal ARIMA model indicate that the model is not specified correctly and most likely over-differenced. According to Dr Chipty, this over-differencing would likely lead to large forecasting errors when the model parameters are used to predict consumption levels during the post-period. Dr Chipty further argues that Professor List's analysis of the 2010 excise tax increase does not bolster confidence in Professor List's conclusion that there is no shift in the downward trend in cigarette sales, because the result of the ordinary least square (OLS) model, intended to corroborate the result for the TPP measures, predicts higher consumption with taxes than without.[33]

33. Dr Chipty also submits that by controlling for prices in the pre-period model and using actual prices in the post-period forecasting, Professor List's event study fails to account for the effect of tobacco control policies that work through price.[34] According to Dr Chipty, failure to control for the 2010 excise tax increase has the effect of crediting the trend with the effect of the tax increase. In her view, Professor List implicitly assumes that the effect of the 2010 tax increase will continue in perpetuity.[35] Dr Chipty reconsiders the dynamic model by controlling for the 2010 excise tax

[30] We note that Professor List's event study is a trend-projection analysis that does not estimate directly the impact of the TPP measures, but rather assesses whether there is a difference between the counterfactual per capita cigarette sales in the absence of the TPP measures and the actual per capita cigarette sales. That is why we have decided to discuss Professor List's event study under the subsection addressing whether there was a change in (per capita) cigarette sales trend in the post-implementation period.

[31] An event analysis consists of (1) estimating the model of per capita cigarette sales during the pre-TPP period; (2) forecasting the per capita cigarette sales that would have been prevailed in the absence of the TPP measures using the estimated model's parameters in the post-implementation period; and (3) determining whether the difference between the observed per capital cigarette sales volumes and the estimated counterfactual cigarette sales volumes is statistically different from zero. See List Report, (Exhibit DOM/IDN-1), paras. 123-132.

[32] See List Report, (Exhibit DOM/IDN-1), paras. 133-134; List Rebuttal Report, (Exhibit DOM/IDN-3), paras. 65-68; List Second Supplemental Report, (Exhibit DOM/IDN-5), paras. 26, 72, 74 and 144-154; and List Third Supplemental Report, (Exhibit DOM/IDN-7), paras. 26-28 and 56-61.

[33] See Chipty Rebuttal Report, (Exhibit AUS-535) (SCI), paras. 37-44; and Chipty Surrebuttal Report, (Exhibit AUS-586), paras. 30-33.

[34] See Chipty Rebuttal Report, (Exhibit AUS-535) (SCI), paras. 33-44; and Chipty Surrebuttal Report, (Exhibit AUS-586), paras. 30-33.

[35] Chipty Third Rebuttal Report, (Exhibit AUS-605), paras. 61—67.

increase and finds a statistically significant shift in the post-implementation downward trend in cigarette sales. However, controlling for other problems, namely correcting the definition of the GHWs variable (from March 2006 onwards instead of between March 2006 and November 2012), accounting for all excise tax changes, and strategic inventory management associated with the tax changes, Dr Chipty concludes that the shift of the post-implementation downward trend in cigarette sales is economically significant (similar in magnitude to the estimates obtained in the preferred before-after consumption model) but not statistically significant. More generally, Dr Chipty contends that Professor List's conclusions regarding the statistical significance of the TPP measures' effect are unreliable because Professor List does not calculate any standard error associated with the estimated policy effect in the first stage. Dr Chipty submits that a before-after analysis, as discussed below, is preferable to Professor List's event study analysis for determining statistical significance.[36]

34. Professor List counters that Dr Chipty does very little to engage with the event study as it pertains to TPP nor with the results reported but tears down an analysis of Australia's 2010 tax policy he never conducted. Professor List submits that when properly done, an event study identifies a negative and statistically significant effect on cigarette consumption of Australia's 2010 tax increase. According to Professor List, Dr Chipty confuses the interpretation of variables in a predictive model with the interpretation of explanatory variables in a causal model. Unlike Dr Chipty's claim that the over-differencing in the ARIMA model would likely lead to large forecasting errors, Professor List contends that the OLS model of cigarette sales does not suffer from this over-differencing problem. Professor List submits that although the predicted values are treated as if they were true, the hypothesis testing would actually be much wider if the uncertainty related to the predictors were taken into account. Professor List concludes that when an event study on the TPP measures finds no effect on cigarette sales, it is not because the data are too underpowered to detect changes induced by effective tobacco control policies, or because specific control variables have been chosen, but because the TPP measures are not working.[37]

### 2.1.2 Aztec Scanner Retail Sales

35. Professor List performs the same event study using the Aztec scanner data. A seasonally adjusted ARIMA model and a dynamic model of per capita cigarette sales volumes are estimated controlling for per capita cigarette consumption during the previous month, price during the previous month, a linear time trend and the 2006 GHWs. The analysis covers the period from July 2008 to May 2015.

36. Professor List concludes that there is no statistical difference between observed per capita cigarette sales volumes and the estimated counterfactual per capita cigarette sales volumes, implying that the post-implementation downward trend in cigarette sales did not shift. However, one of Professor List's specifications finds an upward and statistically significant trend in per capita cigarette sales, suggesting that per capital cigarette sales have increased in the post-TPP period. Professor List explains that this result is reported for the sake of comprehensiveness but he does not place much emphasis on the finding, given that the Aztec market share is increasing over the relevant time period. [38]

37. Australia's expert Dr Chipty submits that the Aztec data should be disregarded because they are confounded by the growing coverage of the retail outlets included in the data. Dr Chipty submits that given this problem with the Aztec data, the analyses built on these data are also problematic.[39]

38. Professor List counters that Dr Chipty favours discarding the Aztec dataset altogether without making any attempt to apply simple empirical tools available to deal with the alleged problem.[40] In

[36] See Chipty Rebuttal Report, (Exhibit AUS-535) (SCI), paras. 33-44; Chipty Surrebuttal Report, (Exhibit AUS-586), paras. 30-33; Chipty Second Rebuttal Report, (Exhibit AUS-591), paras. 39-42, Appendix C; and Chipty Third Rebuttal Report, (Exhibit AUS-605), paras. 61-68.

[37] See List Second Supplemental Report, (Exhibit DOM/IDN-5), paras. 151-154; and List Third Supplemental Report, (Exhibit DOM/IDN-7), paras. 26-28 and 56-61.

[38] See List Third Supplemental Report, (Exhibit DOM/IDN-7), paras. 56-61 and 80.

[39] See Chipty Surrebuttal Report, (Exhibit AUS-586), para. 31.

[40] See List Summary Report, (Exhibit DOM/IDN-9), para. 107.

particular, Professor List referred to IPE's proposal to benchmark quarterly Aztec data to IMS/EOS data in order to adjust for a growing market coverage of Aztec data.[41]

### 2.2 Analysis by the Panel

39. As explained in our analysis of the contribution of the TPP measures on smoking prevalence, we assess the evidence before us on the basis that our task is not to present a unified econometric analysis of this question but to assess the robustness of the results submitted by the parties. In addition, our conclusions in this respect relate exclusively to the data (including the sample period) and econometric results submitted by the parties in these proceedings, and are not intended to be generalized to other datasets or econometric studies.

#### 2.2.1 Professor List's reports

40. After a careful review of the results of Professor List's event study of the TPP measures' impact on per capita cigarette sales, we question the validity of his results for a number of reasons, some of which are specific to one of the models considered by Professor List.

41. We note that in most estimations of the ARIMA model for the pre-TPP period based on the IMS/EOS data that are used to forecast the value of per capita cigarette sale in the post-implementation period, none of the explanatory variables, including the constant, is statistically significant, besides the moving averages parameters.[42] In our view, the lack of statistically significant variables is surprising, given the relatively small number of explanatory variables included in the ARIMA model. We further note that, as pointed out by Dr Chipty, the estimated moving average parameters are not statistically different from -1, which would likely lead to large forecasting errors. This is, in our view, problematic because Professor List's approach relies on the post-implementation forecasting errors to determine whether there was a statistically significant change in the downward trend in per capita cigarette sales. We note that Professor List recognized the issue of over-differencing but failed to address it or propose a solution. Instead, Professor List suggests considering the dynamic model, which, according to him, does not suffer from this over-differencing problem.[43]

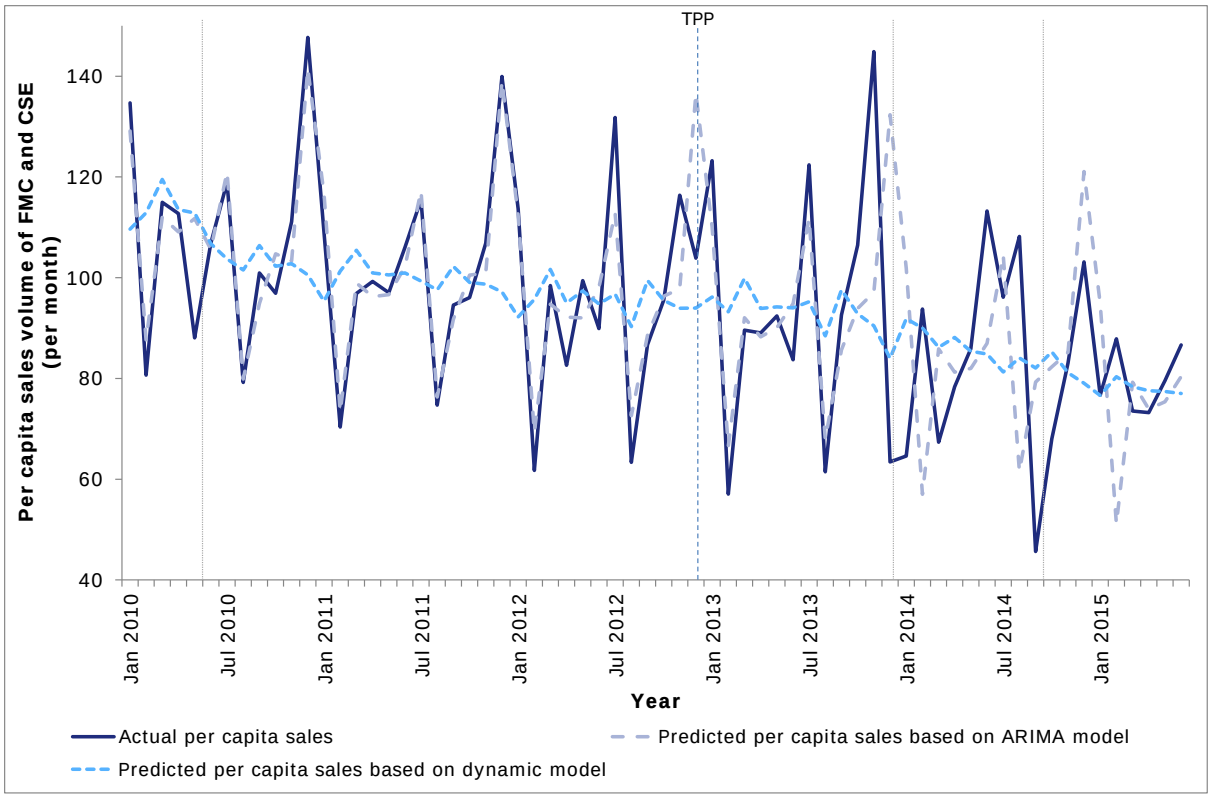
42. Yet, a visual inspection of the results of the ARIMA model and the dynamic model based on the IMS/EOS data shows, as depicted in **Figure D.12**, that most of the estimated per capita cigarette sales associated with the dynamic model are not close at all to the actual values of the per capita sales in the pre-implementation period. In other words, the dynamic model does not fit well the per capita cigarette sales in the pre-TPP period. For instance, the predicted value of per capita sales in December 2010 given by the dynamic model is equal to 100.6, while the actual value of per capita sales amounts to 147.7. Similarly, the dynamic model predicts a value of 95.6 in February 2012, while the observed value is 61.8. We therefore question how the dynamic model specification can be used to accurately forecast the per capita cigarette sales that would have prevailed in the absence of the TPP measures in the post-implementation period, when it already performs relatively poorly in the pre-implementation period.

[41] See IPE Third Updated Report, (Exhibit DOM-375), paras. 169-176.

[42] We note that the same issue applies to Professor List's event study of the 2010 tax change on cigarette sales, where most explanatory variables are not statistically significant, except for some of the moving average parameters. List Second Supplemental Report, (Exhibit DOM/IDN-5), backup material.

[43] See List Second Supplemental Report, (Exhibit DOM/IDN-5), para. 49.

**Figure D.12: Event Study of Per Capita Cigarette Sales Volumes Based on IMS/EOS Data**




Note: Conversion rate of 0.8 grams of fine-cut tobacco per stick. The vertical dotted lines indicate excise tax hikes. The vertical dashed line indicates the introduction of TPP and enlarged GHWs.

Source: List Rebuttal Report, (Exhibit DOM/IDN-3), p. 16.

43. We further note that, similar to the ARIMA model, most explanatory variables in the estimated dynamic models using the IMS/EOS data are not statistically significant (except the constant) at 5%. In particular, the results of the benchmark specification, which includes the price variable in the previous month, suggest that per capita cigarette sales are not explained by the value of per capita sales in the previous month or the linear time trend. This result contradicts the fact that per capita sales follow a dynamic process and a downward trend, as acknowledged by Professor List.[44] While Professor List explains that the predictive models he employs are not designed to precisely describe every causal dynamic in Australia's cigarette market, we note that he does not explain how this justifies the lack of statistically significant explanatory variables.[45] In fact, the only explanatory variable (besides the constant) that is statistically significant is the time trend variable when the price variable is omitted or replaced by a tax level variable. This finding, as explained in the analysis of smoking prevalence, suggests that the price variable and the time trend might be collinear, that is both variables might convey the same information and one of them then becomes redundant. As with the technical exchange between the experts of the Dominican Republic, Honduras, Indonesia and Australia regarding smoking prevalence, they disagree on the use of the price variable. As explained in our review of the econometric studies assessing the impact of the TPP measures on smoking prevalence in Appendix C, we question the results of the dynamic model specification with the price variable based on the IMS/EOS data.[46] Finally, we note that, as pointed out by Dr Chipty, Professor List did not correctly define the March 2006 GHW variable and implicitly assumes that the GHW policy no longer had an effect on consumption as of the beginning

[44] See List Second Supplemental Report, (Exhibit DOM/IDN-5), para. 23.

[45] See List Second Supplemental Report, (Exhibit DOM/IDN-5), para. 149.

[46] We note that the price variable is also included in the ARIMA model but it is expressed in difference (i.e. the difference in price with respect to the previous year). Differentiating the data is often used to address the non-stationary (unit root) of the data. In fact, standard unit root tests suggest that price differential variable is stationary. Therefore, the issue of non-stationary price variable does not apply to the results of Professor List's ARIMA model.

December 2012, when the TPP measures were fully implemented. It is also unclear to what extent defining correctly the dummy variable for the GHW policy would have changed the results.

44. Some of the issues we raised above also apply to the results of Professor List's event study based on the Aztec data. In particular, both the ARIMA model and the dynamic model include the price variable in the previous month, which as explained above, appears to be non-stationary and do not account for the potential impact of the TPP measures on prices.[47] In addition, the results of the dynamic model find a statistically significant upward trend in per capita cigarette sales. As acknowledged by Professor List, Aztec's market share has been increasing over the relevant time period.[48] Yet, Professor List does not control for this feature of the data, so it is unclear to what extent the results of an upward trend shift would prevail, if the increasing market share was accounted for. In any event, we are not persuaded that we can simply ignore or down-weigh this contradictory result, as Professor List suggests, without questioning the model specification and data. Overall, in this context, we consider the results of Professor List's event study to be of limited use in informing whether the TPP measures led to a change in the downward trend of per capita cigarette sales.

### 2.2.2 Dr Chipty's reports

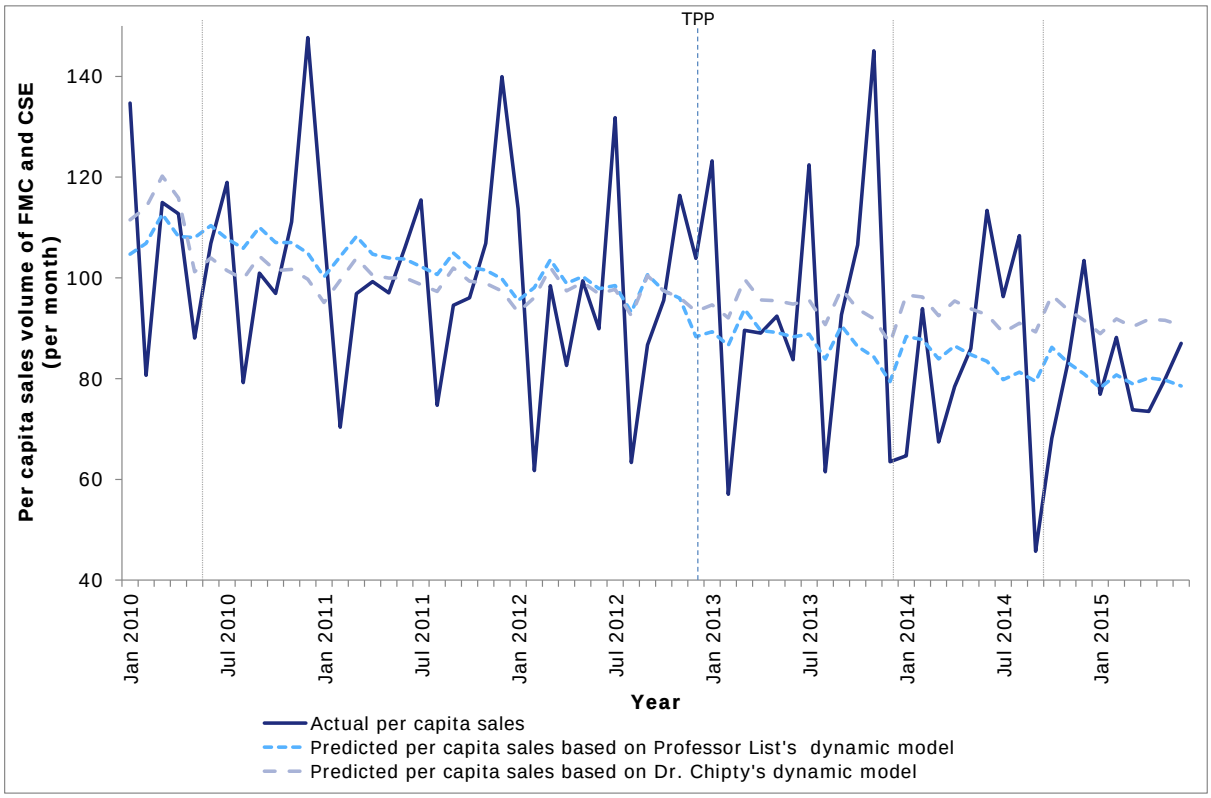
45. We note that Dr Chipty re-estimated Professor List's dynamic model based on the IMS/EOS data by addressing some of the issues we raised above. In particular, replacing the price variable with an excise tax increase dummy reverses Professor List's conclusions and suggests that there was a statistically significant shift in the post-implementation downward trend in per capita cigarette sales. We note, however, that the shift is still negative but no longer statistically significant when the excise tax changes, and strategic inventory management associated with these tax changes are taken into account in the post-implementation estimation. As shown in **Figure D.13**, a visual inspection of the results of Dr Chipty's dynamic model specification based on the IMS/EOS data shows that, although Dr Chipty's specification, with a higher adjusted coefficient of determination, predicts slightly more accurately the values of per capita sales in the pre-implementation period than Professor List, the dynamic model continues to perform relatively poorly in the pre-implementation period.[49] As explained above, we continue to have doubts about how the dynamic model specification can be used to accurately forecast the per capita cigarette sales that would have prevailed in the absence of the TPP measures in the post-implementation period, when it already performs relatively poorly in the pre-implementation period.

[47] We note that Professor List decided to remove the observation associated with 3 August 2008 in his analysis without providing any explanation. We also note that the results of the ARIMA model show that the STATA software is unable to compute the standard error of the second lagged moving average coefficient. In addition, unlike the specification of the ARIMA model based on the IMS/EOS data, Professor List removed the constant term, supposedly as a result of the procedure implemented by Professor List to select the model specification by optimizing the information criteria of the model. See List Rebuttal Report, (Exhibit DOM/IDN-3), fn 14.

[48] See List Rebuttal Report, (Exhibit DOM/IDN-3), fn 15.

[49] The (adjusted) coefficient of determination measures the share of the variance in the dependent variable (e.g. per capita cigarette sales) is predictable from the independent variables included in the model (e.g. price, trend, …). W. H. Green, *Econometric Analysis*, 5th edn (Prentice Hall, 2002).

**Figure D.13: Alternative Event Study of Per Capita Cigarette Sales Volumes Based on IMS/EOS Data**



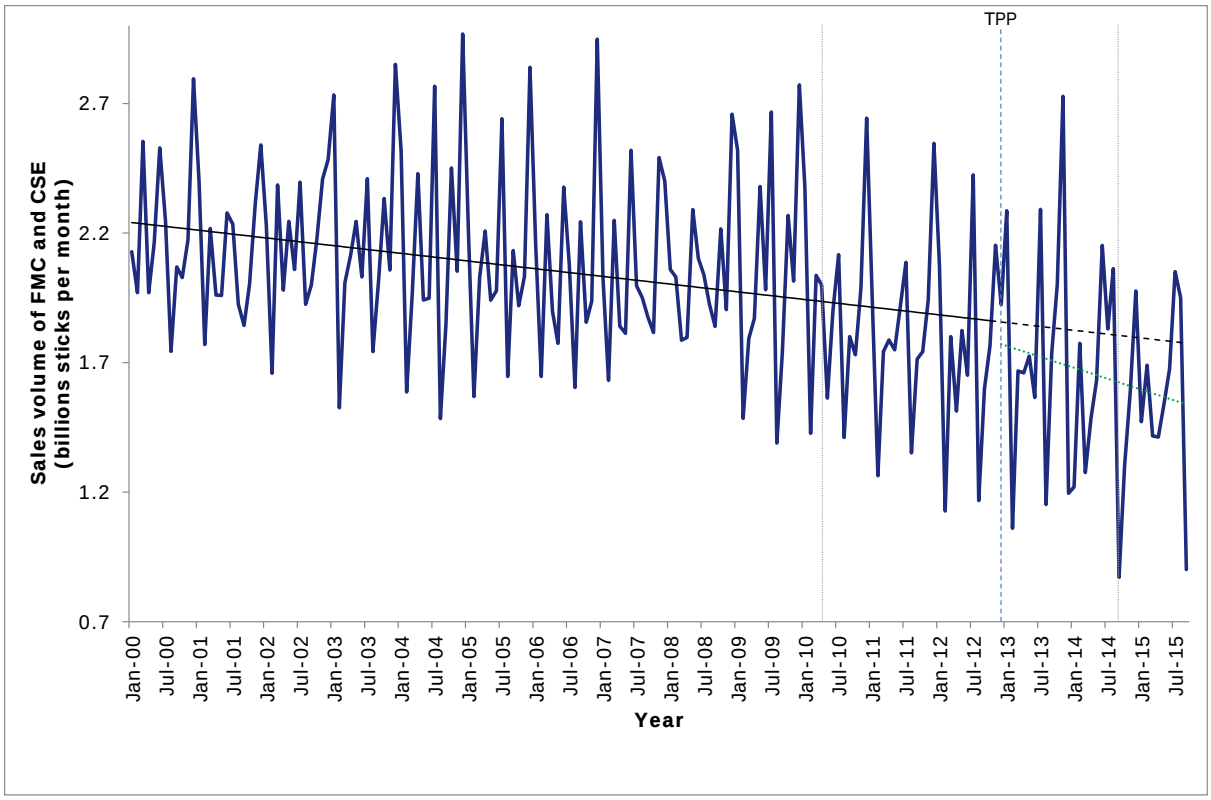

Note: Conversion rate of 0.8 grams of fine-cut tobacco per stick. The vertical dotted lines indicate excise tax hikes. The vertical dashed line indicates the introduction of TPP and enlarged GHWs.

