Comments on Professor Ben Jann's Methodological Report

OxyRomandie, 14 December 2015

Summary

We consider that prof. Jann's re-analysis is an improvement over the approach used in the two working papers under consideration. We also applaud his attempt at presenting an objective and balanced treatment of the issue.

We note that prof. Jann is very critical of the two studies. He considers that the design on which they are based is "weak" and that their co-authors could have been "more careful in pointing out this weakness." He further observes that the two co-authors failed to discuss the limitation of their approach. He adds that it is "hard to find truly convincing arguments" for the key assumption on which their analysis is based, analysis which he finds "somewhat confusing" and even "inconsistent" at places. Prof Jann also agrees "that it is odd to exclude December 2012, as plain packaging came into effect in December 2012" referring to a key decision in the study on adults. He states "that better test approaches exist than the one used by Kaul and Wolf," contradicting the claim made by the two professors that no method would achieve better results than theirs.

While we agree with prof. Jann's critical remarks, we nevertheless respectfully fear that he has failed to identify more crucial defects of the papers, by leaving unchallenged some of the key unfounded assumptions on which the co-authors base their analysis. He has used these invalid assumptions in his own re-calculation of their results, thus replicating their errors.

A re-analysis of the data used in the study on adults has been carried out by P. Diethelm and T.M. Farley, using state-of-the art statistical methodology. They arrived at results which *contradict* the findings of the two professors, thus refuting their main conclusion. Diethelm and Farley's re-analysis has been published in a peer reviewed journal.[1]

Out of the seven errors and seven issues we identified in the two working papers, prof. Jann's report provides answers to two of them which we accept; OxyRomandie has no hesitation withdrawing these two points. However, for the remaining 12 points, his explanations do not provide sufficient answers, as he either simply replicates and thus endorses the errors they contain or, when acknowledging them, he unconvincingly minimizes and relativizes their significance.

[1] Diethelm P and Farley TM. Refuting tobacco-industry funded research: empirical data shows decline in smoking prevalence following introduction of plain packaging in Australia. *Tob. Prev. Cessation* 2015;1(November):6 doi: 10.18332/tpc/60650 Available from: https://doi.org/10.18332/tpc/60650

The essential part of our critique remains therefore unchallenged. We persist in our assessment that the two working papers suffer from serious flaws and design misconception that are collectively damning and make them defective beyond repair. Yet it is these very two papers that are exhibited by Philip Morris[2,3] and other tobacco companies as *the* evidence that plain packaging did not work in Australia.

We still think it of crucial importance that the University of Zürich take its distance with respect to the use made by the tobacco sponsor – their partner – of these two defective studies and publicly denounce without ambiguity the misrepresentation of their results, notably in the tobacco multinational's submission to the UK government in response to the 2014 consultation on plain packaging. The contract that links the University to Philip Morris International gives the University the right to do so. Its status as a public academic institution makes it a moral obligation.

We think that, failing to assume its responsibility, the University of Zürich would set an extremely worrying precedent, institutionalizing the complicity of an academic establishment in the manipulation of science by a corporate sponsor. This is implying that as long as the sponsor pays, he owns the results of the studies produced by the university, which are thus considered purely as deliverables, and his ownership extends to the point of being able to distort and misrepresent their findings, without the university feeling any obligation or responsibility to intervene to prevent or stop the disinformation. Such an approach to partnership between the private sector and the university would wide open the door to all kind of abuses and would inevitably undermine public confidence in academic research.

We reiterate what we said in our letter of 29 January: this affair poses the fundamental question of the integrity of science. The University of Zürich should not let the tobacco industry corrupt science and should protect itself against those who want to take advantage of its influence and reputation, not hesitating to put science at the service of money and not heeding the mission entrusted to this public institution. A mission which consists in particular in disseminating a culture founded on scientific knowledge and raising public awareness of the responsibilities that teachers and academic researchers assume towards society.

---

[2] Philip Morris Limited. Response to the Consultation on "Standardised Packaging" 7 August 2014. Available from: http://www.pmi.com/eng/tobacco_regulation/submissions/Documents/UK%20-%20%20Standardised%20Packaging%20Submission%20PML.pdf

[3] Philip Morris Annex 8.1. Overview of the studies showing that there is no evidence that plain packaging has had the desired effect.pdf. Consultation on proposals for introducing standardized tobacco packaging and implementation of Tobacco Convention Article 5.3 in Norway. Available from: https://www.regjeringen.no/no/dokumenter/horing-av-forslag-til-innforing-av-standardiserte-tobakkspakninger-og-gjennomforing-av-tobakkskonvensjonen-artikkel-5.3-i-norge/id2401022/?uid=4d145cdb-ecc1-46d6-90db-653df39c6f09#

Detailed comments (section by section)

Comments on *2. General remarks on the potential of the given data to identify a causal effect of plain packaging*

We note that prof. Jann considers that the design on which the two working papers are based is weak, making it difficult to conclude that there is a causal relationship, and that the authors could have been more careful in pointing out this weakness. We agree. He further observes that the co-authors failed to discuss the limitation of their approach, in that plain packaging was expected to have a slow and subtle effect which a period of observation of one year might be too short to detect.

In fact, far from exercising this elementary caution, if we believe their financial sponsor, they went to the opposite side, saying that "*if there had been an effect in reality (…) it would have been reflected in the data*." [4] This is a crucial point of OxyRomandie's argument, which we think is reinforced by prof. Jann's report

We have the following specific comments:

[Page 4] *Plain packaging, according to Kaul and Wolf, was introduced in December 2012, so that there are 143 months of pre-treatment observations and 13 months of treatment-period observations (assuming that plain packaging, the treatment, was introduced on December 1).*

Plain packaging was introduced in October 2012, giving two months to retailers to get rid of old packs. If one needs to designate a specific month as the onset of the treatment period, then November appears to be the logical choice: more than half of the packs sold in November were plain and calls to the Quitline peaked during that month.