Source: Chipty Third Rebuttal Report, (Exhibit AUS-605), para. 67.

46. That being said, we find that, according to a standard mean-comparison test, the average cigarette sales volumes based on the IMS/EOS data in the post-implementation period are statistically significantly lower than in the pre-implementation period.[50] This is confirmed by the fact that, as described in **Figure D.14**, the cigarette sales trend in the post-implementation period has become steeper compared to the pre-implementation trend, implying that the fall in cigarette sales has accelerated in the post-implementation period. The same conclusion can be drawn based on the Aztec data. In any case, as for prevalence, the fact that the downward trend in sales from manufacturers to wholesalers and retailers has accelerated in the post-implementation period does not necessarily mean that the TPP measures are having a statistically significant impact, given that other factors, unrelated to the TPP measures, could explain the trend shift. Evidence relating to the extent to which the TPP measures have an impact on cigarette sales or in some reports on cigarette consumption is reviewed next.

[50] For completeness sake, we note that the mean-comparison test applied to the Nielsen data concludes that there is no difference between the pre- and post-implementation period trends. However, the Nielsen data are only available for the February 2011-December 2013 period. When the mean-comparison test is applied to the IMS/EOS data for the same February 2011-December 2013 period, the test result suggests also that there is no shift in the pre- and post-trend, while the conclusion is reversed when the sample period is extended to September 2015. We, therefore, do not consider the result of the mean-comparison test applied to the Nielsen data to be relevant given that other more recent datasets are available and yield a different conclusion.

**Figure D.14: Cigarette Sales Volumes Pre- and Post-TPP Trends**




Note: Conversion rate of 0.8 grams of fine-cut tobacco per stick. The vertical dotted lines indicate excise tax hikes. The vertical dashed line indicates the introduction of TPP and enlarged GHWs.

Source: IPE Summary Report, (Exhibit DOM-379), back-up material.

## 3 WHETHER THE TPP MEASURES CONTRIBUTED TO THE REDUCTION IN CIGARETTE SALES VOLUMES AND CONSUMPTION

47. As explained above, we have determined that the evidence presented to us shows that cigarette sales volumes in Australia have continued to experience a decline, which has accelerated in the post-TPP period. To the extent that there has been a greater reduction in cigarette sales volumes after the entry into force of the TPP measures, the question arises whether, and if so, to what extent, the TPP measures contributed to reducing cigarette consumption.

48. As for smoking prevalence, IPE (for the Dominican Republic), Professor List (for the Dominican Republic and Indonesia), Professor Klick (for Honduras) and Dr Chipty (for Australia) proposed different econometric methods to estimate the TPP measures' impact on cigarette sales or reported cigarette consumption. As mentioned above, all parties recognize, however, that the empirical econometric results submitted do not distinguish between the impact of TPP and the impact of enlarged GHWs on cigarette sales and consumption, because both measures were implemented at the exact same time.[51]

49. Australia submits that, with regard to prevalence, it is too early to look at cigarette consumption to assess the contribution of the TPP measures, for two main reasons. First, reduction in cigarette consumption through smoking prevalence is a long-term objective.[52] Second,

[51] See Australia's first written submission, para. 518; Dominican Republic's response to Panel question No. 8, para. 61; List Report, (Exhibit DOM/IDN-1), para. 113; Honduras's response to Panel question No. 8; and Indonesia's response to Panel question No. 8, para. 8.

[52] See Australia's first written submission, para. 670. Australia considers that in the early stages of introduction of the measures, the most appropriate way to discern their effects is to rely on experiments and surveys which consider drivers of choice, attitudes, and ultimately, the elicitation of behavioural intentions (Australia's first written submission, para. 147).

significant changes in the root behaviours (i.e. initiation, cessation, and relapse) stemming from TPP are likely masked in cigarette consumption because of the stock of current smokers whose behaviours may not be as affected by TPP.[53] Instead Australia considered that the impact of TPP should be investigated through its three mechanisms, i.e. appeal, GHW effectiveness, and the packs' ability to mislead. Australia referred to a series of peer-reviewed studies published in the *Tobacco Control* journal. Notwithstanding this position, Australia engaged in estimating the TPP measures' impact on cigarette sales in response to the Dominican Republic's, Honduras's and Indonesia's submissions, and found on that basis that the TPP measures have already contributed to the reduction of cigarette sales volumes.[54]

50. Just as with the empirical results submitted in the context of prevalence, new methodologies or new model specifications or both were proposed by the experts of the Dominican Republic, Honduras and Indonesia in the course of the proceedings in response to the exchange of arguments between the parties. In some cases, the new models proposed invalidate some of the previous estimations (even though they yielded the same conclusion).[55] For this reason, the approaches and results discussed below are drawn mostly from the most recent expert reports submitted by the parties.[56] Similar to the review of the econometric studies on smoking prevalence, we first describe the relevant results and related discussions by dataset, before presenting our analysis.

## 3.1 Datasets and related studies

### 3.1.1 In-Market-Sales/Exchange of Sales

51. The IMS/EOS data were used by several of the Dominican Republic's and Honduras's experts to estimate econometrically the impact of the TPP measures on cigarette wholesale volumes.

#### 3.1.1.1 IPE Reports

52. The Dominican Republic submitted several reports prepared by IPE, estimating econometrically the TPP measures' impact on the logarithm of cigarette sales using the IMS/EOS data. Throughout the IPE Reports submitted, different econometric approaches and model specifications have been proposed: (1) time series regression analysis of cigarette sales; (2) modified trend analysis of per capita cigarette sales; and (3) ARIMAX model of per capita cigarette sales.[57] The first approach was only presented in the first IPE Report and is based on a model specification of the logarithm of cigarette sales that includes at least a dummy variable for the TPP measures, dummy variables for the 2010 and 2013 excise tax increase, month fixed effects and a trend (in some specifications). The second and subsequent IPE Reports only focused on the last two approaches, which to some extent were adopted to address some of the critiques raised by Australia's expert, Dr Chipty. Under these two approaches, the model specification was modified to focus on the logarithm of per capita cigarette sales and to include at least a TPP measures dummy variable, a tobacco price variable, a 2006 GHW dummy, dummy variables for strategic inventory and a linear trend. The most recent econometric analysis reported in the IPE Report covers the period January 2001 through August 2015 (for the model specifications

[53] See Chipty Report, (Exhibit AUS-17), paras. 32-39.

[54] See Australia's response to Panel question No. 196, para. 214.

[55] For instance, as in the review of the econometric studies on smoking prevalence, IPE initially proposed to control for excise tax increases by including (dummy) indicator variables for each excise tax increase (IPE Updated Report, (Exhibit DOM-303)), but subsequently contended that a more appropriate measure to capture the excise tax increases is the weighted average price per cigarette in Australia. IPE Second Updated Report, (Exhibit DOM-361). Similarly, Professor List and IPE initially applied the STATA software command ivreg2 in order to calculate standard errors that are robust to heteroscedasticity and serial correlation using the automatic bandwidth selection procedure by Newey and West (1994). IPE Updated Report, (Exhibit DOM-303); and List Report, (Exhibit DOM/IDN-1). Subsequently, both Professor List and IPE applied an alternative way of calculating standard errors, that, according to them, is adjusted to reflect more accurately the original suggestion by Newey and West (1994). IPE Second Updated Report (Exhibit DOM-361); and List Rebuttal Report, (Exhibit DOM/IDN-3).

[56] We note that we nonetheless considered all the relevant evidence before us, including all the expert reports, including the methodologies and models contained therein.

[57] See IPE Report, (Exhibit DOM-100), pp. 67-73 and 201-210; and IPE Updated Report, (Exhibit DOM-303), paras. 135-175 and 290-326.

without the price variable) or February 2002 through August 2015 (for the model specifications with the price variable).[58]

53. Overall, IPE concludes that the TPP measures had no statistically significant effect on cigarette sales and per capita cigarette sales. According to IPE, these results are robust across different specifications (e.g. different TPP starting date: October, November and December; excluding the October and November 2012 observations; controlling for the weighted average price of tobacco, tax levels or tax policy change; and strategic inventory management). The results are also robust to Professor List's procedure to compute the explanatory variables' robust standard errors, which is different from the STATA software package ivreg2 used by Dr Chipty.[59] However, when excise tax dummy variables are included in the model specification, some of the results of the modified trend analysis suggest that the TPP measures had a negative and statistically significant impact on cigarette wholesales and per capita cigarette wholesales.[60]

54. Australia's expert, Dr Chipty, rejects the econometric results of the IPE Reports on various technical grounds. Dr Chipty is of the view that the analysis should be extended to the 2000-2015 period, instead of restricting the sample to the period 2001-2015 or 2002-2015, noting that there is no reason for ignoring available data.[61] For similar reasons as for the IPE Reports and Professor List's econometric studies on smoking prevalence, Dr Chipty disagrees with Professor List and IPE's claim that the Stata command, ivreg2, used to estimate standard errors robust to "heteroscedasticity" and "autocorrelation" (Newey-West standard errors) is wrong, noting that both IPE and Professor List used the same Stata command, before Professor List argued he found an error in the Stata programming code.[62] Similarly, Dr Chipty submits that controlling for tobacco prices inclusive of tax while attempting to measure the effects of a tax hike ignores the effect of the TPP measures on price, leading the TPP measures indicator variable to capture only a partial effect of the TPP measures, and the price variable capturing the rest of its effect. Dr Chipty further notes that IPE initially controlled for excise tax increase with indicator variables in their model specification.[63] Dr Chipty further submits that the use of a single tax level variable, as proposed by IPE, is only valid under the proportionality assumption (i.e. the effect of tax changes is proportional to the size of the tax change), which may only be satisfied in some specifications, and is inappropriate in models of consumption/sales.[64] Dr Chipty argues that IPE's consumption analyses rely heavily on the use of time trends by assuming trends in consumption exist in a vacuum without addressing the issue of whether any of the changes in consumption with the time trends around the TPP measures should themselves be interpreted as the TPP measures' effects.[65]

#### 3.1.1.2 Professor Klick's reports

55. Like his analysis of smoking prevalence, Honduras's expert, Professor Klick, proposes comparing Australia's cigarette sales before and after the introduction of the TPP measures relative with another jurisdiction that has not implemented TPP during the same period. Professor Klick posits that although Australia is unique in many ways, it is, notwithstanding Dr Chipty's criticisms, reasonable to use New Zealand as a reliable counterfactual jurisdiction, because (i) both countries share many similarities culturally, historically, and demographically; (ii) both countries are in the same region and share the same seasonality (useful when dealing with sub-annual measures); (iii) the governments of both countries acknowledge that they are comparable in areas such as health behaviour and socio-economic issues; (iv) tobacco prices, including taxes, are comparable between both countries; and (v) smoking rates in both countries are highly correlated (0.95 for

[58] The difference between the two sample periods stem from the fact that the price variable is not available prior to February 2002.

[59] IPE Report, (Exhibit DOM-100), pp. 67-73 and 201-210; IPE Updated Report, (Exhibit DOM-303), paras. 143-155, 290-295; IPE Second Updated Report, (Exhibit DOM-361), paras. 26-31; IPE Third Updated Report, (Exhibit DOM-375), paras. 117-145; and IPE Summary Report, (Exhibit DOM-379), paras. 123-150.

[60] See IPE Summary Report, (Exhibit DOM-379), paras. 144-145 and backup material.

[61] See Chipty Surrebuttal Report, (Exhibit AUS-586), para. 9; and Chipty Third Rebuttal Report, (Exhibit AUS-605), paras. 34-35.

[62] See Chipty Second Rebuttal Report, (Exhibit AUS-591), paras. 13-31.

[63] See Chipty Surrebuttal Report, (Exhibit AUS-586), paras. 12, 45 and 65.

[64] See Chipty Third Rebuttal Report, (Exhibit AUS-605), paras. 22-26 and 29-32.

[65] See Chipty Surrebuttal Report, (Exhibit AUS-586), paras. 68-70.

the period 1964-2012 based on the OECD data on the fraction of residents ages 15 years and older who are daily smokers).[66]

56. Professor Klick reconsidered the IMS/EOS data for Australia and New Zealand and estimated a difference-in-difference model that explains the logarithm of cigarette sales in Australia relative to New Zealand controlling for a TPP measures dummy variable, the logarithm of tobacco price, country fixed effects, time-period fixed effects and country-specific trends. In Professor Klick's first submission, the model only controlled for a TPP measures dummy variable, time-period fixed effects and differential intercept term for New Zealand. Professor Klick subsequently modified his model specification to address some of Dr Chipty's criticisms. In Professor Klick's first report, a differential trends model was also estimated controlling for a TPP measures dummy variable, an Australia monthly trend, and a New Zealand monthly trend. The most recent econometric analysis covers the period February 2011 to September 2015.[67]

57. Professor Klick considers it unlikely that prices may be endogenous to cigarette sales, in other words prices may be determined by cigarette sales, a concern raised by Dr Chipty. This is because tobacco taxes make up the bulk of tobacco prices, and are, theoretically, less likely to be determined endogenously to cigarette sales, at least on a month-to-month basis, since taxes are the result of political decisions that take significant time to reach and implement. Nevertheless, Professor Klick implements an instrument variable (IV) estimation procedure to address the theoretical possibility of an endogeneity bias of the tobacco price variable. The IV estimation procedure involves (1) estimating a model of tobacco price that includes a TPP measures dummy variable, an excise tax dummy variable, country fixed effects and time fixed effects (and country-specific trends in some specifications); and (2) reconstructing the tobacco price variable using the estimated parameters of the model obtained in the first stage. This constructed (instrumented) price variable replaces the original price variable in the model of cigarette sales.[68] The overall impact of the TPP measures corresponds to the sum of the impact of the TPP measures on cigarette sales and the impact of the TPP measures on cigarette price multiplied by the impact of cigarette price on cigarette sales. In addition, Professor Klick contends that New Zealand's 2013 excise tax increase does not invalidate it as a comparator to Australia, because taxes influence behaviour through prices and the model specification controls for price and accounts for its potential endogeneity through the instrumental variables technique.[69]

58. Overall, Professor Klick concludes that the TPP measures have not resulted in a decline in cigarette sales in Australia relative to sales in New Zealand.[70] However, Professor Klick's most recent estimations find a positive and statistically significant impact of the TPP measures on cigarette sales, suggesting that the TPP measures led to an increase in cigarette sales in Australia relative to New Zealand.[71]

59. Dr Chipty contends that Professor Klick's analysis is flawed on various technical grounds. As in the case of Professor Klick's econometric analysis on smoking prevalence, Dr Chipty is of the view that New Zealand is an invalid counterfactual for the purposes of studying the TPP measures' effects, because of New Zealand's January 2013 excise tax increase, a month after the introduction of the TPP measures.[72] In particular, Dr Chipty submits that there is no reliable way to correct for the failure to account for differential country-specific tax policies introduced at different times in Australia and New Zealand because there is only a single month of data in New Zealand between implementation of the TPP measures in Australia and the January 2013 New Zealand

[66] See Klick Report, (Exhibit UKR-5), pp. 3-6; and Klick Third Supplemental Rebuttal Report, (Exhibit HND-166), paras. 30-31.

[67] See Klick Report, (Exhibit UKR-5), pp. 35-36; Klick Rebuttal Report, (Exhibit HND-118), paras. 66-68; and Klick Supplemental Rebuttal Report, (Exhibit HND-122), paras. 107-121.

[68] See Klick Rebuttal Report, (Exhibit HND-118), paras. 52-60.

[69] See Klick Third Supplemental Rebuttal Report, (Exhibit HND-166), para. 32.

[70] See Klick Supplemental Rebuttal Report, (Exhibit HND-122), paras. 117-119; and Klick Third Supplemental Rebuttal Report, (Exhibit HND-166), paras. 44-45.

[71] See Klick Third Supplemental Rebuttal Report, (Exhibit HND-166), paras. 44-45.

[72] See Chipty Supplementary Report, (Exhibit AUS-511), paras. 16-19; and Chipty Rebuttal Report, (Exhibit AUS-535) (SCI), paras. 67-71.

excise tax increase.[73] Dr Chipty also questions the use of a highly flexible time trend that makes it difficult to estimate the effects of the very policies that likely created the trend.[74]

#### 3.1.1.3 Dr Chipty's reports

60. Australia's expert, Dr Chipty, considers the IMS/EOS data to be the only appropriate data to assess the TPP measures' impact on cigarette sales volumes, because the IMS/EOS data capture virtually all sales in Australian market, given that the big three manufacturers account for all but 1% of total sales. In addition, the IMS/EOS data are available for a longer period.[75]

61. Dr Chipty re-estimated the model of cigarette sales developed by IPE in its first report (time series regression analysis) by extending the period of analysis from January 2000 to September 2015 and controlling for a TPP measures dummy variables, tax increases dummy variables (for May 2010, December 2013 and September 2014), a 2006 GHW dummy variable, a linear trend and month fixed effects.[76]

62. Overall, Dr Chipty concludes that the TPP measures had a negative and statistically significant effect on cigarette sales.[77] Most of the results show statistically significant declines in cigarette sales even when using Professor List's own standard error calculation.[78] However, Dr Chipty notes that the results are sensitive to the decision whether to use a tobacco price variable or the tax levels variable instead of the tax indicators, and that specification testing suggests the use of tax levels are not appropriate.[79]

63. IPE counters that Dr Chipty "cherry-picked" a small subset of IPE's analyses for the purposes of undertaking adaptations or "corrections". According to IPE, Dr Chipty engaged in further "cherry-picking", reporting results from only a small subset of specifications, for example by using only one specific measure of tobacco affordability (tax dummies) and simply ignoring the two superior alternatives (tax levels and tobacco prices, respectively). IPE submits that Dr Chipty never even discussed any of the ARIMAX models, thus completely ignoring an entire class of models.[80]

### 3.1.2 Nielsen Retail Sales

64. Similar to the IMS/EOS data, the Nielsen retail sales data were used by the Dominican Republic's and Honduras's experts to estimate econometrically the impact of the TPP measures on cigarette retail volumes.

#### 3.1.2.1 IPE Reports

65. IPE argues that Dr Chipty's decision to dismiss the Nielsen data due to its short sample period is unscientific. According to IPE, despite the relatively shorter period for which the Nielsen data are available, the dataset has two important advantages as compared to the IMS/EOS data and the Aztec data: (1) compared to the IMS/EOS data, the Nielsen dataset provides actual sales to customers and is not affected by wholesalers' strategic inventory management behaviour; and

[73] See Chipty Supplementary Report, (Exhibit AUS-511), paras. 3 and 12-19. Dr Chipty also argued that Professor Klick failed to account for the strategic inventory management associated with excise tax hikes but acknowledged that the failure to account for the buy up problem did not affect the estimates in Professor Klick's analysis. Ibid. paras. 22-24.

[74] See Chipty Supplementary Report, (Exhibit AUS-511), para. 25.

[75] See Chipty Third Rebuttal Report, (Exhibit 605), para. 70.

[76] See Chipty Second Rebuttal Report, (Exhibit AUS-591), paras. 28-31.

[77] See Chipty Surrebuttal Report, (Exhibit AUS-586), paras. 65-66; Chipty Second Rebuttal Report, (Exhibit AUS-591), paras. 29-31, 53-54, Appendix D; and Chipty Third Rebuttal Report, (Exhibit AUS-605), paras. 74-75.

[78] See Chipty Second Rebuttal Report, (Exhibit AUS-591), paras. 28-31.

[79] See Chipty Third Rebuttal Report, (Exhibit 605), paras. 15-28 and 75.

[80] See IPE Summary Report, (Exhibit DOM-379), paras. 156-161.

(2) compared to the Aztec data, which are also at the retail level, the Nielsen data cover virtually the whole market.[81]

66. Similar to the analysis of the IMS/EOS and the Aztec data, IPE presents the results of the (1) modified trend analysis and (2) ARIMAX model of tobacco retail sales per capita controlling for at least a TPP measures dummy variable, a cigarette price variable and a linear trend.[82] The econometric analysis based on the Nielsen data covers the period February 2011 to December 2013.[83]

67. Overall, IPE concludes that the TPP measures had no statistically significant effect on retail sales of cigarettes per capita.[84] According to IPE, these results are robust across different specifications (e.g. different TPP starting date: October, November and December; controlling for the weighted average price of cigarettes, tax levels or tax policy change; and adjusting for seasonality in order to reduce the number of parameters to be estimated). In reality, some of IPE's results of the modified trend analysis (when the TPP measures dummy variable starts in December 2012) suggest that the TPP measures led to an increase in retail sales of cigarettes per capita.[85]

68. Dr Chipty agrees with IPE that the Nielsen data cover "virtually the whole market" and as such are superior to the Aztec data in this respect. However, Dr Chipty contends that the sample size of the Nielsen data is a serious problem for IPE, which attempts to estimate at least 15 different coefficients, including 12-month indicators for seasonality, using 35 monthly observations. Dr Chipty argues that in IPE's Nielsen analysis, the monthly effects are identified off of variation from three data points each, with only two for the January effect. Dr Chipty submits that unlike the IMS/EOS data, which are affected by the issue of strategic inventory management, but can be corrected for it, there is no workaround for the insufficient data issue.[86]

#### 3.1.2.2 Professor Klick's reports

69. Professor Klick also used the Nielsen retail sales data for Australia and New Zealand to estimate a difference-in-difference model of the logarithm of cigarette retail sales per capita controlling for a TPP measures dummy variable, the logarithm of tobacco price, country fixed effects, time-period fixed effects and country-specific trends. In Professor Klick's first submission, the analysis focused on cigarette retail sales and not cigarette retail sales per capita. Professor Klick's first difference-in-difference model also only controlled for a TPP measures dummy variable, time-period fixed effects and differential intercept term for New Zealand. In addition, a differential trends model was also estimated controlling for a TPP measures dummy variable, an Australia monthly trend, and a New Zealand monthly trend. The econometric analysis covers the period February 2011 to December 2013.[87]

70. Overall, Professor Klick concludes that the TPP measures are not associated with a statistically significant reduction in cigarette retail sales and cigarette retail sales per capita in Australia relative to New Zealand.[88] However, two initial specifications, which do not control for price, find a

[81] See IPE Updated Report, (Exhibit DOM-303), para. 157; and IPE Third Updated Report, (Exhibit DOM-375), paras. 158-168.

[82] See IPE Report, (Exhibit DOM-100), pp. 67-73 and 201-210; and IPE Updated Report, (Exhibit DOM-303), paras. 135-175 and 290-326.

[83] See IPE Updated Report, (Exhibit DOM-303), para. 156.

[84] See IPE Updated Report, (Exhibit DOM-303), paras. 156-162; IPE Second Updated Report, (Exhibit DOM-361), paras. 32-35; IPE Third Updated Report, (Exhibit DOM-375), paras. 158-168; and IPE Summary Report, (Exhibit DOM-379), paras. 146-147.

[85] See IPE Third Updated Report, (Exhibit DOM-375), paras. 146-147 and 158-168.

[86] See Chipty Surrebuttal Report, (Exhibit AUS-586), para. 77.

[87] See Klick Report, (Exhibit UKR-5), pp. 12-16; and Klick Rebuttal Report, (Exhibit HND-118), paras. 10, 39-42, 44-48 and 49-51.

[88] See Klick Report, (Exhibit UKR-5), pp. 32-34; and Klick Rebuttal Report, (Exhibit HND-118), paras. 54-64.

positive and statistically significant TPP effect, suggesting that the TPP measures led to an increase in cigarette sales in Australia relative to New Zealand.[89]

71. Professor Klick also presents the results of a "crossfold validation" analysis to test various predictive models using the Nielsen data in terms of predictive accuracy. Professor Klick concludes that the January 2013 tax increase in New Zealand does not disqualify New Zealand as a proper comparator because the best performing model is the one that includes New Zealand's natural log per capita sales and monthly fixed effects and for which the TPP measures' impact on per capita cigarette sales in Australia (relative to New Zealand) is statistically not significant.[90]

72. As described above, Dr Chipty is of the view that New Zealand is an invalid counterfactual for the purposes of studying the TPP measures' effects, because of New Zealand's January 2013 excise tax increase, a month after the introduction of the TPP measures.[91] Dr Chipty further submits that the predictive ability measured by the "crossfold validation" is insufficient to validate a control group given New Zealand's January 2013 tax increase.[92]

### 3.1.3 Aztec Scanner Retail Sales

73. Similar to the analysis of the IMS/EOS data, IPE used the Aztec data and applied different econometric approaches and model specification throughout their reports: (1) time series regression analysis of cigarette retail sales; (2) modified trend analysis of cigarette retail sales per capita; and (3) ARIMAX model of cigarette retail sales per capita.[93] The first approach was only presented in the first IPE Report and is based on a model specification of the logarithm of retail sales of cigarettes that includes at least a TPP measures dummy variable, dummies for the 2010 and 2013 excise tax increase, month fixed effects and a trend. The second and subsequent IPE Reports only focused on the last two approaches, which to some extent were adopted to address some of the criticisms raised by Australia's expert, Dr Chipty. In those two approaches, the model specification was modified to focus on the logarithm of retail sales of cigarettes *per capita* and to include at least a TPP measures dummy variable, a cigarette price variable and a linear trend. IPE argues that Dr Chipty's decision to dismiss the Aztec dataset due to its growing market coverage is unscientific because the use of linear time trends means it is possible to take into account changes in market coverage.[94] The most recent econometric analysis reported in the IPE Report covers the period from 27 July 2008 through 27 September 2015.[95]

74. Overall, IPE concludes that the TPP measures had no statistically significant effect on retail sales of cigarettes per capita.[96] According to IPE, these results are robust across different specifications (e.g. different TPP starting date: October, November and December; controlling for the weighted average price of cigarettes, tax levels or tax policy change; and adjusting for the increasing market share by benchmarking quarterly Aztec data to IMS/EOS data). In reality, when IPE's modified trend analysis specification controls for tax levels (instead of cigarette price or tax dummies), the results suggests that the TPP measures led to an increase in retail sales of cigarettes. Similarly, in a few ARIMAX specifications, the TPP measures' impact on retail sales of cigarettes is positive and statistically significant.

75. As explained above, Dr Chipty argues that not all data are worth studying. She notes that the Aztec scanner data in particular has a growing coverage of retail sales over time, either through the growth of the sales outlets it covers or through the inclusion of additional sales outlets, which

[89] However, Dr Chipty notes that Professor Klick made a transcription error and erroneously reported a statistically significant increase in sale in Professor Klick's two specifications controlling for price. See Chipty Report, (Exhibit AUS-17), para. 53.

[90] See Klick Rebuttal Report, (Exhibit HND-118), paras. 44-51; and Klick Third Supplemental Rebuttal Report, (Exhibit HND-166), para. 32.

[91] See Chipty Rebuttal Report, (Exhibit AUS-535) (SCI), paras. 67-71.

[92] See Chipty Rebuttal Report, (Exhibit AUS-535) (SCI), paras. 70-71.

[93] See IPE Report, (Exhibit DOM-100), pp. 67-73 and 201-210; and IPE Updated Report, (Exhibit DOM-303), paras. 135-175 and 290-326.

[94] See IPE Third Updated Report, (Exhibit DOM-375), paras. 146-147, 156-157 and 169-170.

[95] See IPE Third Updated Report, (Exhibit DOM-375), paras. 240-242.

[96] See IPE Updated Report, (Exhibit DOM-303), paras. 163-172; IPE Second Updated Report, (Exhibit DOM-361), paras. 26-31, 36-38; IPE Third Updated Report, (Exhibit DOM-375), paras. 169-175, 240-242; and IPE Summary Report, (Exhibit DOM-379), paras. 148-150.

could lead to the wrong conclusion that the TPP measures increased cigarette sales even when it did not.[97] Dr Chipty contends that, contrary to IPE's claim that the inclusion of a time trend mitigates the risk of bias, the inclusion of a time trend does not eliminate the risk of bias for at least two reasons: (1) the inclusion of a simple time trend in IPE's model may mitigate some of the problem, but there is no reason to expect that it will eliminate it, because the trend is likely not flexible enough to adequately capture the variety of different business explanations that result in the Aztec data's growing footprint; and (2) more importantly, the simple time trend in IPE's model cannot resolve the fact that the Aztec data are missing information on total sales because they only contain information from 12 retail chains and do not contain information on certain retail channels of tobacco distribution.[98]

### 3.1.4 Nielsen Retail Sales-Aztec Scanner

76. Professor Klick considers that, contrary to Dr Chipty's claim, the Aztec scanner data are valid because there is no evidence that the market share covered by the Aztec data has been growing over time.[99] In that context, Professor Klick combines the Nielsen and Aztec data presented above by omitting the Convenience Independent sample from the Nielsen data, and for the months where the data sources overlap, using the Aztec data.[100] Professor Klick re-estimated the difference-in-difference model of retail sales of cigarettes in Australia relative to New Zealand. In addition to the difference-in-difference analysis of Australia and New Zealand, Professor Klick also estimated a model without New Zealand as a counterfactual control. The model specification controls for a TPP measures dummy variable, the logarithm of tobacco price, the logarithm of the number of stores covered by the Aztec data and a linear trend. The tobacco price variable is instrumented to address any endogeneity bias.[101] In addition, to control for the Aztec data's alleged growing market coverage, a variable reporting the number of stores covered by the Aztec data is included in Professor Klick's most recent model specification. The econometric analysis covers the period February 2011 to September 2015 or is restricted to the January 2012-September 2015 period because the data relating to the number of stores are only available beginning January 2012.

77. Overall, Professor Klick concludes that the TPP measures had no statistically significant effect on the retail sales of cigarettes.[102] However, one of Professor Klick's two difference-in-difference estimations finds a positive and statistically significant TPP effect suggesting that the TPP measures led to an increase in retail sales of cigarettes. Similarly, the results of the model specification without New Zealand as a counterfactual and without a trend variable suggest that the TPP measures led to an increase in the retail sales of cigarettes. Professor Klick explains that he does not view these results as suggesting that smoking has increased under the TPP measures, because of the "multiple comparison problem", which implies that when very many outcomes are examined, there is a relatively high likelihood that one will find statistically significant results even by mere chance.[103]

78. Dr Chipty argues that Professor Klick's analysis based on the Nielsen-Aztec data is flawed because the Nielsen-Aztec data appear to be covering a greater share of total market sales over time. According to Dr Chipty, one reason for the growing share of Professor Klick's Nielsen-Aztec data over time in Australia is the exclusion of downward-trending sales from the Convenience Independent channel. Dr Chipty submits that the IMS/EOS data are decreasing much faster than Professor Klick's constructed Nielsen-Aztec data, suggesting that the Nielsen-Aztec data are not an

[97] See Chipty Report, (Exhibit AUS-17), paras. 55-60; and Chipty Surrebuttal Report, (Exhibit AUS-586), para. 10.

[98] See Chipty Surrebuttal Report, (Exhibit AUS-586), paras. 75-76.

[99] See Klick Fourth Supplemental Rebuttal Report, (Exhibit HND-169), paras. 6-8.

[100] Professor Klick explains that beginning in February 2014, the scanner data are no longer provided by the Nielsen Company. The Nielsen Company also did not continue its Convenience Independent sample from January 2014 onward. However, scanner data, without the Convenience Independent sample, is available from Aztec beginning in March 2012 and continuing through June 2015. See Klick Supplemental Rebuttal Report, (Exhibit HND-122), paras. 122-126.

[101] See Klick Fourth Supplemental Rebuttal Report, (Exhibit HND-169), paras. 9-11.

[102] See Klick Supplemental Rebuttal Report, (Exhibit HND-122), paras. 129-130; Klick Third Supplemental Rebuttal Report, (Exhibit HND-166), paras. 11-23; and Klick Fourth Supplemental Rebuttal Report, (Exhibit HND-169), paras. 12-14.

[103] See Klick Fourth Supplemental Rebuttal Report, (Exhibit HND-169), fn 10.

adequate proxy for total cigarette sales in Australia. In addition, Dr Chipty contends that Professor Klick's Nielsen-Aztec model does not allow for the possibility of estimating the net effect of the TPP measures because the regression model controls for price, which itself embodies some of the policy effect.[104]

### 3.1.5 Commissioned Roy Morgan Research Survey (Australia and New Zealand)

79. As explained in the section discussing smoking prevalence and smoking incidence, the Roy Morgan Research dataset is a survey of individuals who were current or former (in the past 12 months) smokers in both Australia and New Zealand undertaken using random digit dialling sampling techniques. The first wave of the survey was completed prior to the December 2012 implementation of the TPP measures between 2 November 2012 and 26 November 2012 in Australia and between 8 November 2012 and 1 December 2012 in New Zealand. Subsequent waves have been carried out at three-month intervals.

80. Professor Klick used the Roy Morgan Research data and estimated a difference-in-difference linear regression model and negative binomial models that explain the reported number of cigarettes consumed controlling for a TPP measures dummy variable, a dummy variable for an Australian baseline and a dummy variable for a post-TPP implementation time-period. The difference-in-difference analysis covers the period from November 2012 to February 2014.[105]

81. Overall, Professor Klick concludes that the TPP measures had no statistically significant effect on the cigarette consumption in Australia relative to New Zealand.[106] According to Professor Klick, this finding is robust to specifying the number of cigarettes smoked in logarithm (to mitigate outliers), applying different estimators (linear regression, negative binomial regression, robust regression), restricting the sample to only individuals who answered the survey in the 6 waves (to mitigate attrition), restricting the sample to only individuals who answered the survey at least in one post-TPP waves, restricting sample to Wave 1 and Wave 6, restricting sample to smokers planning to quit during Wave 1, and computing robust standard errors or standard errors clustered by individuals.[107] However, some of the results (when respondents who had noticed plain packs in Wave 1 or a pack change of any kind in Wave 1 are discarded) suggest that the TPP measures led to an increase in cigarettes consumption. Professor Klick explains that he does not view these results as suggesting that smoking has increased under the TPP measures, because of the "multiple comparison problem", which implies that when very many outcomes are examined, there is a relatively high likelihood that one will find statistically significant results even by mere chance.[108]

82. Dr Chipty submits that Professor Klick's difference-in-difference study does not provide a reliable estimate of the TPP measures' effect in Australia, because one cannot reliably implement a difference-in-difference estimation strategy without: (1) a pre-period data, yet a majority of smokers were likely smoking plain packs during the survey's first wave; (2) a control group that resembles the treatment group (i.e. Australia) in important dimensions other than the treatment, yet New Zealand had a unique and significant tax change between Waves 1 and 2.[109] Dr Chipty further contends that Professor Klick invalidated his own commissioned survey data analysis by arguing that (1) it makes little sense to talk about trends in very short time spans; (2) examining differential trends with five or six data points is demanding more from the data than is reasonable; (3) wave-to-wave variability is quite large in the data; (4) Dr Chipty's analysis is inappropriate because it involves estimating a trend using data from a single calendar year; and (5) no meaningful analysis can be conducted by seeking a time trend between Waves 2 and 6.[110]

---

[104] See Chipty Surrebuttal Report, (Exhibit AUS-586), paras. 92-96.
[105] See Klick Report, (Exhibit UKR-5), pp. 20-21 and 30-32.
[106] See Klick Report, (Exhibit UKR-5), pp. 30-32.
[107] See Klick Rebuttal Report, (Exhibit HND-118), paras. 25-28, 31-33 and 37-38.
[108] See Klick Rebuttal Report, (Exhibit HND-118), fn 24.
[109] See Chipty Report, (Exhibit AUS-17), para. 24; and Chipty Rebuttal Report, (Exhibit AUS-535) (SCI), paras. 54-71. Dr Chipty initially contended that Professor Klick did not use clustered standard errors except in his fixed-effects models. Dr Chipty further argued that Professor Klick did not allow for heteroscedasticity in the error structure and account for the longitudinal nature of the data by clustering at the respondent level. Ibid. Appendix C.
[110] See Chipty Rebuttal Report, (Exhibit AUS-535) (SCI), paras. 50-53.

### 3.1.6 International Tobacco Control Policy Evaluation Project

83. The International Tobacco Control (ITC) Policy Evaluation Project (ITC Project) is a longitudinal cohort survey to assess the impact, and identify the determinants of, effective tobacco control policies in more than 20 countries, including Australia. The ITC Project covers a number of questions related to GHWs, including attention towards the health warnings, salience of the health warnings and the effect of health warnings on consumers' thoughts, behaviours and intentions towards quitting.

84. Ajzen et al., in their expert report submitted by the Dominican Republic and Indonesia, used the ITC dataset to estimate a generalized estimating equation model of the reported number of cigarettes smoked per day controlling for the survey mode (phone vs. web) and wave of recruitment, as well as respondents' age, sex, income, education and past year quit attempts.[111] The econometric analysis covers the period from September 2011 to May 2013.[112]

85. Overall, Ajzen et al. conclude that TPP and enlarged GHWs had no impact on the average number of cigarettes smoked per day in the first five months following the introduction of the TPP measures.[113] The same result is found when the sample is restricted to the participants involved in the surveys pre- and post-implementation.[114]

86. Australia responds that Ajzen et al. have misunderstood certain features of the ITC data and that their claim that there have been no changes in smoking behaviour post-TPP is without foundation.[115]

### 3.1.7 National Tobacco Plain Packaging Tracking Survey

87. To track the effect of the TPP measures, Australia funded the National Tobacco Plain Packaging Tracking Survey (NTPPTS), a nationwide tracking survey that was conducted by CCV. The NTPPTS is a continuous cross-sectional baseline survey of about 100 interviews per week conducted from 9 April 2012 to 30 March 2014. A follow-up survey of baseline participants then took place approximately four weeks after the initial survey, with the follow-up surveys conducted from 7 May 2012 to 4 May 2014.

88. The NTPPTS were used by scholars as well as several of the experts of the Dominican Republic, Honduras and Indonesia to estimate econometrically the impact of the TPP measures on reported cigarette consumption.

#### 3.1.7.1 Scollo et al. 2015a peer-reviewed study

89. Scollo et al. 2015a[116] used the NTPPTS data to assess changes in reported price paid and changes in reported numbers of cigarettes consumed following the introduction of TPP and enlarged GHWs in the period up to and after the large increase in excise duty on 1 December 2013. The authors estimated logistic and linear models to assess changes between the pre-TPP period (April to September 2012) and three subsequent time periods: the transition phase during which plain packages were being introduced into the Australian market (October and November 2012); TPP year 1 (December 2012 to November 2013); and TPP post-tax (December 2013 to March 2014). All regression models control for sex, age, area socio-economic status, education, past three-month exposure to antismoking campaigns (for cigarette consumption analysis).

[111] We note that Ajzen et al. used the ITC dataset to replicate part of the analysis of Yong et al. 2015, who did not analyse cigarettes consumption. Ajzen et al. argue that Yong et al. 2015 failed to explain why they did not report the impact of the pack changes on cigarette consumption. Ajzen et al. Second Data Report, (Exhibit DOM/IDN-4), para. 34.

[112] See Ajzen et al. Second Data Report, (Exhibit DOM/IDN-4), paras. 21-29 and 48-58.

[113] See Ajzen et al. Second Data Report, (Exhibit DOM/IDN-4), para. 28.

[114] See Ajzen et al. Data Rebuttal Report, (Exhibit DOM/IDN-6), para. 83.

[115] See Australia's response to Panel question No. 196, paras. 225 and 235.

[116] Scollo et al. 2015a, (Exhibits HND-133, DOM-237, DOM-311).

90. Overall, Scollo et al. 2015a conclude, among others things, that the introduction of TPP and enlarged GHWs were not associated with a change in consumption among daily, regular or current smokers or among smokers of brands in any market segment during the first year of implementation of the TPP measures. However, the authors find that reported consumption among regular smokers declined significantly following the December 2013 tax increase.[117]

#### 3.1.7.2 Ajzen et al.'s reports

91. Ajzen et al. re-analysed part of Scollo et al. 2015a[118], applying the approach in Wakefield et al. 2015.[119] Ajzen et al. estimated logistic, ordered logistic and linear regressions of the reported number of cigarettes smoked daily and the proportion of smokers consuming a given number of cigarettes or more per day. The model specification includes a set of variables controlling for age group, gender, education group, socio-economic status group, potential exposure to televised antismoking advertising campaigns in the past three months and cigarette costliness.[120]

92. Overall, Ajzen et al. conclude that the TPP measures had no statistically significant impact on any measure of actual cigarette consumption.[121] Ajzen et al. argue that, unlike Professor Chaloupka's assertion, using a threshold of 20 or more cigarettes per day is not inappropriate because the results continue to find no change in cigarette consumption when thresholds of 5, 10, and 15 cigarettes per day are used.[122] In addition, Ajzen et al. present the quarterly results of the logit model, which suggests that there was an immediate increase in the probability of smoking 20 or more cigarettes per day in the first quarter following the implementation of the TPP measures but no impact in the following quarters.[123] More generally, Ajzen et al. disagree with Professor Chaloupka's assertion that the NTPPTS dataset has less power for detecting statistically significant changes in the more distal outcomes, such as actual tobacco use behaviour. According to Ajzen et al., a power analysis of the NTPPTS data concludes that the statistical power of the NTPPTS data to detect small changes in proximal and distal outcomes is very similar.[124]

93. Australia's expert, Professor Chaloupka, argues that Ajzen et al.'s analyses of the NTPPTS data fail to recognize that the TPP measures' impact should be smaller for the less proximal outcomes when one looks at the impact in the overall sample of smokers and recent quitters, given that smokers and recent quitters for whom the most proximal outcomes were unaffected would not be expected to show any change in the less proximal outcomes.[125] Professor Chaloupka further contends that using an insensitive measure of consumption, whether or not the smoker reports consuming 20 or more cigarettes per day, appears to be an inappropriate threshold given that the average daily consumption of smokers in Australia is well below this level. According to

[117] Scollo et al. 2015a also find that the introduction of TPP and enlarged GHWs was associated with an increase in use of value brands and smaller increases in prices for value relative to premium brands.

[118] Ajzen et al. argue that Scollo et al.'s conclusion that TPP and enlarged GHWs had no significant effect on the average number of cigarettes consumed per day underreports additional results that were not statistically significant. According to Ajzen et al., Scollo et al. 2015a reported on only one of the seven measures on smoking behaviours and the impact of TPP and enlarged GHWs on each of the six unreported variables, dealing with the percentage of daily or weekly smokers, quitting, and relapse was not statistically significant. See Ajzen et al. Data Report, (Exhibit DOM/IDN-2), paras. 225-227. We note, however, that Ajzen et al. do not reassess the analysis by Scollo et al. 2015a of the impact of TPP and enlarged GHWs and tax increase on type of tobacco products and price.

[119] Wakefield et al. 2015, (Exhibits AUS-206, DOM-306).

[120] We note that Ajzen et al. also re-estimate their models by replacing the TPP measures dummies with a monthly time trend.

[121] Ajzen et al. reach the same conclusion with respect to measures related to quitting and relapse, which were unreported by Scollo et al. 2015a. In particular, on quitting, Ajzen et al. conclude that there was no statistical change in the proportion of adult smokers that quit for more than one month or had successfully quit between baseline and follow-up. On relapse, Ajzen et al. conclude that there was no statistical change in the proportion of adult ex-smokers that relapsed, still abstained from smoking at follow-up or had stayed quit for more than one week at follow-up. See Ajzen et al. Data Report, (Exhibit DOM/IDN-2), paras. 221-224, Appendix A, pp. 95-97).

[122] See Ajzen et al. Data Rebuttal Report, (Exhibit DOM/IDN-6), paras. 63-65, Appendix IV.

[123] See Ajzen et al. Data Rebuttal Report, (Exhibit DOM/IDN-6), para. 90.

[124] See Ajzen et al. Data Rebuttal Report, (Exhibit DOM/IDN-6), paras. 50-54, Appendix II.

[125] See Chaloupka Rebuttal Report, (Exhibit AUS-582), para. 4.

Professor Chaloupka, based on the 2013 Australian National Drug Strategy Household Survey, average daily consumption of Australian smokers was 13.7 cigarettes per day.[126]

#### 3.1.7.3 Professor Klick's report

94. Professor Klick also reconsiders the NTPPTS data and estimated a negative binomial model of the average cigarettes consumption per day for daily smokers, smokers in general, or all participants, controlling for a TPP measures dummy variable, gender, age, education, socio-economic status and linear trend. The econometric analysis covers the period from 9 April 2012 to 4 May 2014.[127]

95. Overall, Professor Klick concludes that the TPP measures did not reduce consumption levels. In some of Professor Klick's specifications, the TPP measures' impact on average cigarettes per day for daily smokers and smokers in general is positive and statistically significant. According to Professor Klick, these results are consistent with his findings based on the Roy Morgan longitudinal survey data as well as with Scollo et al. 2015a's findings, which note that the NTPPTS data show no decline in self-reported smoking consumption associated with the TPP measures.[128]

96. As explained above, although Australia's expert, Professor Chaloupka, does not address specifically Professor Klick's analysis of cigarette consumption based on the NTPPTS data, Professor Chaloupka generally contends that because of its cross-sectional nature, the NTPPTS data limits the ability to follow the impact of the TPP measures through the pathway from its impact on the most proximal outcomes, like perceived appeal and noticing of the labels, through less proximal outcomes, such as increased interest in quitting, to the most distal outcomes, such as actual tobacco use behaviour.[129] Professor Chaloupka further argues that the impact of the TPP measures on more distal outcomes should be smaller when the analysis is based on the overall sample of smokers and recent quitters because smokers and recent quitters for whom the most proximal outcomes were unaffected by the TPP measures would not be expected to exhibit any change in more distal outcomes.

### 3.1.8 Cancer Institute New South Wales Tobacco Tracking Survey

97. The Cancer Institute New South Wales Tobacco Tracking Survey (CITTS) is a weekly tracking telephone survey of smokers and recent quitters (who quit in the past 12 months).

98. Professor Klick used the CITTS data to estimate a negative binomial model of the average number of reported tobacco units smoked per week controlling for the TPP measures, annual time trend, gender, individual age fixed effect, week of survey fixed effects and location fixed effects. The econometric analysis covers the period from January 2009 to December 2014.[130]

99. Overall, Professor Klick concludes that the TPP measures led to an increase in cigarette consumption, irrespective of whether the TPP period is taken to start in December 2012 or October 2012.[131] Professor Klick argues that these results are consistent with the finding from his study of the Roy Morgan Research Survey, Nielsen, Aztec, IMS/EOS and NTPPTS data.

100. Although Professor Chaloupka does not address directly Professor Klick's analysis of cigarette consumption based on the CITTS data, Professor Chaloupka contends that Professor Klick failed to appropriately account for the changes in the CITTS methodology that led to a greater number of younger people and males in the survey sample, both of which have relatively higher smoking rates, leading almost certainly to biased estimates towards showing an increase in smoking following the change in method.[132]

---

[126] See Chaloupka Rebuttal Report, (Exhibit AUS-582), para. 8.
[127] See Klick Supplemental Rebuttal Report, (Exhibit HND-122), paras. 44-51.
[128] See Klick Supplemental Rebuttal Report, (Exhibit HND-122), paras. 47-51.
[129] See Chaloupka Rebuttal Report, (Exhibit AUS-582), paras. 2-5.
[130] See Klick Second Supplemental Rebuttal Report, (Exhibit HND-165), paras. 25-26.
[131] See Klick Second Supplemental Rebuttal Report, (Exhibit HND-165), paras. 25-32.
[132] See Chaloupka Second Rebuttal Report, (Exhibit AUS-590), paras. 26-27.