[Page 4] *It is a quasi-experimental design as there is no randomization of the treatment. In general, it is difficult to draw causal conclusions from such a design, as it remains unknown how the counter-factual time trend would have looked.*

We agree. This is something the co-authors should have emphasized in their paper, which would have perhaps made it more difficult for their financial sponsor to misrepresent their results.

[4] Philip Morris Limited. Response to the Consultation on "Standardised Packaging" 7 August 2014. Available from: http://www.pmi.com/eng/tobacco_regulation/submissions/Documents/UK%20-%20%20Standardised%20Packaging%20Submission%20PML.pdf

[Page 4, footnote] *The data appear to stem from weekly surveys, but Kaul and Wolf base their analyses on monthly aggregates. It is not known to me whether Kaul and Wolf had access to the individual level weekly data or only to the monthly aggregates.*

This provides an illustration of our **Issue #5: Contradiction and lack of transparency about the way the data was obtained**.

[Page 5] *Kaul and Wolf assume a linear time trend and hence base their analyses on a linear fit to the pre-treatment data. (…) The assumption behind such an approach is that the time trend would have continued in the same linear fashion as in the pre-treatment period if there had been no treatment. The problem is that it is hard to find truly convincing arguments for why this should be the case (no such arguments are offered by Kaul and Wolf). As argued in the paper by Laverty et al. (forthcoming) it may be equally plausible that the trend would level off (e.g. because the trend has to level off naturally once we get close to zero or because the pre-treatment declines were caused by a series of other tobacco control treatments), or that the trend would accelerate (e.g. due to business cycles or other factors that might influence tobacco consumption). The point is: we simply do not know how the trend would have been like without the treatment.*

This lack of convincing argument for the continuation of the linear trend also applies to the assumption of a linear "*pre-existing*" trend, independent of tobacco control measures, for which it is equally hard to find truly convincing arguments. This is the point we make in our *Error #7: Invalid assumption of long term linearity.*

When they assumed that smoking prevalence in Australia followed of a pre-existing linear trend during the period 2001-2012, Kaul and Wolf contradicted the results presented in two papers by prof. Wakefield and her co-authors, which had clearly shown that tobacco control policy measures were associated with significant and substantial changes in smoking prevalence, the evolution of which was consequently best modelled with a curve that was not linear.[5] Kaul and Wolf do not refer to these two papers, although they were publicly available before they published their working paper on adults, and although they were apparently aware of their existence. Indeed, in their presentation to members of Sir Chantler's review group, they indicated that they knew at least of prof. Wakefield's first paper, when they stated that the Roy Morgan's data set "*has been used, for example, by Wakefield, a very renowned Australian tobacco control advocate, in her research on tobacco control in 2008 and 2009 in published papers, in peer-reviewed published papers.*"[6]

---

[5] Wakefield MA, Durkin S, Spittal MJ, Siahpush M, Scollo M, Simpson JA, et al. Impact of tobacco control policies and mass media campaigns on monthly adult smoking prevalence. *Am J Public Health*. 2008;98:1443-50
and
Wakefield MA, Coomber K, Durkin SJ, et al. Time series analysis of the impact of tobacco control policies on smoking prevalence among Australian adults, 2001–2011. *Bull World Health Organ* 2014;92:413–422

[6] Meeting to discuss "The (Possible) Effect of Plain Packaging on the Smoking Prevalence of Minors in Australia: A Trend Analysis" working paper. Attendees: Kaul A, Wolf M, Cox C, Collis J and Edwards L. King College London, 20 March 2014. Available from: https://www.kcl.ac.uk/health/Packaging-

The lack of convincing arguments has not prevented prof. Wolf from declaring to members of Sir Cyril Chantler's review group: "*I can say from upfront the methodology that we have employed is the one that gives the most leeway to finding an effect, if there had been any."* It has apparently not prevented the two professors from confirming to Philip Morris that *"if there had been an effect in reality (…) it would have been reflected in the data"* [7] (assuming Philip Morris faithfully reported what prof. Kaul and Wolf told them.) This is for us a crucial point (see below).

[Page 6] *…the design on which the working papers by Kaul and Wolf are based on is not particularly strong. Kaul and Wolf cannot be blamed for this as there might have been no better data, but they could have been more careful in pointing out the weaknesses of their design.*

They did not point out the weakness of their design, but on the contrary, prof. Wolf declared to members of Sir Chantler's review group that the method they used was the best: "*I can say from upfront the methodology that we have employed is the one that gives the most leeway to finding an effect, if there had been any."*[8] We agree with prof. Jann that prof. Kaul and Wolf could have been more careful in pointing out the weaknesses of their design. The weakness of the design is puzzling, as the Single Source household survey data from which it is extracted and aggregated is very rich in information. We have contacted Roy Morgan to assess under what conditions OxyRomandie could obtain the data. The first question we were asked was: on which type of smoking we were interested, cigarettes, roll-you own, or all tobacco products. It is strange that prof. Kaul and Wolf did not mention the products being smoked, which could have helped determine whether smokers switched to another product (as is claimed in the press release[9] issued by Philip Morris when they announced the results of prof. Kaul and Wolf's first paper: "*Consumers aren't smoking less, they are just buying cheaper alternatives like roll-you-own cigarettes…*"). They could also have looked at age and sex, which are key variables when studying smoking. Why did they refrain from doing so? Again, the lack of transparency (Issue #5) about the data leaves all these important questions unanswered.