### 3.2 Analysis by the Panel

101. Having determined above that the wholesale market data presented to us suggests an acceleration of the reduction in cigarette sales following the entry into force of the TPP measures, the question before us at this stage of our analysis is whether this acceleration may, in part or in whole, be attributed to the TPP measures.[133]

102. As explained above, while we acknowledge that no data are perfect, we agree with Australia that the IMS/EOS data is the most suitable available market data submitted by the parties to analyse the TPP measures' impact on cigarette sales because the data are available monthly for a long period of time before and after the introduction of the TPP measures. In addition, the IMS/EOS data display a relatively high correlation coefficient with respect to the Nielsen retail cigarette sales data.

103. We consider that the other market data sources considered by the experts of the Dominican Republic, Honduras and Indonesia suffer from a number of drawbacks in comparison with the IMS/EOS data. In particular, the Nielsen and Aztec data are both only available for a shorter sample period. In addition, the Aztec data are characterized by an increasing market coverage making it difficult to distinguish the impact of other factors.

104. At the outset, we note that for a given market dataset, the different conclusions reached by the parties regarding the impact of the TPP measures on cigarette sales stem from the fact that their experts use different model specification (i.e. different independent variables (e.g. cigarette sales or per capita cigarette sales) and different explanatory variables included in the model), estimation approaches and in some cases sample period. Even among the experts of the Dominican Republic and Honduras, different model specifications are used.[134]

105. As explained in our review of the econometric study on smoking prevalence in Appendix C, we consider that our task is not to conduct our own econometric assessment of the contribution of the TPP measures on cigarette sales or cigarette consumption, but rather to review the robustness of the econometric evidence submitted by the parties.

#### 3.2.1 IPE's econometric results

106. A careful review of the econometric results based on the IMS/EOS data reported by IPE leads us to question their robustness on various grounds, many of which are similar to those considered in our review of IPE's analysis of smoking prevalence in Appendix C. In particular, we note that IPE's preferred specification of the modified trend analysis and the ARIMAX model includes both a price variable and a time trend variable, which happen to be highly collinear with each other. Multicollinearity appears to be even more marked when the model specification of the ARIMAX model includes five lags of the logarithm of per capita sales variables and of the price variable.[135] Furthermore, we note that IPE fails to take into account the potential impact of the TPP measures on tobacco prices. Similarly, IPE ignores the fact that the proportionality assumption underlying the use of the tax level in the analysis of the IMS/EOS data is rejected. Moreover, IPE does not address the fact that the price variable and the tax level variable appear to be non-stationary. We note, however, that when the model includes the excise tax dummy variables, which were initially proposed by IPE itself in its first report but later rejected as inferior control variables, most results

[133] We note that the parties did not present any graphical analysis of reported cigarette consumption based on survey data. That is why the conclusions we reached so far only apply to cigarette sales used as a proxy of cigarette consumption.

[134] For instance, IPE initially analysed the impact of the TPP measures on (the logarithm of) cigarette sales (see IPE Report, (Exhibit DOM-100)) but subsequently analysed the impact on (the logarithm of) per capita cigarette sales. See, e.g. IPE Updated Report, (Exhibit DOM-303). Similarly, IPE initially (see IPE Report, (Exhibit DOM-100)) control for excise tax increases by specifying dummy variables for each excise tax hike implemented during the sample period, while IPE later in the proceedings (see, e.g. IPE Second Updated Report, (Exhibit DOM-361)) consider the variable of tobacco price to be a better proxy for excise tax increases. Professor Klick applies an IV estimation procedure to take into account the potential endogeneity of the price variable with respect to cigarette sales while IPE and Professor List do not address the potential endogeneity of the price variable. See Klick Rebuttal Report, (Exhibit HND-118), paras. 52-60.

[135] See IPE Second Updated Report, (Exhibit DOM-361), pp. 38-40; and IPE Third Updated Report, (Exhibit DOM-375), pp. 120-122.

suggest the TPP measures had a negative and statistically significant impact on wholesales cigarette sales.

107. We also question the validity of IPE's results based on the Nielsen data for the same reasons mentioned above, namely the use of the price variable. In particular, we note that in some specifications of the modified trend analysis, which include a trend variable and the price variable in the past month, only a couple of month dummies are statistically significant.[136] The same issue arises in several specifications of the ARIMAX model, where the time trend variable and the logarithm of cigarette sales per capita in the previous month are both not statistically significant, which would suggest that cigarette sales do not follow a dynamic pattern, despite graphical evidence attesting to the contrary.[137] We also note that some of the results of the modified analysis based on the procedure to compute the standard errors applied initially, but that were later rejected by the IPE, find a positive and statistically significant impact of the TPP measures on cigarette sales.[138] As explained in the review of the econometric studies on smoking prevalence, the choice of the maximum lag required to compute robust standard error is not well established in the statistics and econometric literature and it is therefore unclear to what extent these results would have changed for a range of different parameter values.[139] We note also that in some specifications of the ARIMAX model based on the deseasonalized Nielsen data the estimated coefficient of the lagged dependent variable is negative and statistically significant, which is in contradiction with the results of the same model based on the Aztec data.[140] In addition, as explained earlier, the sample period of the Nielsen data is limited with only 35 observations available. Yet, a smaller sample size (i.e. number of observations) makes it, all else being equal, more difficult to estimate more accurately any impacts on cigarette sales.

108. We also question the IPE's results based on the Aztec data.[141] As explained above, the Aztec data are characterized by a growing market coverage, which, in our view, makes it more difficult to distinguish the impact of the explanatory variables, including the TPP measures dummy variable, from the growing market coverage. We note that the IPE proposed as a robustness check to adjust the Aztec data for the increasing market coverage. Yet, we note that the IPE did not provide any evidence that would justify why the transformed Aztec data would reflect more accurately the actual fluctuation of the cigarette retail sales. In fact, we note that the correlation coefficient between the adjusted Aztec data and the Nielsen data or the IMS/EOS data is lower than with respect to the original Aztec data.[142]

### 3.2.2 Professor Klick's econometric results

109. Similarly, a review of the econometric results reported by Professor Klick leads us to question their robustness. In particular, we note that in the first stage of the IV estimation of

[136] We also note that the estimated coefficients of the modified trend analysis based on the Nielsen data are different between IPE Second Updated Report (Exhibit DOM-361), pp. 41-42 and IPE Summary Report (Exhibit DOM-379), p. 54 and backup material, although this is the same model specification and sample period. We also note that the model specification with tax dummies based on the Nielsen data reported in Table 4.2-2 in IPE Summary Report (Exhibit DOM-379) actually do not include the December 2013 tax dummy variable. See IPE Summary Report, (Exhibit DOM-379), backup material.

[137] See IPE Second Updated Report, (Exhibit DOM-361), pp. 45-46.

[138] See IPE Second Updated Report, (Exhibit DOM-361), backup material.

[139] See the email exchange between the parties and STATA developers in the IPE Summary Report, (Exhibit DOM-379), p. 70; and Chipty Second Rebuttal Report, (Exhibit AUS-591), pp. B1-B2.

[140] See IPE Third Updated Report, (Exhibit DOM-375), pp. 76-77. In addition, we note that in some specifications of the modified trend analysis based on the deseasonalized Nielsen data the estimated coefficient of the TPP measures is positive and statistically significant.

[141] We note that in a couple of specifications, which include, for instance, the tax level variable, the results find a positive and statistically significant TPP measures' effect. See IPE Summary Report, (Exhibit DOM-379), p. 54; and IPE Third Updated Report, (Exhibit DOM-375), pp. 124 and 127. Similar findings are sometimes found when the standard errors are computed applying the initial procedure proposed by IPE, namely the IVREG2 command, for the July 2008-May 2015 sample period. See IPE Second Updated Report, (Exhibit DOM-361), backup material.

[142] The weekly adjusted Aztec data were aggregated to monthly data by assuming that each weekly cigarette quantity can be evenly split for each day of the week. The correlation coefficient between the Nielsen data and the original Aztec data is equal to 0.84, while the correlation between the Nielsen data and the adjusted Aztec data is equal to 0.71. The correlation between the original Aztec data and the IMS/EOS data is 0.38, while the correlation between the adjusted Aztec data and the IMS/EOS data is 0.32.

Professor Klick's difference-in-difference analysis based on the IMS/EOS data, surprisingly none of the 54 time-fixed effects are statistically significant. We further note that the second stage of the IV estimation is affected by the multicollinearity between the instrumented price variable and the linear trend variable.[143] In addition, it is unclear how the standard error and confidence interval of the total effect of the TPP measures, composed of the direct estimated TPP measures' impact on cigarette sales obtained in the second stage of the procedure and the indirect estimated TPP measures' impact on cigarette price obtained in the first stage of the procedure through the impact of cigarette price on cigarette sales, were calculated. We also question the results of the difference-in-difference analysis based on the Nielsen-Aztec data because of the use of the price variable as well as the growing market coverage of the data.[144] In particular, we note that in the first stage of the IV estimation procedure, multicollinearity appears to be high between the change in excise tax variable, the country-specific trend variable and the TPP measures dummy variable. Similarly, the multicollinearity between the instrumented price variable, the country-specific trends and several time-fixed effects is particularly high in the second stage of the IV estimation procedure.[145] In addition, we note that surprisingly most of the time-fixed effects in both first and second stages of the IV estimation procedure are not statistically significant. We also question the results of the model specification without New Zealand counterfactual control, because the specification without a time trend finds a positive and statistically significant TPP measures' effect and the specifications with (linear or quadratic) trend find that none of the explanatory variables, except the constant, are statistically significant.[146]

110. Similar to our conclusion regarding Professor Klick's difference-in-difference analysis of smoking incidence based on the Roy Morgan Research Survey data, we question the validity of the results on the reported number of cigarettes consumed on the same grounds, namely the pre-period of the analysis is not valid because Professor Klick is unable to accurately identify the respondents, who had noticed plain packs in the pre-period. Furthermore, we note that the instrumented tax used in the second stage of the IV estimation procedure displays signs of multicollinearity with respect to the TPP measures dummy and several time fixed effects.[147]

111. We also question Professor Klick's results on the average reported number of cigarettes smoked per day based on the NTPPTS data because the majority of the results suggest that the TPP measures' effect is positive and statistically significant.[148] It is unclear to what extent this finding is the result of the model not being specified correctly given that Scollo et al. 2015a using the same NTPPTS data but applying a different model specification find that cigarette consumption did not change during the first year of the implementation of the TPP measures.[149] For instance, we note that unlike Scollo et al. 2015a, Professor Klick does not control for the past three-month exposure to anti-smoking campaigns that aired on television during the survey period.

112. Similarly, we question the validity of Professor Klick's results on the reported number of tobacco units consumed per week based on the CITTS data. We note that most of the 49 fixed effects associated with the week of survey are surprisingly not statistically significant in most specifications.[150] In addition, as noted in Appendix B, the survey data, such as the CITTS data, may, as suggested by Australia, be more suited to analysing the impact of the TPP measures on

[143] Evidence of multicollinearity is confirmed by the variance inflation factors statistic.

[144] We note that the Aztec data used by Professor Klick is available monthly, while IPE and Professor List used the Aztec data reported at the weekly level.

[145] We also note that without the country-specific trend variable, the change in excise tax variable continues to display a high variance inflation factor statistics, when the country-specific trend variable is removed (see Klick Fourth Supplemental Rebuttal Report, (Exhibit HND-169), backup material).

[146] See Klick Third Supplemental Rebuttal Report, (Exhibit HND-166), backup material; and Klick Fourth Supplemental Rebuttal Report, (Exhibit HND-169), backup material. We also note that according to standard unit root tests, the variable reporting the number of stores covered by the Aztec data is not stationary.

[147] We also note that in the first stage of the IV estimation procedure, surprisingly the majority of the time fixed effects are not statistically significant. The same issues arise in a couple of specifications considered in Professor Klick's "crossfold validation". In one specification, the only variable statistically significant is the constant term or the level of cigarette sales in New Zealand (see Klick Rebuttal Report, (Exhibit HND-118), backup material).

[148] See Klick Supplemental Rebuttal Report, (Exhibit HND-122), pp. 19-21.

[149] See Scollo et al. 2015a, (Exhibits HND-133 DOM-237, DOM-311), p. ii73.

[150] See Klick Second Supplemental Rebuttal Report, (Exhibit HND-165), pp. 11-12. We also note that most of the results suggest that the TPP measures' effect on reported tobacco units consumed per week is positive and statistically significant.

proximal outcomes, such as appeal, GHWs and ability of the pack to mislead than more distal outcomes, such as smoking behaviours.

### 3.2.3 Ajzen et al.'s econometric results

113. A careful review of Ajzen et al.'s analysis of the reported number of cigarettes smoked daily and the proportion of smokers consuming at least a given number of cigarettes per day reported in the NTPPTS data leads us to conclude that Ajzen et al. corroborate Scollo et al.'s (2015a) findings that there were no changes in daily cigarette consumption during the first year of implementation of the TPP measures.[151] We note, however, that in several specifications, Ajzen et al. do not analyse directly the quantity of cigarettes smoked reported by the respondent, but, instead, focus on the probability the respondent consumes a given number of cigarettes or more per day. We further note that Ajzen et al.'s quarterly results find that the probability of reporting smoking 20 cigarettes or more per day increased in the first quarter following the introduction of the TPP measures.[152] Yet, Ajzen et al. do not provide any explanation that would explain such a result.

114. The review of Ajzen et al.'s analysis of the number of cigarettes smoked per day reported in the ITC dataset also raises a number of questions. At the outset, we note that Ajzen et al. do not analyse directly the number of cigarettes smoked per day, but a categorical variable distinguishing between 0-10, 11-20, 21-30 and more than 30 cigarettes per day. We further note that the model specification does not control for anti-smoking advertising in mass media, as acknowledged by Ajzen et al.[153], and for cigarette costliness. It is therefore unclear if the results would be similar if these explanatory variables had been taken into account. In addition, we note that the results of the weighted prevalence estimates are based in a specification in which only a few variables, namely the variables for high education level, quit attempt and survey mode, are statistically significant, besides the constant.[154] It is therefore unclear why most explanatory variables in the model considered by Ajzen et al. are not statistically significant.

### 3.2.4 Dr Chipty's econometric results

115. Turning to the econometric results analysing wholesale cigarette sales based on the IMS/EOS data submitted by Australia's expert, Dr Chipty, we note that some concerns that we raised regarding the experts of the Dominican Republic's, Honduras's and Indonesia's approaches and results of the market data have been to some extent addressed by Dr Chipty. In particular, Dr Chipty specifies excise tax increases dummy variables and thus avoids the problems of endogeneity associated with the inclusion of the price variable as well as the unit root problem of the price or tax level variables. We further note that part of Dr Chipty's model specification is based on the first specification proposed by the IPE but modified to account for strategic inventory management and the 2006 GHWs regulation.

116. A careful review of Dr Chipty's econometric results further shows that the negative and statistically significant impact of the TPP measures on wholesale cigarette sales is robust to alternative specifications, including different TPP starting date (October, November and December 2012), and to Professor List's procedure to compute standard errors.[155] We note however that the TPP measures' effects are no longer statistically significant when the set of tax hikes dummies are replaced by a tax levels variable. Yet, we note as explained earlier that

[151] We note that the model specification considered by Ajzen et al. is different from the one developed by Professor Klick. For instance, Ajzen et al. include dummy variables for two age groups, while Professor Klick includes a single age variable. Ajzen et al. also includes dummy variables for different groups of education and socio-economic status, respectively, while Professor Klick includes a single variable for education and another variable for socio-economic status.

[152] See Ajzen et al. Data Rebuttal Report, (Exhibit DOM/IDN-6), p. 22.

[153] See Ajzen et al. Second Data Report, (Exhibit DOM/IDN-4), para. 18.

[154] See Ajzen et al. Second Data Report, (Exhibit DOM/IDN-4), backup material. When the sample is restricted to participants involved in the pre- and post-implementation surveys, the only statistically significant variables are two age categories, quitting attempt and the constant. In the weighted prevalence estimates results, the only statistically significant variables are two age group categories. Ajzen et al. Data Rebuttal Report, (Exhibit DOM/IDN-6), backup material.

[155] See Chipty Surrebuttal Report, (Exhibit AUS-586), p. 40; and Chipty Second Rebuttal Report, (Exhibit AUS-591), pp. 19-20, 33, and D1.

specification testing suggests tax levels are not appropriate in the model specification.[156] Furthermore, the tax levels variable is likely to be non-stationary.

117. Overall, based on the most recent econometric evidence submitted by Australia, there is some econometric evidence suggesting that TPP and enlarged GHWs contributed to the reduction in wholesale cigarette sales in Australia.[157]

## 4 EVIDENCE RELATING TO THE EVOLUTION OF IMPORTS OF CIGARS

118. The parties did not present an econometric analysis of the post-implementation impact of the TPP measures on cigar consumption. Rather, the parties addressed a graphical and descriptive analysis of trade data.

119. According to data analysed by Australia[158], the value and volume of Australian cigar and cigarillo imports shows fluctuations within a set range throughout the period between March 2005 and March 2015.[159] Australia notes, in any event, that the values represented in this dataset are "customs values", and may not represent the actual value of the products.[160]

120. Australia also submits the HoustonKemp Report[161], which notes that all cigars and cigarillos sold in Australia are currently imported.[162] The HoustonKemp Report concludes that imports of cigars and cigarillos have fluctuated over time and have fallen significantly since 20[[***]]: annual imports of cigars and cigarillos fell by [[***]]% from 20[[***]] to 20[[***]], following the introduction of the TPP measures in December 2012, whereas annual import volumes fell at a faster rate prior to the TPP measures being introduced, by [[***]]% from 20[[***]] to 20[[***]].[163] As regards monthly cigar and cigarillo imports to Australia, the HoustonKemp Report finds that the level of monthly imports varies substantially, although, on average, these have been falling since the TPP measures came fully into force. The HoustonKemp Report adds that some of the increases in imports for a particular month coincide with the introduction of the TPP measures and increases in excise tax rates.[164]

121. Based on data from the International Trade Centre's trade statistics database, the complainants argue that cigar imports to Australia have remained relatively stable over the period 2010-2014, and point out that the difference between the evolution of cigarette and cigar imports to Australia can be explained by the fact that the reduced-fire risk requirements did not apply to cigars.[165]

122. As regards large cigars (i.e. excluding cigarillos and little cigars), the HoustonKemp Report adds that total imports into Australia from all countries have varied within the range of [[***]],000 kg to [[***]],000 kg from 20[[***]] to 20[[***]], and the level of imports fell from 20[[***]] to 20[[***]], immediately before the TPP measures came into force, but have risen in 20[[***]].[166] The HoustonKemp Report concludes that "[i]t follows that there is no evidence of the TPP measure[s] causing a fall in the imports of large cigars".[167]

---

[156] See Chipty Third Rebuttal Report, (Exhibit AUS-605), paras. 27-29.

[157] See Chipty Third Rebuttal Report, (Exhibit AUS-605), pp. 22 and 51; and Chipty Second Rebuttal Report, (Exhibit AUS-591), pp. 33, D1, and D2.

[158] Supplementary Graphs, Import Volumes, Value and Share of the Market, (Exhibit AUS-512).

[159] Supplementary Graphs, Import Volumes, Value and Share of the Market, (Exhibit AUS-512), Figures 7, 8, 13, and 14.

[160] Supplementary Graphs, Import Volumes, Value and Share of the Market, (Exhibit AUS-512), Figures 7, 8, 13, and 14.

[161] HoustonKemp Report, (Exhibit AUS-19) (SCI).

[162] HoustonKemp Report, (Exhibit AUS-19) (SCI), p. 45.

[163] HoustonKemp Report, (Exhibit AUS-19) (SCI), p. 45.

[164] HoustonKemp Report, (Exhibit AUS-19) (SCI), p. 45.

[165] Dominican Republic's response to Panel question No. 5. See also Honduras's response to Panel question No. 5; Cuba's response to Panel question No. 5 (annexed to its response to Panel question No. 138); and Indonesia's response to Panel question No. 5.

[166] HoustonKemp Report, (Exhibit AUS-19) (SCI), p. 47.

[167] HoustonKemp Report, (Exhibit AUS-19) (SCI), p. 47.

123. We have difficulty drawing conclusions from the above evidence on the evolution of Australian cigar, large handmade (LHM) cigar, or cigarillo consumption and imports. Based on the above evidence, we conclude that, despite fluctuations, overall imports of cigars and cigarillos have experienced a downward trend in recent years. However, it is unclear whether and to what extent this is attributable to the TPP measures.

124. Some of the evidence submitted by the parties relates to cigar imports per complainant. In particular, the HoustonKemp Report submitted by Australia explains that the complainants have accounted for a small proportion of Australia's annual imports of cigars and cigarillos in general. From 19[[***]] to 20[[***]], Cuba accounted for between [[***]]% and [[***]]% of total annual imports, the Dominican Republic accounted for between [[***]]% and [[***]]% of imports, and the other complainants collectively accounted for less than [[***]]% of imports.[168]

125. Furthermore, the Panel explored with Cuba, Honduras, and the Dominican Republic whether they had data on cigar sales to Australia and how these were affected by the TPP measures. Cuba does not provide a direct answer to the Panel's question concerning how volumes and values of sales of Cuban cigars in Australia, including Cuban LHM cigars, and Cuban LHM cigars carrying the Habanos GI and/or the Cuban Government Warranty Seal might have changed as a result of the TPP measures, and whether the price of cigars in the above product categories has changed as a result of the TPP measures.[169] Cuba states that a reply to this question would require access to the data on retail sales of Cuban LHM cigars in Australia as well as other markets for purposes of comparison, and these data would have to be analysed and compared so as to take account of the specific features of each market (such as the level of taxes) in order to draw reasonable conclusions; however, "[a]t present, Cuba does not have access to such data".[170] Australia points out that Cuba's response does not provide the Panel with any information or evidence about the volume and value of sales of Cuban cigars in Australia following the introduction of tobacco plain packaging.[171]

126. Nonetheless, Cuba provides data as regards sales volumes of Cuban LHM cigars in Australia, namely data relating to wholesale sales of LHM cigars in Australia between January 2009 and July 2014 by Pacific Cigar Company (PCC).[172] Cuba considers such data as a relevant measure of consumption of LHM cigars as it is estimated that PCC have held a market share of approximately 70% of total LHM cigar sales in Australia over the relevant period.[173] According to Cuba, these data show that monthly sales of LHM cigars in Australia largely fluctuated within a set range of sticks per month for the majority of the relevant period, rising to the upper end of this range from 2013 onwards.[174] Cuba adds that there does not appear to have been any decrease in monthly sales of LHM cigars after December 2012 when the TPP measures were introduced, not least when the post-implementation sales volumes are compared to monthly sales in the two years immediately prior to the introduction of the TPP measures.[175] According to Cuba, indeed, it appears that sales of LHM cigars in Australia have marginally increased since December 2012.[176] Cuba draws similar conclusions from the same sales information in half-yearly, rather than monthly, format. According to Cuba, the same sales data in half-yearly format illustrate the consistency of sales volumes over time and the apparent increase in sales of LHM cigars in the post-implementation period.[177]

127. Australia points out that the HoustonKemp Report demonstrates that although overall volumes of cigar and cigarillo imports into Australia have been declining since the introduction of tobacco plain packaging, in 20[[***]] imports of Cuban cigars into Australia "rose to their highest level since 19[[***]], after the TPP measure was introduced".[178] Indeed, the HoustonKemp Report

[168] HoustonKemp Report, (Exhibit AUS-19) (SCI), pp. 46-47.
[169] Cuba's response to Panel question No. 193.
[170] Cuba's response to Panel question No. 193.
[171] Australia's comments on Cuba's response to Panel question No. 193.
[172] Cuba's first written submission, paras. 159-161.
[173] Cuba's first written submission, para. 159.
[174] Cuba's first written submission, para. 160.
[175] Cuba's first written submission, para. 160.
[176] Cuba's first written submission, para. 160.
[177] Cuba's first written submission, para. 161.
[178] Australia's comments on Cuba's responses to Panel question Nos. 192, 194, and 195 (citing HoustonKemp Report, (Exhibit AUS-19) (SCI), p. 47).

finds that approximately [[***]],000 kg of cigars and cigarillos were imported from Cuba each year from 20[[***]] to 20[[***]]; however, in 20[[***]], Cuba's imports into Australia rose to the highest level since 19[[***]], after the TPP measures were implemented.[179] As regards large cigar imports into Australia (i.e. excluding cigarillos and other cigars), the HoustonKemp Report finds that Cuba's share has increased from 19[[***]] to 20[[***]].[180]

128. Based on the data concerning PCC and the data available in the HoustonKemp Report, we conclude that there has been an increase in the volume of Cuban LHM cigar and cigar/cigarillo sales in Australia. Of note, the underlying data is limited and indirect in that it does not provide any information on the factors driving the evolution of sales of Cuban cigars in Australia or address the role of the TPP measures in that regard.