[Page 6] *…a treatment period of one year might be too short for the effect to fully unfold. Smoking habits are hard to change, especially with "soft" measures such as plain packaging, and it would be surprising to see a strong and immediate effect. Such an effect would only be expected if accessibility were suddenly restricted (e.g. restaurant bans) or if prices suddenly increased dramatically. The argument, I think, is equally true for existing smokers and those taking up smoking. The idea of plain packaging, as far as I can see, is to influence consumption*

review/packaging-review-docs/meetingsandbriefings/Professors-Kaul--Wolf-%28University-of-Zurich%29-20-March-2014.pdf

[7] Philip Morris Limited. Response to the Consultation on "Standardised Packaging" 7 August 2014. Available from: http://www.pmi.com/eng/tobacco_regulation/submissions/Documents/UK%20-%20%20Standardised%20Packaging%20Submission%20PML.pdf

[8] Ibid.

[9] Philip Morris International. Researchers Find No Evidence Plain Packaging 'Experiment' Has Cut Smoking, Press release, 2014.
http://www.pmi.com/eng/media_center/Pages/plain_packaging_experiment.aspx

behavior by changing the "image" of tobacco brands and smoking in general. Such an approach probably has a very slow and subtle effect that might not be observed in just one year. (...) I think it is an issue that might have deserved some discussion in the papers by Kaul and Wolf.

We agree that this is an issue that might have deserved some discussion, which one normally expects to read in a balanced treatment of the subject. This was unfortunately lacking in prof. Kaul and Wolf's papers.

One further consequence of this remark is that low level effects observed during the first year could nevertheless be considered as successful outcomes of plain packaging. It is such low levels which need to be taken into account in the power calculations – see below.

[Page 6] *For example, samples might be non-representative (no specific information on sampling is given by Kaul and Wolf) and non-response or social-desirability bias or other measurement errors might distort the data. Furthermore, the data analyzed by Kaul and Wolf has been aggregated from individual-level measurements and errors might have been introduced during this process (e.g. inadequate treatment of missing values).*

This is a serious issue, which provides yet another illustration of our *Issue #5: Contradiction and lack of transparency about the way the data was obtained.* In their reply to the critique by Laverty et al. in The Lancet, prof. Kaul and Wolf said: "*The data we have worked with are publicly available, and our analyses are described in detail and can be replicated.*" The data are not publicly available, they are proprietary and researchers who want to have access to the underlying data set will have to pay Roy Morgan's a high fee, amounting to tens of thousands of Australian dollars. It is therefore difficult to assess how the data were aggregated.

Comments on *3. A re-analysis of the data*

[Page 8] *...a simpler and more straightforward approach would be to directly estimate the treatment effect by including additional parameters in the model*.

We agree. This is actually what Diethelm and Farley have done in their re-analysis (see attached paper). However, the model should not be limited to taking into account the effect of *only one* tobacco control measure at the end of the 13-year period. If it is assumed that plain packaging may have an effect on prevalence as a tobacco control measure then, for the sake of consistency, it should be equally assumed that the other tobacco control measures implemented during the 13-year period could also have had an effect. A good model should therefore consider at least the prominent tobacco control measures that are susceptible of having an effect on prevalence.

A model that ignores the influence of tobacco control measures may miss their confounding effect on the intervention under consideration. Prof. Kaul and Wolf analysis (and for that matter prof. Jann's re-analysis) is entirely based on the assumption that these other measures had no effect and that smoking prevalence followed a "*pre-existing*" linear decline which was similar in all OECD countries, and independent of tobacco control measures. In our opinion, the failure to take into account these important tobacco control measures reduces the pertinence of prof. Kaul and Wolf's analysis and of prof. Jann's re-analysis. It could very well be that different results would be obtained when these other measures are taken into account in the model and Diethelm and Farley's re-analysis shows that this is indeed the case.

We also note that prof. Jann finds that there is a simpler and more straightforward approach than the approach used by prof. Kaul and Wolf.

[Page 8] *Given that the dependent variable is dichotomous, however, a more appropriate model for the data might be logistic regression (...). Logistic regression, for example, has the advantage that effects level off once getting close to zero or one by construction, so that predictions outside 0 to 1 are not possible.*

We agree. Logistic regression is indeed more appropriate here. This again contradicts the claim made by prof. Wolf when he declared to members of Sir Cyril Chantler's review group: "*I can say from upfront the methodology that we have employed is the one that gives the most leeway to finding an effect, if there had been any.*" [10]

[Page 11] *An issue with the adult data is that a linear model does not fit the pre-treatment period very well. For example, a quadratic model indicates curvature (significant coefficient of month squared in the first table of the output below). Based on graphical inspection of a nonparametric smooth, Kaul and Wolf (2014a) decided to use only observations from July 2004 on to estimate the baseline trend in the pre-treatment period.*

[10] Chantler C. Standardised packaging of tobacco Report of the independent review undertaken by Sir Cyril Chantler. April 2014 Available from http://www.kcl.ac.uk/health/10035-TSO-2901853-Chantler-Review-ACCESSIBLE.PDF

We are glad prof. Jann recognizes that this is an issue, although we would have been less willing to accept prof. Kaul and Wolf's treatment of the issue. If the linear model does not fit the data well, it is the model that needs to be reconsidered and adjusted, not the data by truncating a significant part of them. Two studies by Wakefield et al., published in peer-reviewed journals and which were available at the time Kaul and Wolf posted their two papers on the website of the University of Zürich, explain that smoking prevalence in Australia is not simply the result of a linear time trend, but varies in a way that reflects the adoption of tobacco control measures, notably strict smoke-free policies and tobacco tax increase. Although they were aware of these important studies, Kaul and Wolf fail to refer to them in their papers, and prof. Jann similarly ignores them.

[Page 13] *The most straightforward approach to estimate the treatment effect of plain packaging is to apply the above models to all observations and include an indicator variable for the treatment.*

We agree. This is what has been done by Diethelm and Farley in their re-analysis of the data presented in the attached paper. We indeed think that there is no reason to restrict the indicator variables to only one tobacco control measure. Indicator variables should also be created to take into account other prominent tobacco control measures implemented during the 13-year period that could act as confounders for the plain packaging effect.