129. As regards sales of Dominican Republic cigars in Australia, the Dominican Republic notes, as a preliminary matter, that general international trade statistics that capture trade reported by domestic customs authorities, such as the UN Comtrade Database, do not provide a reliable source of information. In particular, for the purposes of tracking Dominican Republic cigar sales in Australia, this data is not helpful for a number of reasons, not least because it can mis-specify the true country of origin if a company is domiciled in a third country and because Dominican Republic cigars may be mis-classified if the product is imported into, and then re-exported by, a third country (e.g. Belgium, the Netherlands, Singapore or Hong Kong).[181]

130. Rather, the Dominican Republic submits data from the Dominican Republic's cigar industry for sales of Dominican Republic cigars in Australia.[182] According to the Dominican Republic, this data shows that following the implementation of the TPP measures, there has been a decline in sales of Dominican Republic cigars, which are predominantly premium hand-rolled products.[183] The Dominican Republic notes that it has not been able to obtain data regarding the price and value of Dominican Republic cigar sales in Australia following the imposition of the TPP measures.[184]

131. Australia responds that the Dominican Republic's data show that the volume of sales of Dominican Republic cigars in Australia has declined in the period following the introduction of tobacco plain packaging; however, the Dominican Republic provides no evidence of the impact of the TPP measures on the value and prices of sales of Dominican Republic cigars in Australia.[185] Further, Australia notes, although requested by the Panel, the Dominican Republic has not provided any information on the impact of the measure on the subset of cigars that were permitted to use the GI "Cigarro Dominicano" prior to the implementation of tobacco plain packaging."[186]

132. We agree with the Dominican Republic that the UN Comtrade Database has limitations, for establishing export patterns from an individual exporter. We also concur with the limitations of the Dominican Republic's sales data highlighted by Australia. Importantly, none of the data explains the role of the TPP measures in any changes to sales of Dominican Republic cigars in Australia.

133. Based on data from the UN Comtrade Database, Honduras argues that the volumes and values of Honduran cigars imported into Australia have "drastically decreased".[187] According to Honduras, "at face value", the value and volume of trade in Honduran cigars to Australia has "drastically decreased" as there has been "a drop by 98%" between 2010 and 2014.[188] Honduras adds that, in terms of the combined value and quantity of imports of Honduran cigars into Australia, if one compares 2010 and 2011 to 2013 and 2014 (i.e. discounting the year of 2012 as the year when plain packaging was introduced), import values have decreased by 91% and import volume by 97%.[189] Honduras points out that in the same period, its exports of cigars to the

[179] HoustonKemp Report, (Exhibit AUS-19) (SCI), p. 47.
[180] HoustonKemp Report, (Exhibit AUS-19) (SCI), p. 47.
[181] Dominican Republic's response to Panel question No. 194.
[182] Dominican Republic's response to Panel question No. 194.
[183] Dominican Republic's response to Panel question No. 194.
[184] Dominican Republic's response to Panel question No. 194.
[185] Australia's comments on Cuba's response to Panel question No. 194.
[186] Australia's comments on Cuba's response to Panel question No. 194.
[187] Honduras's response to Panel question No. 195.
[188] Honduras's response to Panel question No. 195.
[189] Honduras's response to Panel question No. 195.

world substantially increased, which confirms that there was not a general downward trend of Honduran cigar exports in the period 2010 to 2014 but that the drastic decline is specific to the Australian market with tobacco plain packaging.[190]

134. Australia notes that Honduras relies on the UN Comtrade Database, which the Dominican Republic has qualified as "unreliable" and "not helpful"[191], and does not provide any data on the prices of its cigars in the Australian market.[192] Australia adds that notwithstanding the disagreement among Honduras and the Dominican Republic as to the reliability of the UN Comtrade Database, the data presented by Honduras appears to show a rapid decline in sales of Honduras cigars in Australia beginning in 2012 (the year that the TPP measures were introduced), and such decline in imports is consistent with Australia's argument that the packaging changes introduced in late 2012 led to a decline in cigar smoking prevalence.[193]

135. As noted, we agree with Australia and the Dominican Republic concerning the serious limitations of the UN Comtrade Database highlighted by the Dominican Republic for establishing total imports (direct and indirect) from an individual exporting country. Accordingly, we have difficulty drawing conclusions from such data specifically for all imports (direct and indirect) of Honduran cigars to Australia.

136. In the light of the above, we conclude that the evidence before us on cigars, cigarillos and LHM cigars allows us to draw limited conclusions on the evolution of certain consumption trends. We note that despite fluctuations, overall imports of cigars and cigarillos to Australia have followed a downward trend in recent years. However, we have no evidence to link these conclusions to the TPP measures that would allow us to draw conclusions on the basis of this evidence on the effect of the TPP measures on cigar consumption in Australia.

## 5 OVERALL CONCLUSION ON POST-IMPLEMENTATION EVIDENCE ON TOBACCO PRODUCT SALES VOLUMES

137. Overall, based on the most recent data available and evidence submitted by the parties, we find that:

a. There is some evidence that cigarette sales in Australia continued to decrease following the introduction of the TPP measures.

b. The downward trend in cigarette sales in Australia appears to have accelerated in the post-TPP period.

c. Although it is impossible to distinguish between the impact of TPP and enlarged GHWs, there is some econometric evidence suggesting that the TPP measures, in combination with the enlarged GHWs implemented at the same time, contributed to the reduction in wholesale cigarette sales, and therefore cigarette consumption, after their entry into force.

d. The evidence before us on the evolution of consumption of cigars in the post-TPP period is more limited and does not allow us to draw clear conclusions on the effect of the TPP measures on cigar consumption in Australia.

138. We note, however, that no post-implementation empirical evidence has been presented to us on the impact of the TPP measures on cigars and cigarillos consumption.

[190] Honduras's response to Panel question No. 195.
[191] Australia's comments on Cuba's response to Panel question No. 195 (referencing Dominican Republic's response to Panel question No. 194).
[192] Australia's comments on Cuba's response to Panel question No. 195.
[193] Australia's comments on Cuba's response to Panel question No. 195.

## APPENDIX E:

## POST-IMPLEMENTATION EVIDENCE ON DOWNWARD SUBSTITUTION IN THE CIGARETTE MARKET

1. The Dominican Republic and Indonesia submitted empirical statistical and econometric studies analysing the impact of the TPP measures (and enlarged GHWs) on the change in consumption patterns from premium to lower-priced brands.[1] Different terms have been used to refer to this phenomenon: downward substitution[2] or downtrading.[3]

2. One of the only points of agreement among the parties is that it is not possible, on the basis of the available data, to distinguish between the impact of plain packaging and the impact of the enlarged GHWs on cigarette sales or consumption, because both measures were implemented at exactly the same time.[4] Unless specified otherwise, the impact of plain packaging in this section therefore refers to the impact of plain packaging <u>and</u> enlarged GHWs.

3. The Dominican Republic and Indonesia argue that all the empirical statistical and econometric studies undertaken by their experts, the IPE (for the Dominican Republic), and Professor List (for the Dominican Republic and Indonesia), point to the fact that the TPP measures led consumers to replace higher-priced cigarettes with low-priced cigarettes and to shift their preferences from higher-priced to low-priced cigarettes.[5]

4. Similar to the discussion on smoking prevalence and cigarette consumption, we note that the evidence presented to us to analyse substitution can be grouped in three main approaches, that are reviewed separately next:

- First, the Dominican Republic and Indonesia have submitted descriptive statistics analysis aimed at determining whether the shares of higher- to low-priced cigarettes sales and smokers have increased following the implementation of the TPP measures;

- Second, the Dominican Republic and Indonesia have submitted statistical analysis to determine whether there was a break in the trend in downward substitution following the implementation of the TPP measures, and in particular, whether the reduction in the shares of higher- to low-priced cigarette sales and smokers have accelerated following the implementation of the TPP measures;

- Finally, some of the parties have submitted econometric analyses to determine whether the TPP measures have contributed to the downward trend of the relative quantities and preferences by isolating and quantifying the different factors that can explain the evolution of the share of sales volume of higher- to low-priced cigarettes as well as the evolution of the share of higher-priced cigarette smokers relative to smokers of low-priced cigarettes.

5. Unlike the discussion on smoking prevalence and cigarette consumption, however, Australia did not engage in estimating the impact of plain packaging on downward substitution in cigarettes. Australia argues that the Dominican Republic and Indonesia have failed to demonstrate that any

[1] See IPE Report, (Exhibit DOM-100); IPE Updated Report, (Exhibit DOM-303); IPE Second Updated Report, (Exhibit DOM-361); IPE Third Updated Report, (Exhibit DOM-375); IPE Summary Report, (Exhibit DOM-379); and List Rebuttal Report, (Exhibit DOM/IDN-3).

[2] See IPE Report, (Exhibit DOM-100), pp. 16-17.

[3] See IPE Updated Report, (Exhibit DOM-303), paras. 212 and 228-235; and List Rebuttal Report, (Exhibit DOM/IDN-3), paras. 104 and 111.

[4] See Australia's first written submission, para. 518; Dominican Republic's response to Panel question No. 8, para. 61; Honduras's response to Panel question No. 8; and Indonesia's response to Panel question No. 8, para. 8.

[5] See IPE Report, (Exhibit DOM-100), paras. 4 and 61-67; IPE Updated Report, (Exhibit DOM-303), paras. 176-180; IPE Second Updated Report, (Exhibit DOM-361), paras. 43-59; IPE Third Updated Report, (Exhibit DOM-375), paras. 23 and 245-258; IPE Summary Report, (Exhibit DOM-379), paras. 43-45; and List Rebuttal Report, (Exhibit DOM/IDN-3), paras. 104-111.

downward substitution effects that have occurred in the Australian market are attributable to the TPP measures and not to other factors.[6]

## 1 WHETHER DOWNWARD SUBSTITUTION IN CIGARETTES INCREASED FOLLOWING THE IMPLEMENTATION OF THE TPP MEASURES

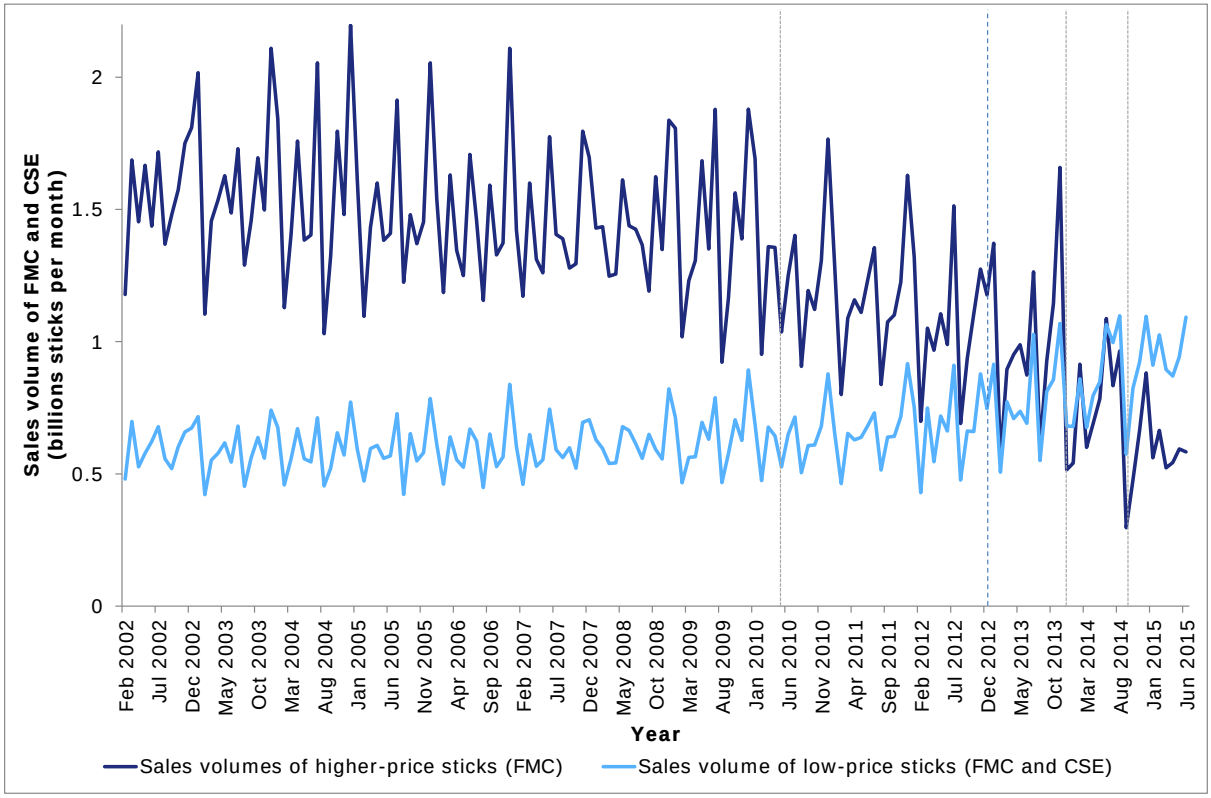
6. Two market data sources tracking cigarette sales volumes in Australia were used by the experts of the Dominican Republic, Honduras and Indonesia to compute the share of sales volume of higher- to low-priced cigarettes: (1) In-Market-Sales/Exchange of Sales (IMS/EOS) data and (2) Aztec data. In addition, the experts of the Dominican Republic and Indonesia used the Roy Morgan Single Source (RMSS) survey data to compute the difference between the share of higher-priced cigarette smokers and the share of low-priced cigarette smokers.

### 1.1 Datasets and related studies

#### 1.1.1 In-Market-Sales/Exchange of Sales

7. The IPE classifies monthly tobacco product sales from manufacturers to wholesalers and retailers reported in the IMS/EOS data by classifying the cigarette market into two distinct price segments, derived from information provided from the Aztec data: (1) higher-priced cigarettes comprising "premium" and "mid-price" factory-made cigarettes, and (2) low-priced cigarettes comprising "low-price" and "deep discount" factory-made cigarettes and cigarette stick equivalents (CSE) from fine-cut tobacco.[7]

**Figure E.1: Monthly Sales Volumes of Higher- and Low-Priced Sticks Based on IMS/EOS Data**




Note: Conversion rate of 0.8 grams of fine-cut tobacco per stick. The vertical dotted lines indicate excise tax hikes. The vertical dashed line indicates the introduction of plain packaging and enlarged GHWs.

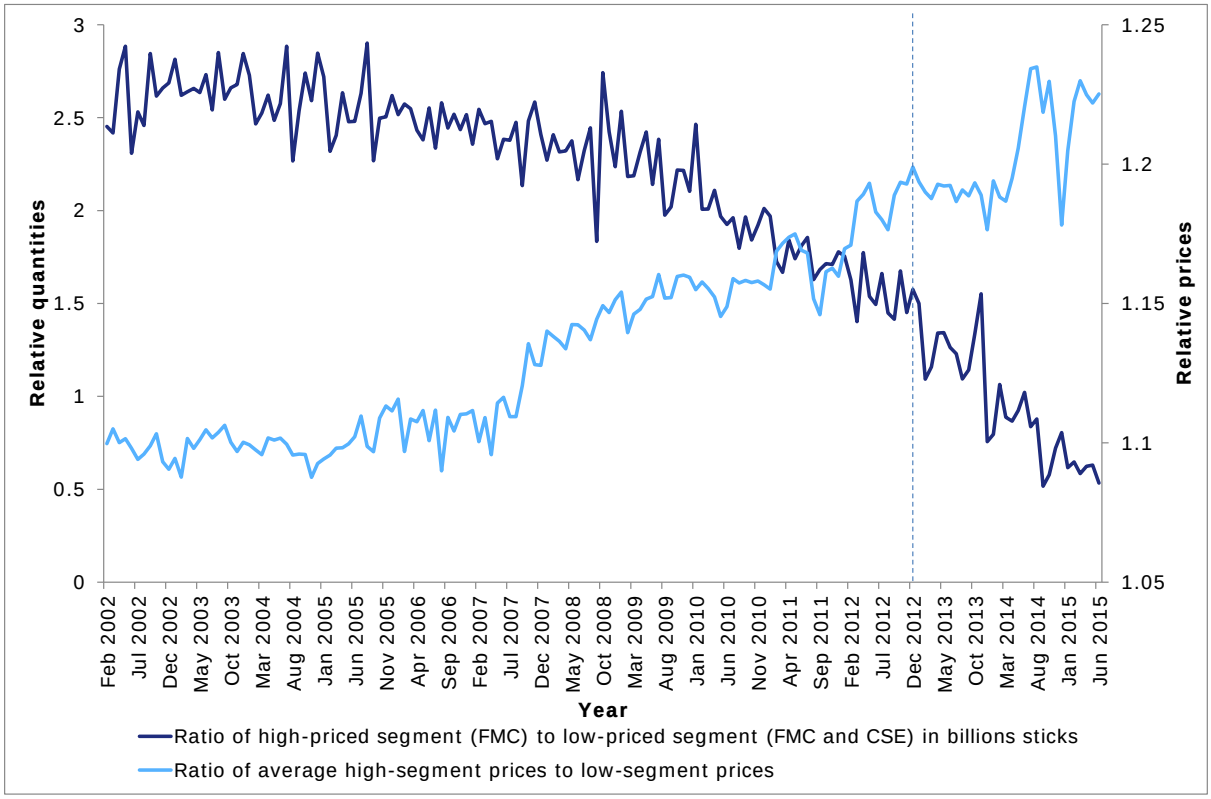
Source: IPE Updated Report, (Exhibit DOM-303), p. 73.

[6] Australia's first written submission, paras. 542-546.
[7] See IPE Report, (Exhibit DOM-100), p. 76.

8. The IPE notes, as depicted in **Figure E.1**, that sales volumes of both higher- and low-priced cigarettes are marked by strong seasonal patterns. According to the IPE the overall relationship between sales of higher-priced and low-priced cigarettes changes over time and is not linear.[8] In particular, between 2002 and up to 2008, sales volumes for higher-priced cigarettes and low-priced cigarettes remained relatively stable, with volumes of higher-priced products being more than twice as large as those for low-priced products. From mid-2007 on, sales volumes of higher-priced cigarettes started to decline, but volumes of low-priced goods remained largely stable. From late 2012 onwards and following the entry into force of the TPP measures, volumes of low-prices sticks began to increase, and from late 2013 on, more low-priced cigarettes were sold than higher-priced cigarettes.

9. The IPE further compares the evolution of the ratio of higher- to low-priced cigarettes with respect to the ratio of cigarettes' average high prices to low prices, as shown in **Figure E.2**.[9] The IPE notes that the relative sales volumes were quite stable up to mid-2007, with a weak linear downward trend. During the same period, relative prices increased only slightly. From mid-2007 onwards, relative prices started to increase much faster than before, meaning that higher-priced cigarettes became more and more expensive relative to low-priced alternatives. Coinciding with the relative price increase (from mid-2007 onwards), relative quantities (quantities of higher-priced sticks relative to low-priced sticks sold) started to decrease from mid-2007. Since the introduction of the TPP measures, the sales ratio of higher- to low-priced cigarettes continued to decrease, while the relative prices continued to increase.[10]

**Figure E.2: Monthly Sales Volumes Ratio of Higher-to Low-Priced Sticks and Ratio of Average Higher to Low Prices Based on IMS/EOS Data**




Note: Conversion rate of 0.8 grams of fine-cut tobacco per stick. The vertical dashed line indicates the introduction of plain packaging and enlarged GHWs.

Source: IPE Updated Report, (Exhibit DOM-303), p. 75.

[8] See IPE Report, (Exhibit DOM-100), p. 80.

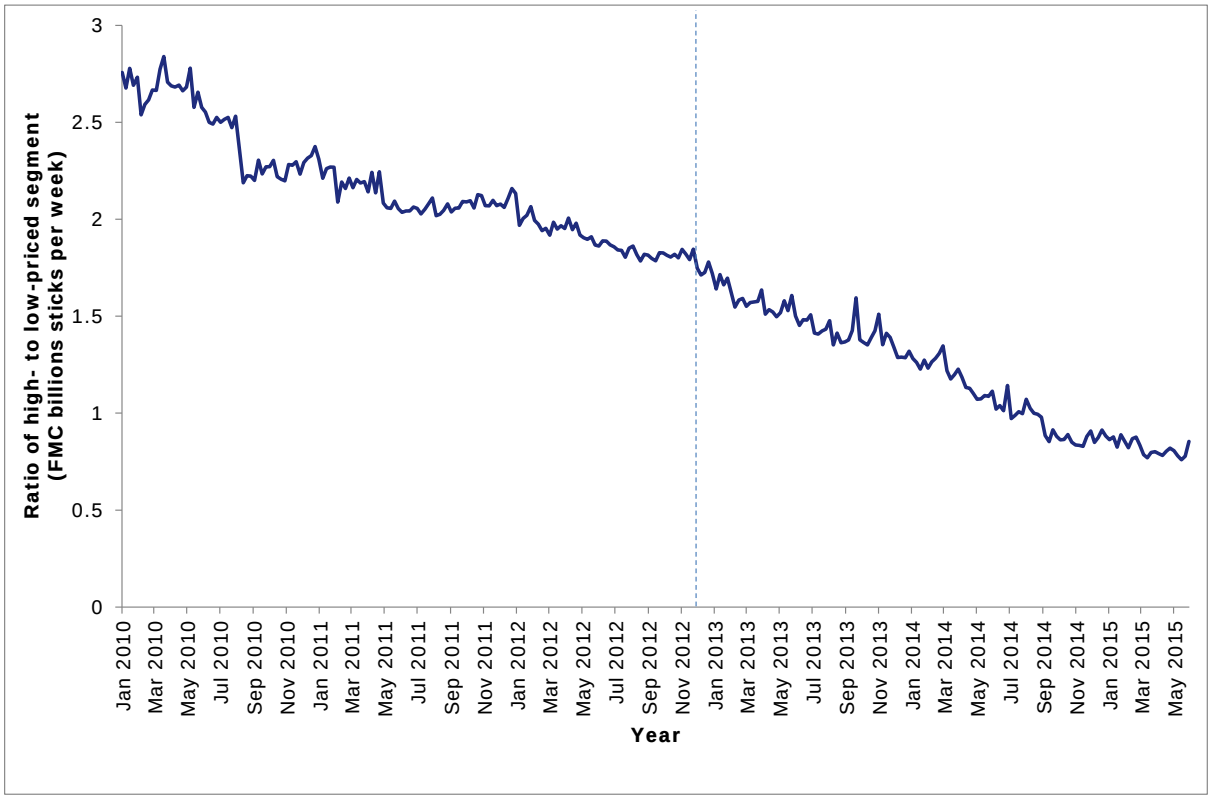
[9] We note that Professor List reports also the evolution of the ratio of higher- to low-priced cigarettes based on the IMS/EOS data. List Rebuttal Report, (Exhibit DOM/IDN-3), pp. 27-28. We note, however, that unlike IPE, Professor List does not include the sales of RYO tobacco in his analysis of the Aztec data.