[Page 14] *However, if we employ a more gradual interpretation of statistical results without resorting to strict (and somewhat arbitrary) cutoffs, we can acknowledge that the effects at least point in the expected direction. For example, using a one-sided test, the p-value from the logistic regression for minors is p = 0:062, which is not far from the conventional 5% level.*

We note that prof. Jann has found an effect for minors which is statistically significant at the 10% level. We would have expected prof. Kaul and Wolf to arrive at a similar conclusion and acknowledge that their data was suggestive of an effect of plain packaging on minors. Instead, their conclusion leaned in the opposite direction, i.e. in the direction of no evidence of an effect: "*Altogether, we have applied quite liberal inference techniques, that is, our analysis, if anything, is slightly biased in favor of finding a statistically significant (negative) effect of plain packaging on smoking prevalence of Australians aged 14 to 17 years. Nevertheless, no such evidence has been discovered. More conservative statistical inference methods would only reinforce this conclusion.*"[11]

[Page 17] *From July 2004 (highlighted) on there is not much change and all effects are clearly insignificant.*

While this may be true of the model used by prof. Kaul and Wolf, which rests on the questionable assumption that all tobacco control interventions are ineffective (except perhaps the last one, plain packaging) and do not decrease smoking prevalence, our own re-analysis of the data (using

[11] Kaul A and Wolf M. The (Possible) Effect of Plain Packaging on the Smoking Prevalence of Minors in Australia: A Trend Analysis. University of Zurich Department of Economics Working Paper Series. May 2014; Available from: http://www.econ.uzh.ch/static/workingpapers.php?id=828

the logistic regression analysis described by Diethelm and Farley in their paper) shows that the effect associated with plain packaging remains substantial and statistically significant even when the onset of the period of analysis is July 2004.

[Page 17] *To the left of July 2004 the effect systematically grows and eventually becomes significant. However, in this region there is considerable misfit of the linear model (...), which inflates the treatment effect estimate.*

We are pleased to see that prof. Jann agrees that the linear model is not suitable and does not fit the data very well. As said above, the poor fit of the linear model invoked by prof. Kaul and Wolf led the two authors to truncate their data by eliminating the first 42 months rather than to question the validity of their linearity assumption.

[Page 27-28] *The overall tests for adults are similarly weak (with two-sided p-values of 0.922 and 0.555), although there is a significant effect in December 2012 (with p-values of 053 and .008). Yet, as indicated by the overall tests, it is not very surprising to see a significant effect among 13 monthly estimates (for example, using a 5% level, we would expect one significant result per twenty test). It is thus hard to say whether the December 2012 effect is systematic. Such a conclusion could only be drawn if there had been a strong and well justified prior expectation of a specific effect in December 2012 only, but not in other months.*

We agree, and this is an important point. However, prof. Jann draws back from the full conclusion that his observation implies, i.e. that prof. Kaul and Wolf failed to show that there was "*strong and well justified prior expectations of a specific effect in December 2012*" and that consequently their special treatment of December 2012 was *ad-hoc* and unfounded. A recent paper on the introduction of plain packaging in Australia says: "*The new standardised packs were available and likely already exerting an impact in the Australian market from October 2012 onwards, well before the 1 December mandated introduction date.*"[12] According to the same source, in November, more than half of the packs of cigarettes sold were plain. December has not been mentioned by any expert nor by the Australian health authorities as a month during which smoking prevalence would suddenly drop, to return to its previous level in January 2013. Nevertheless, prof. Kaul and Wolf conclude that "*if one is willing to accept a relatively low level of statistical significance (10%), then there is evidence for a very short-lived plain packaging effect on smoking prevalence, namely in December 2012 only (after which smoking prevalence is statistically indistinguishable from its pre-existing trend).*" When one knows the nature of nicotine addiction, this is almost as absurd as saying: "there was a short-lived drop of the number of people with one leg in December 2012". Excluding December from the plain packaging period was clearly a *post hoc* decision, which enabled the two authors to avoid considering the prevalence observed in December 2012 as "evidence for a plain packaging effect" – this is the point we raised under *Issue #1 - Avoiding evidence by post-hoc change to the method.*

[12] Scollo M, Lindorff K, Coomber K, et al. Standardised packaging and new enlarged graphic health warnings for tobacco products in Australia—legislative requirements and implementation of the Tobacco Plain Packaging Act 2011 and the Competition and Consumer (Tobacco) Information Standard, 2011 *Tob Control* 2015;24: ii9–ii16.

[Page 36-37] *I report the results from one-sided tests because we are interested in whether plain packaging decreases smoking prevalence. After all, this is why plain packaging has been introduced. The choice of the test is not a matter of whether an opposite effect is possible or not. It rests on an a-priori decision to look only for evidence for an effect in one direction, which seems appropriate in the present context.*

We agree and this is the point we made under *Error #6 - Invalid significance level due to confusion about one-tail vs. two-tail test.* Prof. Jann does not consider this an error, but labels it an "*a priori decision,*" and we accept his view. For us, what matters is the direction taken by such a decision. It seems that each time prof. Kaul and Wolf were confronted, implicitly or explicitly (and consciously or not) with such types of decisions, they almost consistently leaned on the side where the chance of finding evidence of an effect of plain packaging was reduced.

[Page 37] *I conclude that, for minors, only a very strong treatment effect could be detected with these data with reasonable power.*

We agree. This is the point made in the paper co-authored by OxyRomandie's president with prof. McKee.[13] This has not prevented prof. Kaul and Wolf from accepting to be quoted in the press release by Philip Morris International as saying "*We used statistical methodology that gave every possible leeway for detecting a possible plain packaging effect. Nevertheless, the data does not support any evidence of an actual effect of the Australian Plain Packaging Act on smoking prevalence of minors."*[14]

[Page 38] *I conclude that the chance of detecting an effect for adults based on the given data would have been reasonably high, if the effect was around 1 percentage point or larger.*

Prof. Jann should have added at the end of his sentence: "*… and, in fact, one such effect was found.*" Indeed, in his calculations presented on page 27, he *found* a significant effect with both the WLS and Logit models (on page 27, he says: "*there is a significant effect in December 2012 (with p-values of .053 and .008*"). Correctly, prof. Jann refrained from excluding December 2012 from his analysis and consistently used the 13-month period from December 2012 to December 2014 inclusive for his power calculations. We suggest that his power results presented in Table 2 on page 39 (wrongly entitled "*Summary of power analyses for* minors*"*) are not applicable in the present situation. To be meaningful in the present context, his simulation should have estimated the probability of finding *at least two months* with a significant effect for actual effects of specific sizes; the power values produced by such simulation will be much lower than those shown in Table 2.

---

[13] Diethelm P, McKee M. Tobacco industry-funded research on standardised packaging: there are none so blind as those who will not see! *Tob Control* Published Online First: 7 July 2014 doi:10.1136/tobaccocontrol-2014-051734

[14] Philip Morris International. Researchers Find No Evidence Plain Packaging 'Experiment' Has Cut Smoking, Press release, 2014. http://www.pmi.com/eng/media_center/Pages/plain_packaging_experiment.aspx

Furthermore, it should be observed that a 1% effect (i.e. about a reduction of 6% of the rate of smokers) is an unjustifiably high target. A reduction of 0.5% (or even 0.25%) would already be a considerable success - a 0.5% reduction would correspond to a reduction of 3% of the number of smokers and would result in about 600 lives saved, i.e. close to half of the deaths caused by road accidents in Australia.

For practical purposes, a statistical test should be assessed for its ability to detect an effect which is sufficient to justify the intervention. Even if the introduction of plain packaging had further reduced smoking prevalence in Australia during the first year by 0.25%, this could legitimately be considered an important success, sufficient to justify the measure, as about 300 lives would be saved. If a test has a high probability of not detecting such an effect, it is the limitation of the test that should be emphasized, not the lack of the evidence it is unable to detect.

At the 0.5% level, all the methods examined by prof. Jann (those of prof. Kaul and Wolf and his own methods) are associated with low power. This means that they have a high chance of not detecting any effect when in fact there was one. This is completely at the opposite of what prof. Kaul and Wolf "*confirmed*" to Philip Morris: "*if there had been an effect in reality (...) it would have been reflected in the data*" (again assuming PMI said the truth).

Comments on *4. Remarks on the errors and issues raised by OxyRomandie*

Two of the errors we initially raised have been addressed in prof. Jann's report in a way that we consider satisfactory and we are ready to withdraw them from our list of errors. The other errors, however, either have been confirmed by prof. Jann or we consider that his explanations did not provide a satisfactory rebuttal. In our opinion the essential part of our initial criticism remains unchallenged.

Comments on *4.1 Error #1: Erroneous and misleading reporting of study results*

[Page 38] *OxyRomandie does not claim that Kaul and Wolf misrepresented the results in their working papers.*

This is right. However, like prof. Jann, we think prof. Kaul and Wolf could have be more careful in pointing out the weakness and limitations of their design, methods and results. The way they presented some results paved the way for their misrepresentation by others, and notably by their contractual partner in the project, tobacco multinational Philip Morris.

[Page 38] *I agree that some of the quotes provided by OxyRomandie read as "evidence for no effect" instead of "no evidence for an effect."*

This is a crucial point for us and we are pleased to see that prof. Jann recognizes it. This is also called "absence of evidence" vs. "evidence of absence".

[Page 38] *The difference between the two formulations* ["evidence for no effect" and "no evidence for an effect."] *is subtle and my experience is—based on teaching statistics—that people without statistical training are typically not aware of the difference. Of course, it is better to always use the correct formulation, but I do not think that it really makes a big difference (statistically trained people will appreciate the correct meaning either way, while others will probably not understand the difference).*

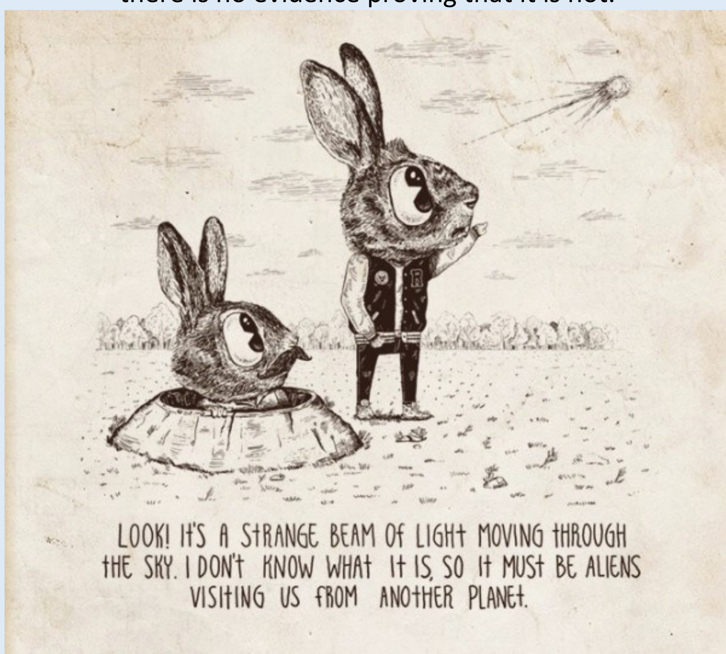
Confusing "evidence of absence" and "absence of evidence" is a well-known logical fallacy called *appeal to ignorance* (or argument *ad ignorantiam*), which is documented in textbooks on informal logic. For instance, in Copi and Cohen's *Introduction to Logic*[15], we read: "*It is fallacious to argue that some proposition is true simply because it has not been proved false. It is equally fallacious to argue that some proposition is false simply because it has not been proved*