[10] See IPE Updated Report, (Exhibit DOM-303), pp. 192-194.

### 1.1.2 Aztec Scanner

10. The experts of the Dominican Republic and Indonesia do not present a graphical analysis of downward substitution using the Aztec data. Instead, the IPE and Professor List use the Aztec scanner data in the context of econometric studies, which will be reviewed in detail next. But in that context, Professor List presents the evolution of the ratio of higher- to low-priced cigarettes based on the Aztec data, which, as shown in **Figure E.3**, displays a downward trend and continues to decrease in the post-TPP period.[11]

**Figure E.3: Weekly Sales Volumes Ratio of Higher-to Low-Priced Sticks Based on Aztec Data**




Note: Conversion rate of 0.8 grams of fine-cut tobacco per stick. The vertical dashed line indicates the introduction of plain packaging and enlarged GHWs.

Source: List Rebuttal Report, (Exhibit DOM/IDN-3), p. 29.

### 1.1.3 Roy Morgan Single Source

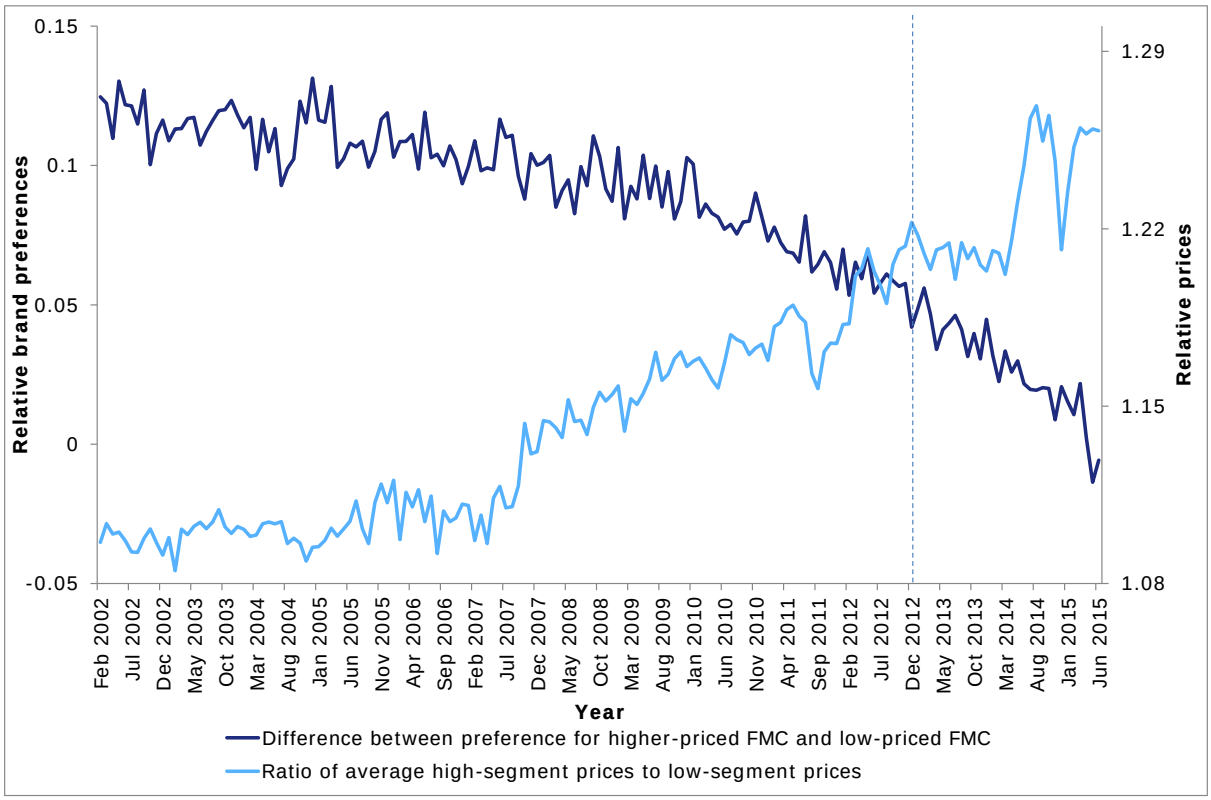
11. In addition to market data, the IPE considers the RMSS survey data and presents, as shown in **Figure E.4**, the evolution of the difference between the share of smokers preferring brands from the higher-priced segment and the share of smokers preferring brands from the low-priced segment.[12] According to the IPE, the share of higher-priced cigarette smokers relative to smokers of low-priced cigarettes has been declining, with very little development in relative brand preferences and relative prices between 2002 up to mid-2007. Thereafter, relative prices rose and relative brand preference declined, implying an inverse relationship between relative brand preferences and relative prices between mid-2007 and late 2012. In the post-TPP period, relative brand preference has continued to decline.[13]

[11] See List Rebuttal Report, (Exhibit DOM/IDN-3), pp. 28-29. We note, however, that unlike IPE, Professor List does not include the sales of RYO tobacco in his analysis of the Aztec data.

[12] See IPE Report, (Exhibit DOM-100), p. 90; and IPE Updated Report, (Exhibit DOM-303), para. 89.

[13] See IPE Report, (Exhibit DOM-100), pp. 87-90; and IPE Updated Report, (Exhibit DOM-303), paras. 217-221.

**Figure E.4: Monthly Brand Preference for Higher-Priced versus Low-Priced Brands and Ratio of Average Higher to Low Prices Based on RMSS Data**




Note: The vertical dashed line indicates the introduction of plain packaging and enlarged GHWs.

Source: IPE Updated Report, (Exhibit DOM-303), p. 89.

## 1.2 Analysis by the Panel

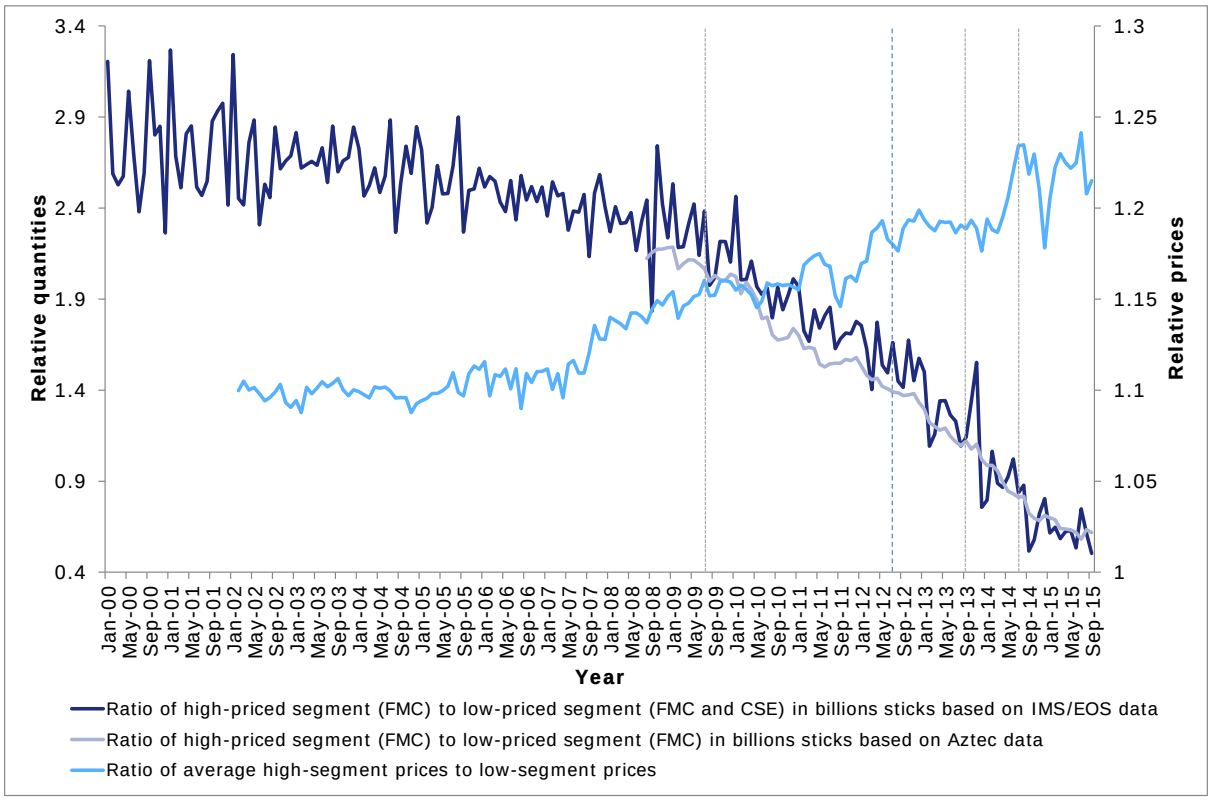
12. As in the case of analysing smoking prevalence and cigarette consumption, we acknowledge the importance of analysing the trends of the ratio of higher- to low-priced cigarettes and the difference in share of higher-priced cigarette smokers relative to smokers of low-priced cigarettes with the most recent available data. As explained in the analysis of cigarette consumption, we recognize that there is no perfect market sales data in terms of market coverage, frequency (weekly, monthly or annually) and period covered.[14] We also note, at the outset, that neither the IPE nor Professor List subsequently updated their graphical analysis of the ratio of higher- to low-priced cigarettes with more recent data.

13. After a careful review of the most recent available IMS/EOS and Aztec datasets, we notice that despite fluctuations, the ratio of higher- to low-priced cigarettes sales has experienced a downward trend in the period following the introduction of plain packaging, as depicted in **Figure E.5**.[15] Conversely, we note that the ratio of the average higher-segment prices to low-segment prices has on average increased over the same period. Furthermore, we note that the co-relation between the ratio of higher- to low-priced cigarettes sales and the ratio of the average higher-segment prices to low-segment prices is strong and negative.

[14] See IPE Updated Report, (Exhibit DOM-303), para. 137; and IPE Summary Report, (Exhibit DOM-379), paras. 131-133. In particular, the IMS/EOS data cover sales from manufacturers to wholesalers and retailers, while the Aztec data cover sales from retailers to consumers. Similarly, the Aztec dataset is only available from 27 July 2008 to 27 September 2015, while the IMS/EOS data cover a larger sample period, from January 2000 to September 2015. Likewise, the IMS/EOS data cover almost 99% of the Australian market, while the Aztec data cover 67% of the Australian market.

[15] The weekly Aztec data were aggregated to monthly data by assuming that each weekly cigarette quantity can be evenly split for each day of the week.

**Figure E.5: Graphical Assessment of Monthly Sales Volumes Ratio of Higher- to Low-Priced Sticks Based on IMS/EOS and Aztec Data**




Note: Conversion rate of 0.8 grams of fine-cut tobacco per stick. The vertical dotted lines indicate excise tax hikes. The vertical dashed line indicates the introduction of plain packaging and enlarged GHWs.

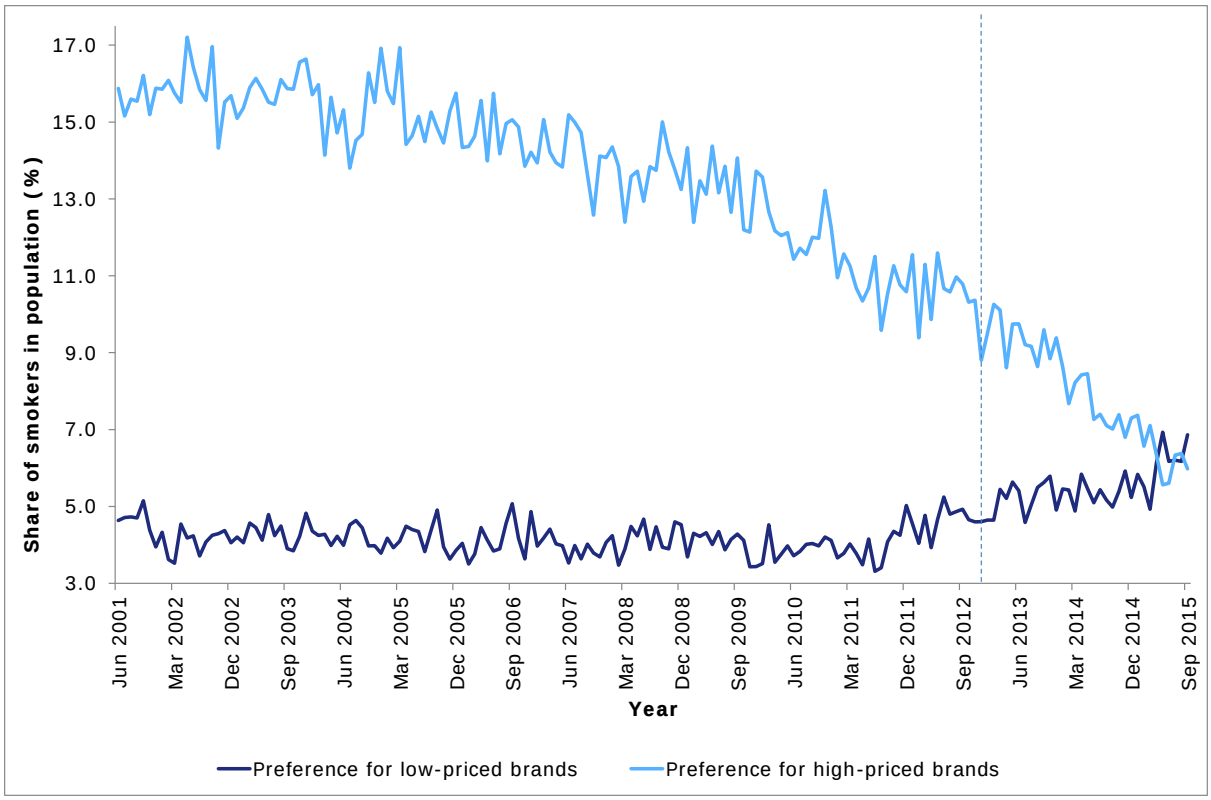
Source: IPE Third Updated Report, (Exhibit DOM-375), backup material.

14. The evolution of the shares of smokers preferring higher-priced brands and those preferring low-priced brands, as depicted in **Figure E.6**, leads us to qualify our previous conclusion. In particular, we note that the share of smokers preferring higher-priced brands has decreased, while the share of smokers preferring low-priced brands has experienced a small but positive increase. In addition, the share of smokers preferring higher-priced brands has, on average, decreased at a much faster rate than the share of smokers preferring low-priced brands has increased, confirming the decrease in smoking prevalence.[16] This explains why, as shown in **Figure E.7**, the difference between the shares of smokers preferring higher-priced brands and low-priced brands has, on average, continued to decrease after the introduction of the TPP measures, such that in May, June and September 2015, the share of smokers preferring low-priced brands is higher than the share of smokers preferring higher-priced brands.

15. We further note that the downward trend of relative preferences based on the RMSS data is slightly more pronounced that the downward trend of relative quantities based on the IMS/EOS and Aztec data. We next review whether the downward trend of the ratio of higher- to low-priced cigarettes sales and the difference between the shares of smokers preferring higher-priced brands and those preferring low-priced brands have accelerated in the post-TPP period.

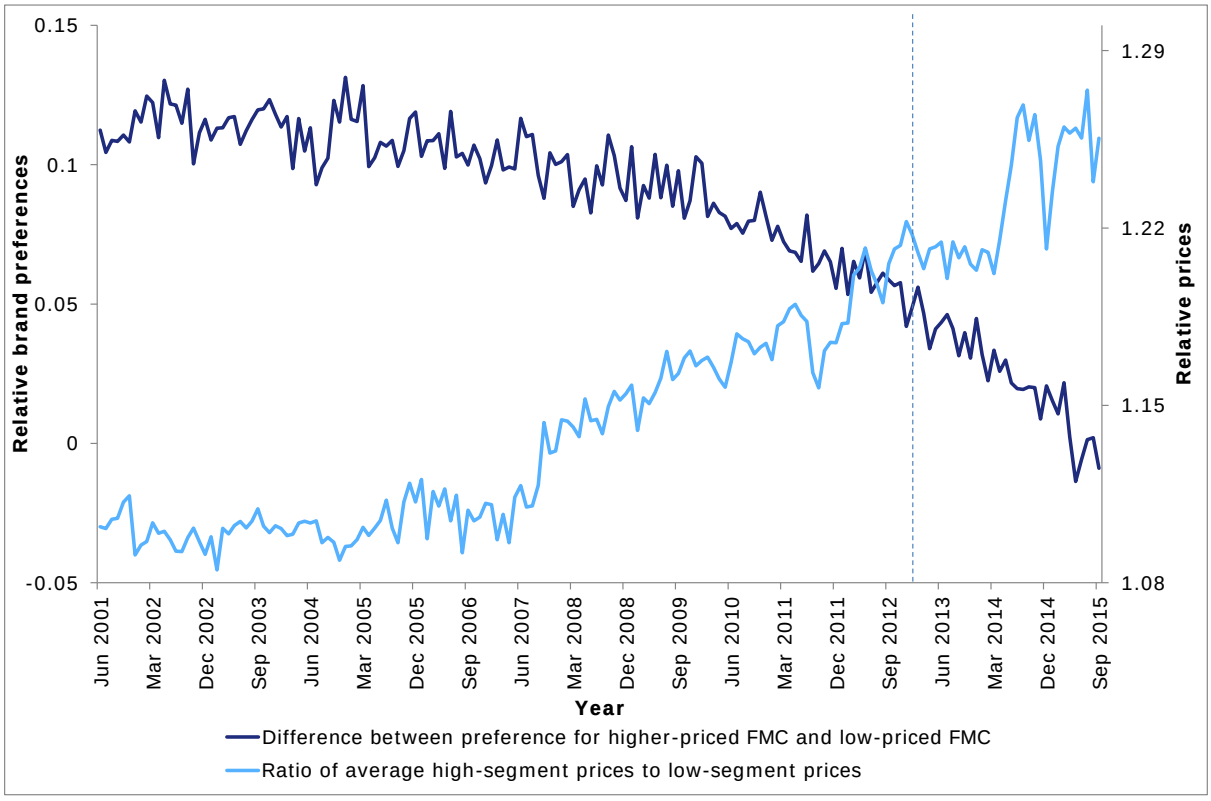
[16] The growth rate differential between the shares of smokers preferring higher-priced brands and those preferring low-priced brands is equal to 2.7 for the period from June 2001 to September 2015. The growth rate differential is equal to 2.3 for the period from June 2007 to September 2015. See IPE Third Updated Report, (Exhibit DOM-375), back-up material.

**Figure E.6: Graphical Assessment of Brand Preference Based on RMSS Data**




Note: The vertical dashed line indicates the introduction of plain packaging and enlarged GHWs.
Source: IPE Third Updated Report, (Exhibit DOM-375), back-up material.

**Figure E.7: Graphical Assessment of Difference in Brand Preference Based on RMSS Data**




Note: The vertical dashed line indicates the introduction of plain packaging and enlarged GHWs.
Source: IPE Third Updated Report, (Exhibit DOM-375), back-up material.

## 2 WHETHER DOWNWARD SUBSTITUTION IN CIGARETTES ACCELERATED AFTER THE ENTRY INTO FORCE OF THE TPP MEASURES

16. We determined that, in the period following the introduction of the TPP measures, the most recent market and survey data show two trends. First, there is a continuing decrease of the ratio of higher- to low-priced cigarettes sales. Second, there is a decrease in the share of smokers preferring higher-priced brands to those preferring low-priced brands. We now turn to consider whether there was a shift in these trends in the post-TPP implementation period. The Dominican Republic's and Indonesia's experts examine this issue by applying different methodologies, which are described next. We then present our analysis.

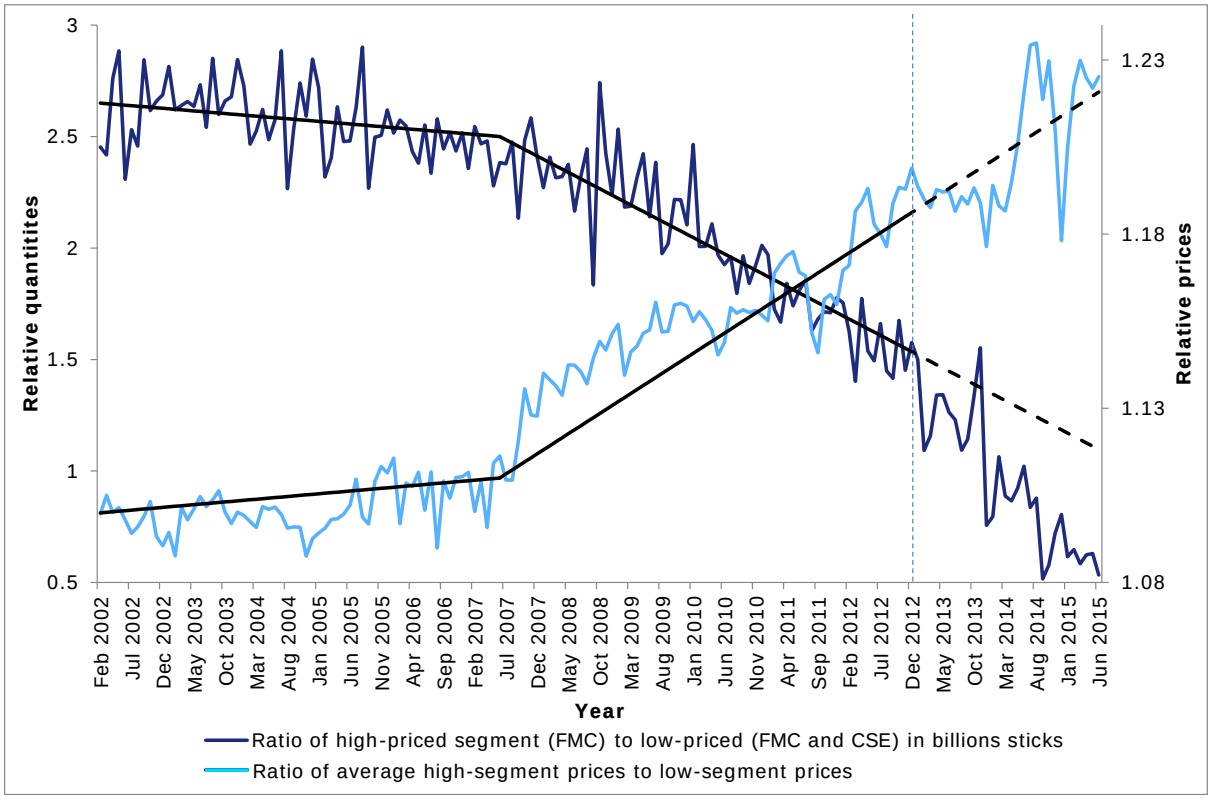
### 2.1 Datasets and related studies

#### 2.1.1 In-Market-Sales/Exchange of Sales

##### 2.1.1.1 IPE Reports

17. Based on the graphical assessment discussed above[17], the IPE submits that since the implementation of the TPP measures in December 2012, relative quantities of higher- to low-priced cigarettes sales have been significantly lower than the interpolated pre-plain packaging trend, as depicted in **Figure E.8**.[18]

**Figure E.8: Trend Assessment of Monthly Sales Volumes Ratio of Higher- to Low-Priced Sticks and Ratio of Average Higher to Low Prices Based on IMS/EOS Data**




Note: Conversion rate of 0.8 grams of fine-cut tobacco per stick. The vertical dashed line indicates the introduction of plain packaging and enlarged GHWs.