[15] Copi IM, Cohen C and McMahon K. Introduction to Logic. 14th Edition. Pearson,Upper Saddle River, New Jersey (USA), 2010

true. (…) In science, the fallacious appeal to ignorance crops up when claims for which there is no evidence are held false for that reason." (Prof. Douglas Walton has even written a book on *appeal to ignorance*.[16]) It is an argument often used by charlatans (see illustration below) with which they take advantage of human credulity – the fact that many people do not fully appreciate the difference makes them vulnerable to it. This is why it is one of the favourite devices in the panoply of the tobacco industry deception techniques. As the tobacco companies' credibility level is very low, they need respected scientists from reputed universities to provide them with the premise of the argument ("it has not been proved true") and they will provide the rest ("therefore it is false"). This is essentially what Philip Morris has done with prof. Kaul and Wolf and the University of Zürich.

Appeal to Ignorance

Such an argument assumes a proposition to be true simply because there is no evidence proving that it is not.



Source: http://www.visualnews.com/2013/09/18/logical-fallacies-explained-fun-animal-illustrations-illustrated-book-bad-arguments/

It is important to note, however, that PMI, in its reporting of the results of prof. Kaul and Wolf's studies to the UK government, went *beyond* the fallacy of appeal to ignorance. In its simple form, this fallacy has the following structure:

[16] D Walton. Argument from ignorance. Pennsylvania State University Press (1 Jan. 1995); see also D Walton. The appeal to Ignorance, or *Argumentum Ad Ignorantiam.* Available from http://www.dougwalton.ca/papers%20in%20pdf/99ignorantiam.pdf

Premise: *There is no evidence for* ***A***

Conclusion: *Therefore* ***A*** *is false*

Here ***A*** is "Plain packaging is effective". Perhaps realizing that this could be too obviously identified as a fallacious argument, in its submission to the UK government, Philip Morris changed it to:

Major premise: *There is no evidence for* ***A***

Minor premise: *If* ***A*** *were true there would be evidence for* ***A***

Conclusion: *Therefore* ***A*** *is false*

("*Kaul and Wolf confirmed that if there had been an effect in reality (...) it would have been reflected in the data. According to the study, however, no effect was found.*") Adding the minor premise transforms the previous fallacious argument into a *logically* valid one, by removing the *false dichotomy* that lies at the core of an argument *ad ignorantiam*. But this is done at the cost of introducing *a false statement* (the minor premise is false – it is equivalent to saying that the power of the tests used by prof. Kaul and Wolf was 1.0 which we know is not true). As the argument is now logically valid, the conclusion is inescapable and can be left implicit (this is what PMI did). In this latter form, the argument will be very powerful before judges and arbitrators. Philip Morris presented Kaul and Wolf's findings as *proof* that plain packaging is ineffective; to do so they manipulated the argument by introducing a false statement, pretending that it had been confirmed by the two professors. *This is a very serious scientific irregularity.*

The target audience of the papers is not just students: Philip Morris International has not commissioned the University of Zürich to produce a paper with the aim of simply adding a contribution to the large body of academic teaching material. We respectfully suggest that the target audience of the two papers may be much larger than prof. Jann perhaps realizes and it is mostly controlled by PMI (according to the contract between the University of Zürich and the tobacco multinational, the latter is "*the sole and exclusive owner in all countries of Work Product from the time of its creation.*") Key members of the target audience are *policy decision makers, legislators, lawyers, arbitrators and judges*. Of course, prof. Kaul and Wolf's working papers will not be read by most of these people but will be (and are) produced before them by PMI as key pieces of evidence supporting the misrepresentation of their results. And unfortunately, as the authors of the two papers were not very careful about expressing the limitations of their analysis, they made it easier for the tobacco sponsor to misrepresent their findings.

Unless one is engaged in the tobacco control field, it can be easy to miss the context in which prof. Kaul and Wolf two papers were published. This context is of a very serious legal battle between the tobacco multinational Philip Morris International and two countries which have either adopted plain packaging (Australia) or have very large graphic health warnings on cigarette packs making the pack look virtually the same as a plain pack (Uruguay). Philip Morris has launched two legal actions, one against Australia using a Bilateral Investment Treaty (BIT) between Hong Kong and Australia and one against Uruguay using a BIT between Switzerland and that country. These complaints are currently being processed by arbitration tribunals. Furthermore, the tobacco industry has lobbied a number of countries which have filed a complaint before the World Trade Organization against Australia for its plain packaging decision.

While these legal actions are rather technical in their arguments, one of their recurring central theme refers to the principle of proportionality. The tobacco companies and their lawyers argue that plain packaging is a disproportionate measure because they claim it is ineffective, drawing extensively on prof. Kaul and Wolf's two working papers as key pieces of evidence of its ineffectiveness.

It is not therefore possible to brush away the difference between "absence of evidence" and "evidence of absence" and fall into the trap of appeal to ignorance. The use of the argument *ad ignorantiam* by the tobacco sponsor as proof of the ineffectiveness of plain packaging is a very serious issue, which is the central point on which OxyRomandie's intervention in the present affair was based. This has very important public health and legal implications.

The studies under consideration here fall outside the usual realm of academic science: they are typical examples of *post-academic research*[17]. The two professors would probably never have studied the effect of plain packaging in Australia if they had not been approached and paid by Philip Morris International to do it. Preferably, the University of Zürich should have refused to engage with such a company, whose interest is in fundamental conflict with public health. Given that it has, it must apply the greatest sensitivity to the implication towards society of such research work – with which the name of the University of Zürich is associated - and assume its responsibility. It must categorically condemn the misuse of the research work produced by prof. Kaul and Wolf and the manipulation by the tobacco multinational of science and the two professors' work.

[Page 39] *I am very skeptical of whether researchers can be held responsible for monitoring the use and interpretation of their results by others. This would be an obligation that is impossible to fulfil and it would strongly discourage researchers from publishing anything. Of course, we can expect researchers to pay attention to a correct representation of their results in press releases or similar materials, if they are given the chance to do so. But we cannot make them responsible for what is published by others and we cannot expect them to actively watch out for material misinterpreting their results.*

If prof. Jann includes the contractual partner under "*others*" in the above statement, then this would be an extremely worrying evolution of academic science: this would institutionalize the complicity of the university in the manipulation of science by corporate sponsors. This is implying that as long as the sponsor pays, he owns the results of the studies produced by the university, which are considered purely as deliverables, and his ownership extends to the point of being able to distort or misrepresent their findings, without the university feeling any obligation or responsibility to intervene to prevent or stop the disinformation. This is prone to saying: "What the corporate sponsor does with my study is none of my business. I have delivered my work product and my role as an academic scientist stops there."

[17] See J Ziman. Is science losing its objectivity ? Nature. Vol. 382 751-754 29 August 1996

Such an approach to partnership between the private sector and the university would wide open the door to all kind of abuses. Imagine a multinational pharmaceutical company selling a drug with a huge market but which is strongly suspected of causing cancer. The easy way for it would be to mandate the University of Zürich to study whether *headache* is among the suspected side effects. Getting the results of the study which naturally would have found no effect (this is why headache was chosen in the first place), the company could then use these results to claim "it's proved - and the University of Zürich confirms it - our drug is perfectly safe!" This sounds a bit ridiculous, but it is not far from what Philip Morris International has done with the studies carried out at their request by prof. Kaul and Wolf. This would indeed be a very serious and very worrying development, which would undermine public confidence in academic research.

In the present case, the University of Zürich has signed a contract with Philip Morris International that gives it a high degree of control over the communication about the studies by the tobacco multinational (see extract below). In particular, PMI cannot use the name of the university nor the name of the professors in press releases and other external communication "*without the prior express written approval*" of the university. Having such a degree of control over the communication by the tobacco sponsor, the university can be considered as a full stakeholder in PMIs' communication about the studies and consequently must assume its part of the responsibility for the misrepresentation of the results of the studies that is found in PMI's communication – misrepresentation that is recognized by prof. Jann.

9. CONDUCT, PUBLICITY AND THIRD PARTY CONTACTS

9.1 Neither Party nor its Personnel shall, without the prior express written approval of the other Party, (i) advertise or otherwise publicize the existence or terms of this Agreement or any other aspect of the relationship between the Parties or (ii) use the others Party's name or that of any of its Personnel name or any trade name, trademark or service mark or brand imagery belonging to that Party and/or its Affiliates in any press release, any form of advertising, or any of its business communications (internal or external) except those necessary to provide the Services.

9.2 If at any time either Party or either Party's Personnel is contacted by a third party, including any news organization, concerning the Services provided under this Agreement, such Party and/or such Party's Personnel shall make no comment, notify the other Party of the third party contact, and refer the third party to such other Party and/or coordinate the information provided to the third party with such other Party.

Comments on *4.2 Error #2: Power is obtained by sacrificing significance*

We take due note that, although our observation was true, it does not invalidate prof. Kaul and Wolf's claim concerning the power of their tests. However, our confusion came the fact that, when interpreting their results, the two professors have changed the emphasis of their test, shifting from testing for the presence of the expected effect ($H_o: \mu \geq 0$ vs. $H_{1:}\mu < 0$) to testing for the absence of the expected effect ($H_o: \mu < 0$ vs. $H_1: \mu \geq 0$), sacrificing usual significance and using the power of the test as an indicator of significance, inversing the roles of Type I and Type II errors. They did all this while still retaining the usual test criteria – thus the confusion. In this new setting, a power of 0.8 corresponds to p-value of 0.2 and is *not* sufficient to achieve significance – this is something that prof. Kaul and Wolf should have pointed out in their papers. This was mostly the point we made, which therefore still remains valid. We consider this an important defect of the two papers.

Comments on *4.3 Error #3: Inadequate model for calculating power which introduces a bias towards exceedingly large power values*

We are pleased that prof. Jann agrees with the point we made and accept that prof. Kaul and Wolf can choose other assumptions. It is nevertheless important to acknowledge that the sudden effect model inflates the estimate one obtains when calculating the power of the test and therefore introduces a bias towards large power values (yet another bias which goes in the same direction as the others). This appears clearly in prof. Jann's re-analysis – notably in his Tables 1 and 2. We therefore still consider this a flaw of the papers.

Comments on *4.4 Error #4: Ignorance of the fact that disjunctive grouping of two tests results in a significance level higher than the significance level of the individual tests*

We are pleased that prof. Jann acknowledges that the point we made was true and accept that this does not affect prof. Kaul and Wolf's power calculations. We withdraw this point.

Comments on *4.5 Error #5: Failure to take into account the difference between pointwise and uniform confidence intervals*

We are pleased that prof. Jann considers our argument "undoubtedly true". He again observes that this does not render the power values reported by Kaul and Wolf invalid, and we accept his explanation. We withdraw this point.

Comments on *4.6 Error #6: Invalid significance level due to confusion about one-tail vs. two- tail test*

Prof. Jann's explanations confirm and reinforce the criticism we made under this point.

[Page 41] *Kaul and Wolf (2014a) report tests for monthly deviations based on 90% confidence intervals and based on 95% confidence intervals. OxyRomandie argues that watching for a 90% confidence interval to be entirely negative is equivalent to a one-sided test at the nominal 5% significance level. I agree.*

We are pleased that prof. Jann agrees with OxyRomandie.