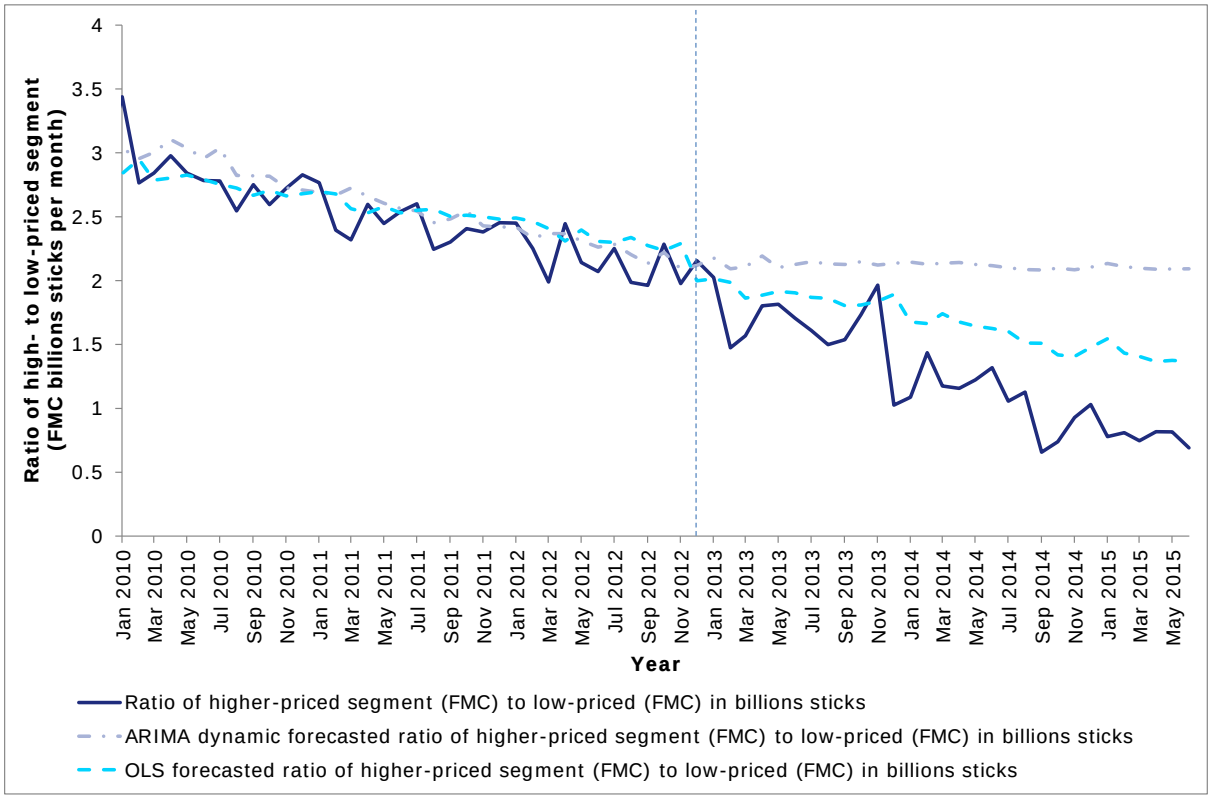
Source: IPE Updated Report, (Exhibit DOM-303), p. 75.

[17] See paras. 8 and 9 above.
[18] See IPE Updated Report, (Exhibit DOM-303), para. 194.

#### 2.1.1.2 Professor List's report

18. Professor List presents the results of an event study analysing whether there has been a shift in the evolution of the ratio of higher- to low-priced cigarettes sales following the introduction of plain packaging and enlarged GHWs.[19] Specifically, a seasonally adjusted Autoregressive Integrated Moving Average (ARIMA) model is estimated controlling for the change in the ratio of higher- to low-priced cigarettes sales during the previous month and the change in the 2006 GHWs. A dynamic model of the ratio of higher- to low-priced cigarettes sales is also estimated controlling for the quantity ratio of higher- to low-priced cigarettes sales during the previous month, the price ratio of higher-to low priced cigarettes in the previous month, a linear time trend, and the 2006 GHWs.[20] The most recent analysis covers the period from February 2002 to June 2015.

**Figure E.9: Event Study on Sales Volumes Ratio of Higher- to Low-Priced Sticks Based on IMS/EOS Data**




Note: Conversion rate of 0.8 grams of fine-cut tobacco per stick. The vertical dashed line indicates the introduction of plain packaging and enlarged GHWs.

Source: List Rebuttal Report, (Exhibit DOM/IDN-3), p. 28.

[19] We note that Professor List's event study is a trend-projection analysis that does not estimate directly the impact of the TPP measures, but rather assesses whether there is a difference between the counterfactual ratio of higher- to low-priced cigarettes sales in the absence of the TPP measures and the actual ratio of higher- to low-priced cigarettes sales. That is why we have decided to discuss Professor List's event study under the subsection addressing whether there was a change in downward substitution in cigarettes in the post-implementation period.

[20] An event analysis consists of (1) estimating the model of the ratio of higher- to low-priced cigarettes sales during the pre-plain packaging period; (2) forecasting the ratio of higher- to low-priced cigarettes sales that would have been prevailed in the absence of the TPP measures using the estimated model's parameters in the post-implementation period; and (3) determining whether the difference between the observed ratio of higher- to low-priced cigarettes sales volumes and the estimated counterfactual ratio of higher- to low-priced cigarettes volumes is statistically different from zero. See List Rebuttal Report, (Exhibit DOM/IDN-3), para. 109.

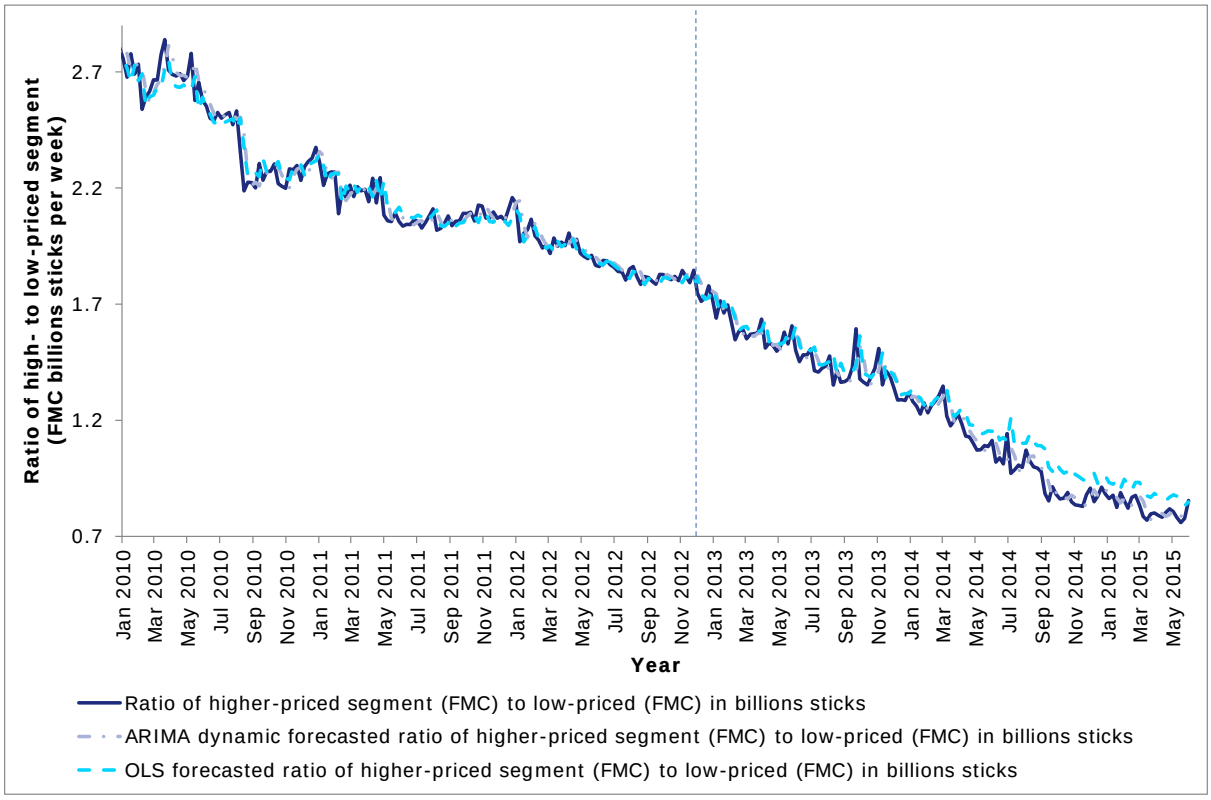
19. Professor List concludes that according to both models, there is a negative and statistically significant difference between the observed ratio of higher- to low-priced cigarettes sales and the estimated counterfactual ratio of higher- to low-priced cigarettes sales, as depicted in **Figure E.9**. According to Professor List, the fact that the ratio of higher- to low-priced cigarettes sales did decrease faster in the post-implementation period provides evidence that the TPP measures have led to important changes to the composition of sales in the Australian market, which is consistent with Australian consumers "downtrading" when making their cigarette choices.[21]

### 2.1.2 Aztec Scanner

20. Professor List carried out the same event study using the Aztec scanner data. A seasonally adjusted ARIMA model and a dynamic model of the ratio of higher- to low-priced cigarettes sales volumes are estimated controlling for change in the quantity ratio of higher- to low-priced cigarettes sales during the previous month as well as the price ratio of the higher- to low-priced cigarettes during the previous month and the linear time trend variable for the dynamic model. The analysis covers the period from July 2008 to May 2015.

21. Professor List reaches the same conclusion that there is a statistically significant decrease in the ratio of higher-priced to low-priced cigarettes sales in the post-implementation period relative to the predicted ratio, as shown in **Figure E.10**, which is consistent with downward substitution from higher-priced to low-priced cigarettes.[22]

**Figure E.10: Event Study on Sales Volumes Ratio of Higher-to Low-Priced Sticks Based on Aztec Data**




Note: The vertical dashed line indicates the introduction of plain packaging and enlarged GHWs.

Source: List Rebuttal Report, (Exhibit DOM/IDN-3), p. 29.

22. Although Professor List acknowledges that it is impossible to separate the GHW effect from the plain packaging effect without making an assumption on the size of one of the effects, he

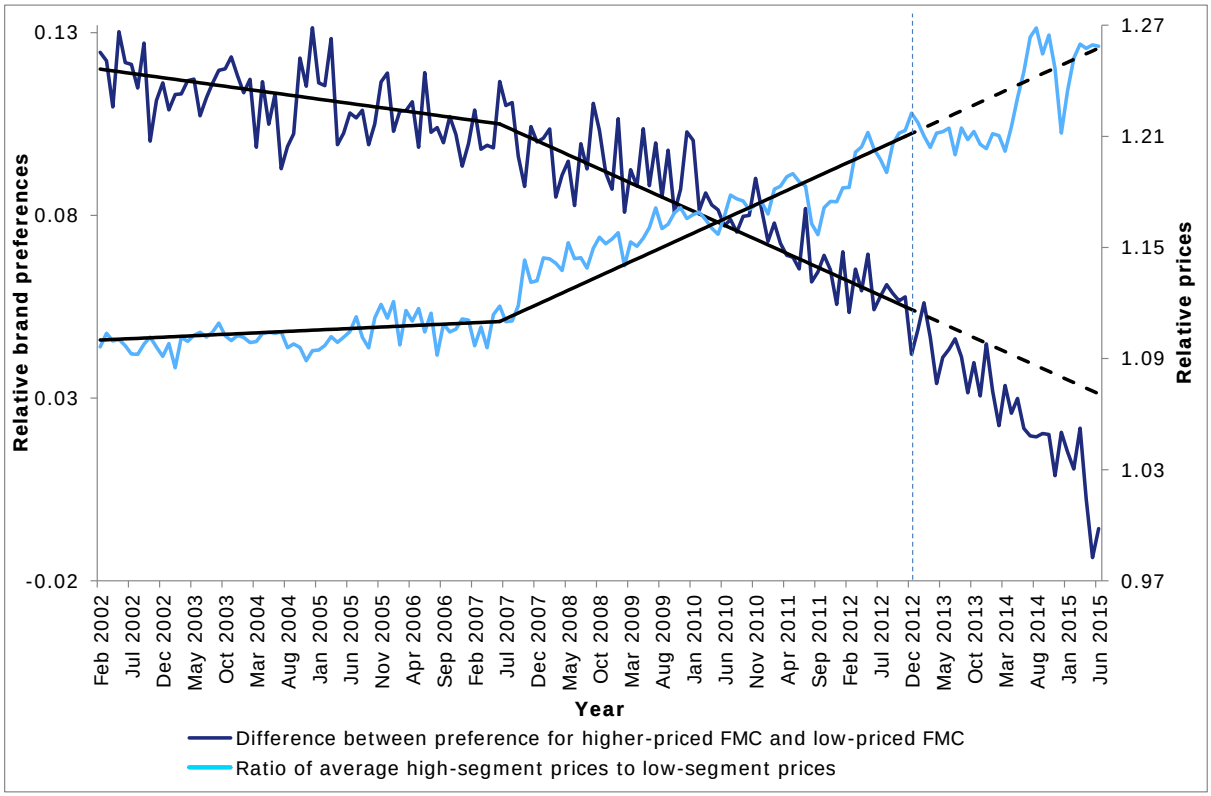
[21] See List Rebuttal Report, (Exhibit DOM/IDN-3), paras. 110-111.
[22] See List Third Supplemental Report, (Exhibit DOM/IDN-7), paras. 56-61 and 80.

proposes to examine the effect of Canada's GHW policy on the composition of higher- and low-priced cigarette sales. In September 2011, Canada mandated an increase in the GHW to 75% of the front and back of the pack, which is, according to Professor List, similar to the GHW increase that coincided with Australia's implementation of its TPP measures in December 2012. Professor List presents the econometric results of an event study of the ratio of higher- to low-tier cigarettes in Canada controlling for the sales composition in the previous month, the price ratio in the previous month, a linear time trend and a dummy variable for the 2011 GHW change. Professor List concludes that the 2011 GHW change in Canada had no impact in the decrease in the ratio of higher- to low-priced cigarette sales. Professor List argues that the evidence of no impact of the GHW on cigarettes sales composition in Canada is consistent with the plain packaging element and not the enlarged GHWs having an important role on the change in the composition of higher- and low-priced cigarette sales observed after the introduction of the TPP measures in Australia.[23]

### 2.1.3 Roy Morgan Single Source

23. Based on the graphical assessment of the RMSS data presented above[24], the IPE submits that shortly after the introduction of the TPP measures, relative brand preferences started to shift downwards to a much greater extent, and at a much greater rate, compared to the pre-plain packaging trend, as depicted in **Figure E.11**.[25]

**Figure E.11: Monthly Preference for Higher-Priced versus Low-Priced Brands and Ratio of Average Higher to Low Prices Based on RMSS Data**




Note: The vertical dashed line indicates the introduction of plain packaging and enlarged GHWs.

Source: IPE Updated Report, (Exhibit DOM-303), p. 89.

[23] See List Rebuttal Report, (Exhibit DOM/IDN-3), paras. 112-118.
[24] See para. 11 above.
[25] See IPE Updated Report, (Exhibit DOM-303), para. 194.

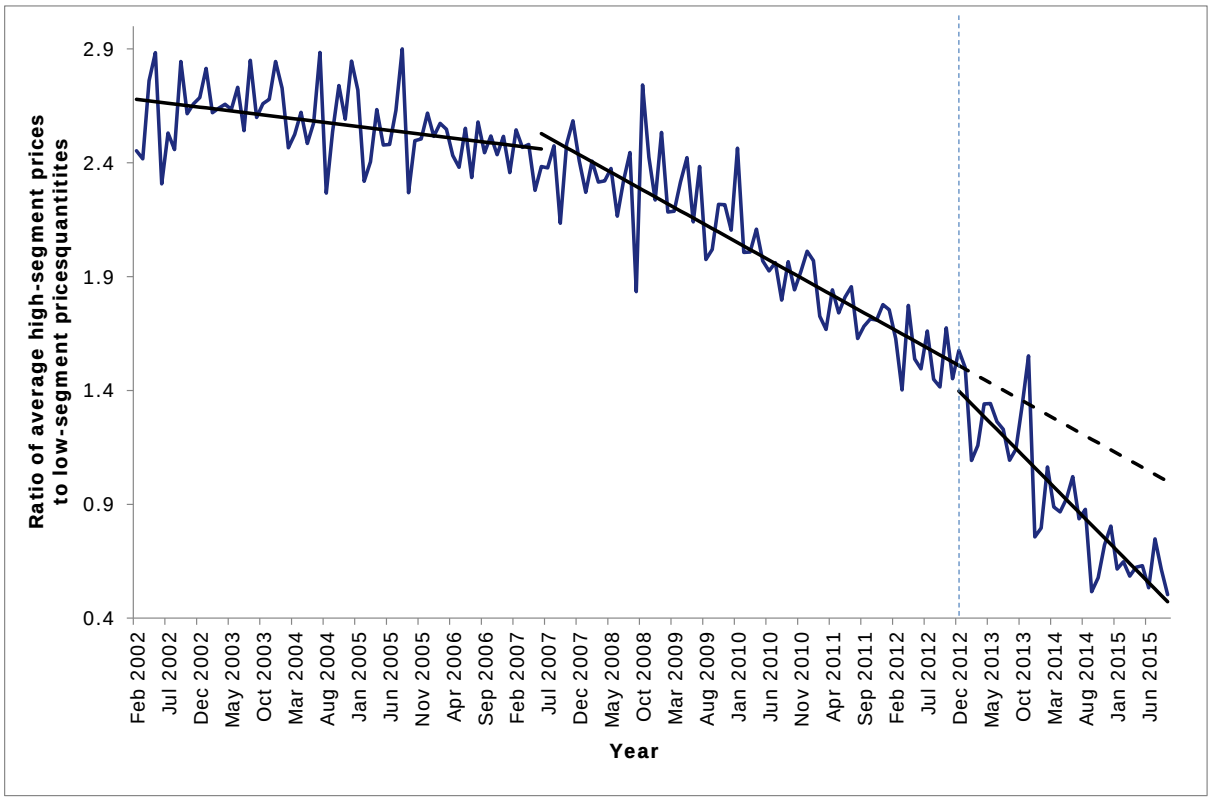
## 2.2 Analysis by the Panel

24. As explained in our analysis of the TPP measures' impact on smoking prevalence and cigarette consumption[26], we are not required to present a unified econometric analysis but to assess the robustness of the results submitted by the parties. In addition, our conclusions apply exclusively to the data (including the sample period), and econometric results submitted by the parties, and cannot in any way be generalized to other datasets and econometric studies.

### 2.2.1 IPE Reports

25. A careful review of the IPE's graphical analysis of the ratio of higher- to low-priced cigarettes sales leads us to the same conclusion reached by the IPE. In particular, extending the analysis to the most recent available data provided by the parties, we find that, according to a standard mean-comparison test, the ratio of higher- to low-priced cigarettes sales based on the IMS/EOS data in the post-implementation period is statistically significantly lower than in the pre-implementation period. This is confirmed by **Figure E.12**, which shows that the trend of relative quantities in the post-implementation has become steeper compared to the pre-implementation trend, implying that the fall in the ratio of higher- to low-priced cigarettes sales has accelerated in the post-implementation period. The same conclusion can be drawn with the Aztec data.

**Figure E.12: Preference for Higher-Priced versus Low-Priced Brands Pre- and Post-TPP Trends Based on RMSS Data**




Note: Conversion rate of 0.8 grams of fine-cut tobacco per stick. The vertical dotted lines indicate excise tax hikes. The vertical dashed line indicates the introduction of plain packaging and enlarged GHWs.

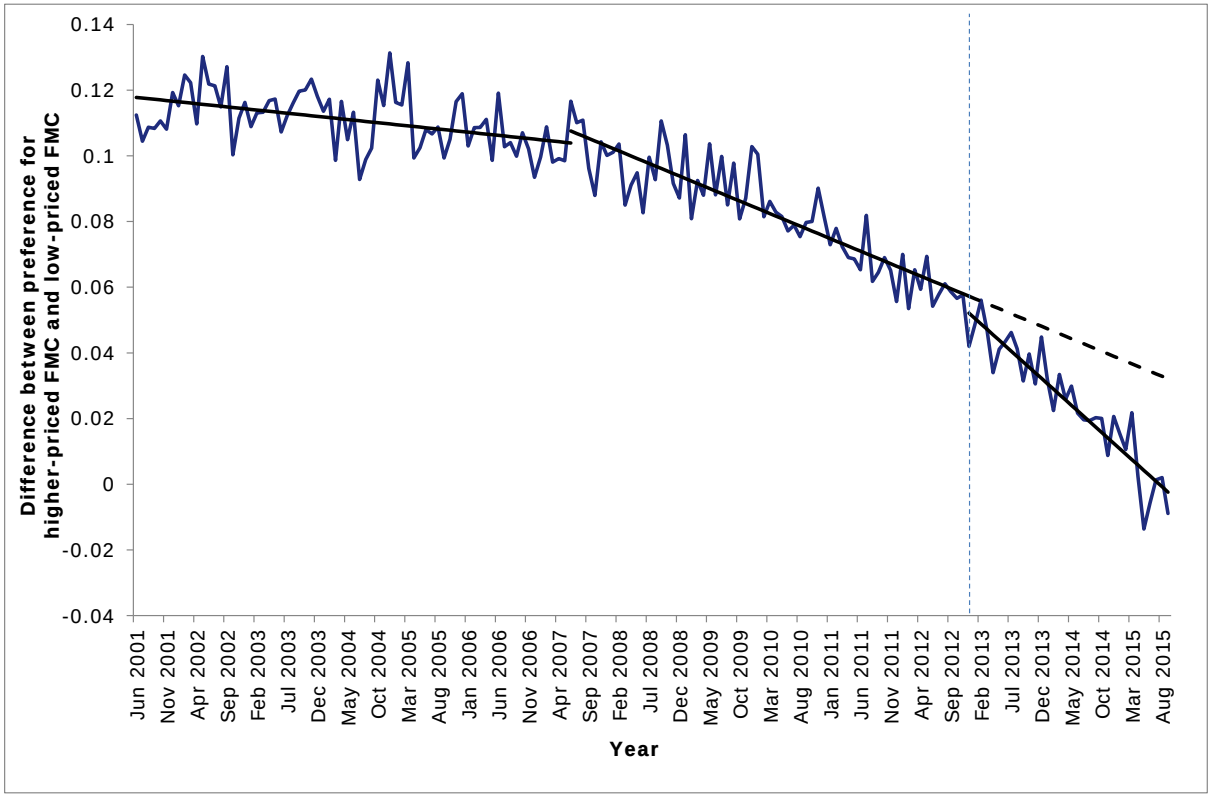
Source: IPE Third Updated Report, (Exhibit DOM-375), backup material.

26. We reach a similar conclusion with respect to the difference between the share of smokers preferring higher-priced brands and those preferring low-priced brands. We note that the downward trend of the difference between smokers preferring higher-priced brands and those

[26] See Appendices C and D.

preferring low-priced brands has accelerated in the post-implementation period, as shown in **Figure E.13**.

**Figure E.13: Sales Volumes Ratio of Higher-to Low-Priced Sticks Pre- and Post-TPP Trends Based on IMS/EOS Data**




Note: Vertical dashed line indicates the introduction of plain packaging (and enlarged graphic warnings).

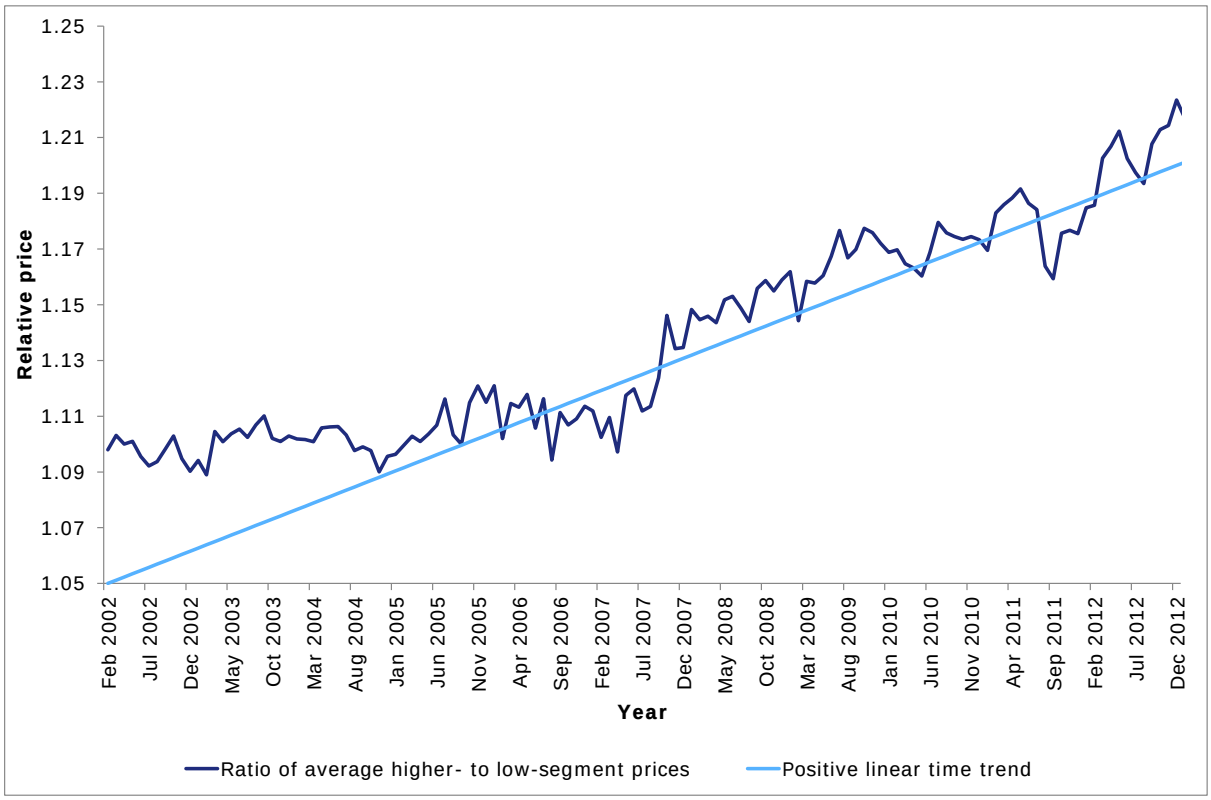
Source: IPE Updated Report, (Exhibit DOM-303), p. 89.

### 2.2.2 Professor List's report

27. After a careful review of the results of Professor List's event study of the impact of the TPP measures on the ratio of higher-priced to low-priced cigarettes sales, we question the validity of Professor List's results for a number of reasons, some of which are specific to one model considered by Professor List.

28. We note that in most estimations of the ARIMA model for the pre-plain packaging period based on the IMS/EOS and Aztec data that are used to forecast the value of the ratio of higher-priced to low-priced cigarettes sales in the post-implementation period, none of the explanatory variables is statistically significant, besides the moving averages parameters. The lack of statistically significant variables is surprising to us given that the ARIMA model only includes one or two explanatory variables. We further note, as in the ARIMA model analysing smoking prevalence, that the estimated lagged moving average parameter of the ARIMA model based on the IMS/EOS data is not statistically different from -1, which would likely lead to large forecasting errors. This is, in our view, problematic because Professor List's approach relies on the post-implementation forecasting errors to determine whether there was a statistically significant change in the downward trend in the ratio of higher-priced to low-priced cigarettes sales.

**Figure E.14: Ratio of Average Higher to Low Prices and Linear Time Variable Based on RMSS Data**




Source: List Rebuttal Report, (Exhibit DOM/IDN-3), backup material.

29. We also note that the dynamic model faces multicollinearity problems between the time trend variable and the price ratio variable when using the IMS/EOS data and between the time trend variable and the ratio of higher-priced to low-priced cigarettes sales during the previous month when using the Aztec data.[27] In other words, the price ratio variable and the time trend variable seem to convey the same information, as shown in **Figure E.14**, and as a result one of them becomes redundant in the model specification with the IMS/EOS data. A similar conclusion applies to the time trend variable and the ratio of higher-priced to low-priced cigarettes sales during the previous month with the Aztec data. In both cases, this multicollinearity problem could explain why the dynamic model finds that the lagged price ratio variable is not statistically significant. This is particularly surprising given that one of the Dominican Republic's experts, the IPE, considers relative prices to be one of the factors leading to downward substitution from higher-priced to low-priced cigarettes.[28] We also question why Professor List implicitly assumes in the models based on the IMS/EOS data that the effect of the 2006 GHW disappears following the introduction of the TPP measures. More generally, Professor List did not take into account the impact of the excise tax hikes in the pre- and post-plain packaging periods. Yet, according to Australia, tobacco manufacturers operating in Australia have attributed downtrading effects to increases in excise taxes.[29]

30. We also question the relevance of using Canada's GHW policy change to infer that plain packaging and not the enlarged GHWs had an impact on the ratio of higher- to low-tier cigarettes in Australia. Professor List did not provide evidence that would justify making such an inference, although he acknowledges that the empirical estimates for Canada rest on strong assumptions.[30]

[27] Evidence of multicollinearity is confirmed by the variance inflation factors statistic.
[28] See IPE Updated Report, (Exhibit DOM-303), para. 194.
[29] See Australia's first written submission, paras. 544-545; Australia's second written submission, paras. 418-419; BATA Media Release, (Exhibit AUS-255); and BAT, Half-Year Results 2015, (Exhibit AUS-556).
[30] List Rebuttal Report, (Exhibit DOM/IDN-3), para. 118.

31. Overall, and in this context, we consider the results of Professor List's event study to be of limited use in helping to answer the question whether the TPP measures led to a change in the downward trend of the ratio of higher-priced to low-priced cigarettes sales. That being said, and as explained above, we find that, according to a standard mean-comparison test, the average cigarette sales volumes based on the IMS/EOS data and the Aztec data in the post-implementation period are statistically significantly lower than in the pre-implementation period. Similarly, the downward trend of the difference between the shares of smokers preferring higher-priced brands and those preferring low-priced brands based on the RMSS data has accelerated in the post-plain packaging period.

32. In any event, as in respect of smoking prevalence and smoking consumption, the fact that the downward trends of the ratio of higher- to low-priced cigarettes sales and the difference between the share of smokers preferring higher-priced brands and those preferring low-priced brands have accelerated in the post-implementation period does not necessarily imply that the TPP measures are having a statistically significant impact, given that other factors, unrelated to the TPP measures, could explain the respective trend shift. The extent to which the TPP measures have an impact on the ratio of higher-priced to low-priced cigarettes sales and on the difference between the shares of smokers preferring higher-priced brands and low-priced brands is reviewed next.

## 3 WHETHER THE TPP MEASURES CONTRIBUTED TO DOWNWARD SUBSTITUTION IN CIGARETTES

33. As explained above, we have determined that the ratio of higher-priced to low-priced cigarettes sales and the difference between the shares of smokers preferring higher-priced brands and those preferring low-priced brands in Australia have continued to experience a decline, which has accelerated in the post-plain packaging period. To the extent that there has been a greater reduction in these metrics after the entry into force of the TPP measures, the question arises whether, and if so, to what extent, plain packaging contributed to reducing these metrics.

34. As for the analyses of smoking prevalence and cigarette sales and consumption, the Dominican Republic's and Indonesia's experts proposed different econometric methods to estimate the impact of the TPP measures on relative quantities and preferences of higher-priced cigarettes with respect to low-priced cigarettes. As mentioned above, the Dominican Republic's and Indonesia's experts recognize, however, that it is impossible to technically distinguish between the impact of plain packaging <u>and</u> the impact of the enlarged GHWs on relative quantities and preferences of higher-priced cigarettes with respect to low-priced cigarettes, because both measures were implemented at the exact same time.[31]

### 3.1 Datasets and related studies

#### 3.1.1 In-Market-Sales/Exchange of Sales

35. Several reports by the IPE have been submitted, estimating econometrically the impact of the TPP measures on the logarithm of the ratio of higher- to low-priced cigarette wholesale sales using the IMS/EOS data. Throughout these reports, different econometric approaches and model specifications have been proposed: (1) time series regression analysis; (2) modified trend analysis; and (3) Autoregressive Integrated Moving Average with Explanatory Variable (ARIMAX) model.[32]

36. The first approach was only presented in the first IPE Report and is based on a model specification of the logarithm of the ratio of higher- to low-priced cigarette sales that includes a TPP measures dummy variable, dummy variables for the 2010 and 2013 excise tax increases, a 2006 GHW dummy, the logarithm of the higher- to low-priced cigarettes price ratio, month fixed effects and a time trend variable. The second and subsequent IPE reports only focused on the last

[31] See Dominican Republic's response to Panel question No. 8, para. 61; List Report, (Exhibit DOM/IDN-1), para. 113; Honduras's response to Panel question No. 8; and Indonesia's response to Panel question No. 8, para. 8.

[32] See IPE Report, (Exhibit DOM-100), pp. 67-73 and 201-210; and IPE Updated Report, (Exhibit DOM-303), paras. 135-175 and 290-326.

two approaches. The modified trend analysis estimates a standard time series regression model of the logarithm of the ratio of higher-to lower-priced cigarette sales using (heteroscedastic) autocorrelation-robust standards errors and controlling for a TPP measures dummy variable, dummy variables for the 2010, 2013 and 2014 excise tax increases, a 2006 GHW dummy, the (current or lagged) logarithm of the higher- to low-priced cigarettes price ratio, month fixed effects and a time trend variable. The ARIMA model includes the same set of explanatory variables as well as the (one up to four) lagged logarithm of the ratio of higher-to lower-priced cigarette sales and the (one up to four) lagged logarithm of the price ratio. The most recent econometric analysis covers the period February 2002 through August 2015.

37. Overall, the IPE concludes that the TPP measures had a negative and statistically significant effect on the logarithm of the ratio of higher- to low-priced cigarette sales. According to the IPE, these results are robust across different specifications (e.g. different plain packaging starting date: October, November and December; alternative time trend specification: linear, quadratic and cubic; and alternative methods of calculating standard errors of the modified trend analysis).[33]

38. Although the IPE acknowledges that there is no perfect way to separate the plain packaging and enlarged GHW effects, it argues that there are indirect ways to study any effect of the enlarged GHW requirement on downtrading in Australia, such as estimating the downtrading effect of another tobacco control policy introduced in isolation in Australia, without the simultaneous introduction of a confound. The IPE proposes using the introduction of GHWs in Australia in 2006 as a proxy for the implementation of the enlarged GHW requirement in December 2012. Given that the econometric results of the analysis of the logarithm of the ratio of higher- to low-priced cigarette wholesale sales find that the impact of the 2006 GHW is negative but not statistically significant, the IPE argues that the plain packaging component is likely responsible for the vast majority of the downtrading effect found from the December 2012 policy change.[34]

### 3.1.2 Aztec Scanner

39. The IPE performed the same (1) time series regression analysis; (2) modified trend analysis; and (3) ARIMAX model analysis using the Aztec scanner data.[35] The exact same set of explanatory variables is included, except the 2006 GHW dummy. The analysis covers the period from July 2008 to May 2015.

40. Overall, the IPE finds the same conclusion as the analysis based on the IMS/EOS data, namely the impact of the TPP measures on the logarithm of the ratio of higher- to low-priced cigarette retail sales is negative and statistically significant. According to the IPE, these results are robust across different specifications (e.g. different plain packaging starting date: October, November and December; alternative time trend specification: linear, quadratic and cubic; and alternative methods of calculating standard errors of the modified trend analysis).[36] In reality, the estimated coefficient of the TPP measures dummy variable is not statistically significant in a couple of specifications.[37]

### 3.1.3 Roy Morgan Single Source

41. The RMSS data were used by the Dominican Republic's and Indonesia's experts to estimate econometrically the impact of the TPP measures on the relative preferences of higher-priced cigarettes with respect to low-priced cigarettes.

---

[33] See IPE Report, (Exhibit DOM-100), pp. 83-86 and 223-233; IPE Updated Report, (Exhibit DOM-303), paras. 195-201; IPE Second Updated Report, (Exhibit DOM-361), paras. 45-48; and IPE Third Updated Report, (Exhibit DOM-375), paras. 249-251.

[34] See IPE Updated Report, (Exhibit DOM-303), paras. 228-234. IPE also presented a graphic analysis of the relative quantities with respect to the introduction of GHWs in 2006 and plain packaging in 2012.

[35] See IPE Report, (Exhibit DOM-100), pp. 80-83; and IPE Updated Report, (Exhibit DOM-303), paras. 205-206.

[36] See IPE Report, (Exhibit DOM-100), pp. 234-244; IPE Updated Report, (Exhibit DOM-303), paras. 202-208; IPE Second Updated Report, (Exhibit DOM-361), paras. 49-52; and IPE Third Updated Report, (Exhibit DOM-375), paras. 252-254.

[37] This is the case when the starting date of the TPP measures dummy variable is set to October 2012. See IPE Third Updated Report, (Exhibit DOM-375), p. 140.

#### 3.1.3.1 IPE Reports

42. The IPE presented in various reports the results of a standard probit model of the probability of an individual cigarette smoker consuming higher-priced cigarettes. The model controls for a TPP measures dummy variable, (current or lagged) relative price of higher-priced versus low-priced cigarettes, gender dummy, polynomial of degree 4 of the age variable, polynomial of degree 4 of the education variable, polynomial of degree 4 of the income group variable, dummy variables for the 2010, 2013, 2014 and 2015 excise tax increases, and a trend variable.[38] In its first report, the IPE also estimated a linear probability model, whose model specification included the same demographic variables as well as a social class variable but without any of the tax hike dummy variables and the trend variable.[39] The analysis covers the period from July 2006 through September 2015.

43. Overall, the IPE concludes that the impact of the TPP measures on the probability that an individual smoker prefers higher-priced brands is negative and statistically significant.[40] According to the IPE, this finding is robust to different model specifications (e.g. different plain packaging starting date: October, November and December, inclusion of the first lagged of the price ratio, and excise tax increases dummies).

44. The IPE further submits that plain packaging and not the enlarged GHWs is likely responsible for the vast majority of the downtrading effect from the introduction of the TPP measures because a modification of the model specification enabling to extend the sample period from February 2002 to September 2015 and controlling for the 2006 GHW finds a negative but not statistically significant impact of the 2006 GHWs on the probability of smoking cigarettes from the higher-priced segment.[41]

#### 3.1.3.2 Professor List's Report

45. Professor List presented econometric results of a two-stage micro-econometric model based on the RMSS data. The first stage estimates the likelihood of a representative person smoking a high or medium-priced segment brand of cigarettes, conditional on being a smoker, in a given month and controlling for demographic characteristics (e.g. age, fourth order polynomials of gender, education, and income position). The second stage conducts a before and after analysis on the likelihood for each month computed in the first stage by estimating a linear probability model controlling for the TPP measures, higher- to low-priced cigarettes price ratio and/or linear trend, and weighting changes by Roy Morgan Research. Professor List re-estimated the same model for a representative person smoking a low-priced segment brand of cigarettes. The econometric analysis covers the period from July 2006 to June 2015.

46. Overall, Professor List concludes that the TPP measures had a statistically significant negative effect on the likelihood of consuming higher-priced cigarettes but a positive and statistically significant effect on the likelihood of consuming lower-priced cigarettes.[42]

### 3.2 Analysis by the Panel

47. Having determined above that the decrease in the ratio of higher- to low-priced cigarette sales and in the difference between the share of smokers preferring higher-priced brands and the share of those preferring low-priced brands has accelerated following the entry into force of the TPP measures, the question before us at this stage of our analysis is whether this acceleration may, in part or in whole, be attributed to the TPP measures.

---

[38] See IPE Updated Report, (Exhibit DOM-303), paras. 245-250.

[39] See IPE Report, (Exhibit DOM-100), pp. 245-250.

[40] See IPE Report, (Exhibit DOM-100), pp. 90-94; IPE Updated Report, (Exhibit DOM-303), paras. 223-227, p. 134; IPE Second Updated Report, (Exhibit DOM-361), paras. 53-55; and IPE Third Updated Report, (Exhibit DOM-375), paras. 255-256.

[41] See IPE Updated Report, (Exhibit DOM-303), paras. 231-235, p. 134.

[42] See List Rebuttal Report, (Exhibit DOM/IDN-3), paras. 106-107.

48. At the outset, we note, as explained above, that Australia did not provide any econometric evidence on downward substitution. We also note that Dominican Republic's and Indonesia's experts use the same data sources but consider different model specifications and approaches.[43]

49. As explained in our review of the econometric study on smoking prevalence and cigarette sales and consumption, it is not our task to conduct our own econometric assessment but rather review the robustness of the econometric evidence submitted by each party.

### 3.2.1 IPE's econometric results

50. A careful review of the econometric results based on the IMS/EOS data provided in the IPE reports leads us to conclude that there is some evidence, albeit limited, that the TPP measures, together with the englarged GHWs introduced on the same date, appear to have had a negative impact on the ratio of higher- to low-priced cigarette wholesale sales. We note, however, that the price ratio has a much larger impact on the ratio of higher- to low-priced cigarette wholesale sales. This finding is in line with the strong negative correlation between price ratio and quantity ratio, as well as the evidence submitted by Australia regarding the impact of the tobacco industry's own marketing and pricing strategies on downtrading in the Australian market.[44] Similarly, the econometric results show that the overall impact of the excise tax hikes on the ratio of higher- to low-priced cigarette wholesale sales is larger the TPP measures.

51. That being said, we note that some specifications of the modified trend analysis are characterized by multicollinearity between the quadratic or cubic trend variables and the current and/or lagged price ratio. We note that the issue of multicollinearity is even more severe in all of the specifications of the ARIMAX model based on the IMS/EOS data. The multicollinearity is particularly high between the quadratic or cubic time trend variables, the lagged price ratio and lagged quantity ratio.[45] We also question the results of most ARIMAX models, because the results find that the higher- to low-priced cigarettes price ratio has either no impact or a statistically positive impact. It is unclear what drives such results, although we note that the price ratio is negative and statistically significant in the specifications, which are less affected by multicollinearity by including only a linear time variable. The results of some specifications also suggest that the trend variable has a positive and statistically significant impact, which is in contradiction with other econometric results and the graphical analysis discussed above. We therefore reject the results of the ARIMAX model based on the IMS/EOS data.

52. We also question the validity of the econometric results based on the Aztec retail data on the same grounds. We note that the results of the modified trend analysis are affected by high collinearity between the linear trend variable and the current price ratio as well as the price ratio in the previous month. The collinearity problem also arises in the ARIMAX model between the logarithm of the ratio of higher- to low-priced cigarette retail sales in the previous month, the trend variable(s) and the price ratio in the previous (second to fourth) months. The collinearity is even more severe when the current price ratio is included in the model specification. We also note that all the results of the ARIMAX model find that the impact of the logarithm of the ratio of higher- to low-priced cigarette wholesale sales in the previous month is positive and statistically significant, which is in contradiction with the negative and statistically significant impact found in the econometric analysis based on the IMS/EOS data.

53. Our review of the IPE's micro analysis of the effect of the TPP measures on smokers' relative preference for higher-priced versus low-priced cigarettes based on the RMSS data leads to question the validity of some of its results. Besides the issue of collinearity between the price ratio and the time trend variable, we note that the impact of the relative price of higher-priced cigarettes with respect to low-priced cigarettes in the current or previous month on the probability of choosing higher-priced cigarettes is never statistically significant. This finding is, in our view, rather surprising given that as acknowledged by the IPE itself, the individual choice of higher-priced versus low-priced cigarettes is likely to be influenced by the relative price of the two

[43] For instance, Professor List does not include dummy variables for the excise tax increases unlike IPE. Similarly, Professor List presents the results of a two-stage econometric model applied to the RMSS data, while IPE presents the results of one-stage micro-econometric analysis.

[44] See Australia's second written submission, paras. 418-419.

[45] Evidence of multicollinearity is confirmed by the variance inflation factors statistic.

product groups.[46] Yet, the IPE did not provide any explanations that would justify such surprising results.

### 3.2.2 Professor List's econometric results

54. We also question the validity of Professor List's two-stage econometric model of the probability of a representative person smoking a high or medium price segment brand of cigarettes based on the RMSS data on various grounds. First, we note that, in the first stage of the model of the representative smoker, more than 60% of the individual monthly estimations report that either none or only one or two of the explanatory variables are statistically significant.[47] Similarly, there is only up to two explanatory variables that are statistically significant in more than 43% of the individual monthly estimations in the first stage of the model of the representative smoker of low-tier cigarettes. It is therefore unclear to what extent the results of the second stage of the analysis are reliable given the lack of statistical significance in the first stage. Second, as explained above, issues of multicollinearity arise between the ratio of higher- to low-priced cigarettes prices and the time trend. Third, several estimation results find surprisingly that the price ratio is statistically not significant. Finally, we note that Professor List does not, unlike the IPE, take into account the impact of the excise tax hikes on the probability of smoking, which is considered by the parties to be one of the most efficient tobacco control policies.

55. Although we question the validity of Professor List two-stage econometric model analysis, we cannot rule out that plain packaging and enlarged GHWs contributed to the reduction in the ratio of higher- to low-priced cigarette wholesale sales in Australia based on the most recent econometric evidence on wholesale data submitted by the IPE. That being said, the extent to which this reduction in the ratio of higher- to low-priced cigarette wholesale sales in Australia that could be attributed to the TPP measures, represents only downtrading is unclear to our eyes. As highlighted in the graphical analysis presented above, the reduction in higher-priced segment wholesale sales has decreased at a much faster rate than the sales of low-priced cigarettes, which implies that at least part of the reduction in the ratio of higher- to low-priced cigarette wholesale sales is due to the overall reduction in the total wholesale sales volume following and due to the introduction of the TPP measures, as we concluded earlier.

## 4 OVERALL CONCLUSION ON POST-IMPLEMENTATION EVIDENCE ON DOWNWARD SUBSTITUTION IN CIGARETTES

56. Overall, based on the most recent data available and econometric evidence submitted by the parties, we find that:

a. There is evidence that the ratio of higher- to low-priced cigarette wholesale and retail sales and the difference between the share of smokers preferring higher-priced cigarettes and the share of those preferring low-priced cigarettes in Australia continued to decrease following the introduction of the TPP measures.

b. The downward trend in the ratio of higher- to low-priced cigarette wholesale and retail sales and the difference between the share of smokers preferring higher-priced cigarettes and the share of those preferring low-priced cigarettes in Australia appears to have accelerated in the post-plain packaging period.

c. Although it is impossible to distinguish between the impact of plain packaging and GHWs, there is some econometric evidence suggesting that the TPP measures contributed to the reduction in the ratio of higher- to low-priced cigarette wholesale sales. That said, it is unclear to what extent this reduction in the quantity ratio attributable to the TPP measures represents only downward substitution. In fact, given that the reduction in higher-priced segment wholesale sales has decreased at a much faster rate than the increase in sales of low-priced cigarettes, at least part of the reduction in the quantity ratio is due to the overall reduction in the total wholesale sales volume following and due to the introduction of the TPP measures and enlarged GHWs,

[46] See IPE Updated Report, (Exhibit DOM-303), paras. 231-235 and 218.

[47] 12 demographic characteristics and the constant term are included in the first-stage model specification.

as we have concluded in our previous analysis on cigarette consumption. In addition, the econometric results show that the increase in relative cigarette price and excise tax hikes have had a negative and greater impact on the the ratio of higher- to low-priced cigarette wholesale sales than the TPP measures.

57. We note, however, that no post-implementation empirical evidence has been presented to us on the impact of the TPP measures on the ratio of higher- to low-priced cigars and cigarillos.

_____