[Page 41] *…the dispute is over whether a two-sided test or a one-sided test is appropriate. Like OxyRomandie I consider a one-sided test appropriate in the present context (see above), but it is essential to realize that there is no right or wrong here.*

We are again pleased that prof. Jann agrees with OxyRomandie. We should like to add, however, that the question of "*right or wrong*" might be considered in the totality of the issues reviewed here. It is not a trivial question, particularly when such choices - which are indeed not wrong or right – appear to always be made *in the same direction*. Prof. Kaul and Wolf's choice here is the one that will make it more difficult to declare an effect statistically significant, putting a stone in the "no effect found" pan of the balance. It is the accumulation of the stones in the same pan of the balance which for us constitutes a major flaw of Kaul and Wolf's papers.

[Page 41] *However, note that the analyses performed by Kaul and Wolf are (mostly) in line with a one-sided setting, not a two-sided setting. In particular, in their power simulations only significantly negative deviations are counted. (…) This may all be somewhat confusing, but the bottom line is that indeed there is an inconsistency in Kaul and Wolf (2014a): The power analyses they perform use one-sided tests, but their interpretation of the December 2012 effect is in terms of a two-sided test.*

This was our point, better expressed than we did. What prof. Jann says is that prof. Kaul and Wolf used a two tailed test for interpreting their results, which allowed them to report low statistical significance (and reporting absence of evidence at the standard 5% level of significance), while using the one-tailed test for their power calculations, which inflated the power of their test. Again two stones put in the "no effect found" pan of the balance. The *inconsistency* reported by prof. Jann is also clearly an error.

One could have expected prof. Kaul and Wolf to be very careful and present their results with some caveats. Instead, the media release signed by the two professors and issued by *IPE-Institute for Policy Evaluation*, a private German consulting firm, announced the results of prof. Kaul and Wolf's second paper as follows: "*The experts found no evidence for a plain packaging effect on smoking prevalence using standard techniques for statistical analysis, in particular requiring a statistical significance level of 5%, which is the standard in applied research.*"[18] This was

[18] IPE Institute for Policy Evaluation: Research Released on Smoking Prevalence in Australia Following Plain Packaging, Media release, 1 July 2014. Available from http://www.ipe-

accompanied on the same by a release from the Reuters agency which reproduced IPE's media release, given it a worldwide audience[19]. Both releases quoted "*the lead author of the report,*" prof. Kaul: "*Using standard analytic techniques that are easy for other researchers to replicate, we found no solid evidence for a plain packaging effect in any month. Only when using statistical techniques biased in favour of finding a plain packaging effect could we detect weak evidence for a one-time effect on smoking prevalence in December 2012 itself, after which smoking prevalence is statistically indistinguishable from the pre-existing trend.*"

[Page 42] *However, I would not consider this a "fundamental flaw".*

This error may not be "*fundamental*" by itself, however, what its particular relevance comes from the importance that the two professors gave to this point in their public communication and the fact that it leans on the same side as the other errors, supporting the authors' conclusion of absence of evidence.

Comments on *4.7 Error #7: Invalid assumption of long term linearity*

[Page 42] *From my own analysis I cannot support the claim that a linear fit is inappropriate. There is some extra variation in the monthly estimates in addition to what we would expect from a binomial distribution, but the deviations appear unsystematic (i.e. not suggesting a specific alternative longtime trend) and it would be difficult to come up with a convincing alternative trend model.*

This question is discussed in Diethelm and Farley's re-analysis (see attached paper). Their findings confirm what was said by Sir Cyril Chantler. It is important when analysing the Australian data to take into account "*known confounding factors that influence smoking and prevalence*", including tax rises.

Contrary to what prof. Jann fears, it is not a great challenge to come up with a model which is convincing and fits the data much better than the rudimentary linear model. It is just a matter of *not excluding* that tobacco control measures adopted by Australia during the 13-year period of observation have had an effect on smoking prevalence, a fact which is actually well documented in two studies by Wakefield and her colleagues.[20] In the most recent of these two studies, the

---

saarland.de/app/download/8738690094/Media+Release+-+University+of+Zurich+and+Saarland+Report+-+July.pdf?t=1424102445

[19] IPE Institute for Policy Evaluation: Research Released on Smoking Prevalence in Australia Following Plain Packaging. Reuters. 1 July 2014. Available from http://uk.reuters.com/article/2014/07/01/ipe-idUKnBw015108a+100+BSW20140701

[20] Wakefield MA, Durkin S, Spittal MJ, Siahpush M, Scollo M, Simpson JA, et al. Impact of tobacco control policies and mass media campaigns on monthly adult smoking prevalence. *Am J Public Health*. 2008;98:1443-50
and

authors found that "*stronger smoke-free laws, tobacco price increases and greater exposure to mass media campaigns independently explained 76% of the decrease in smoking prevalence from February 2002 to June 2011*." This contradict the assumption made by prof. Kaul and Wolf of a "*pre-existing*" linear decline which is independent of tobacco control measures. As noted above, Prof. Kaul and Wolf were aware of the existence of prof. Wakefield's work.

As shown by Diethelm and Farley in their re-analysis (see attached paper), the assumption of linearity throughout the 12-year "*pre-treatment*" period is a fundamental error which invalidates prof. Kaul and Wolf's findings. It leads to a simplistic and flawed model, which ignores the well documented effect of tobacco control measures on smoking prevalence in Australia and fails to take into account important confounding factors that could mask the effect of plain packaging.

Rather than questioning the validity of the model and revising it, prof. Kaul and Wolf have questioned the data, arbitrarily amputating them by cutting off the lower third, to force them to fit their model. We think that this is wrong and a major flaw of their approach. We respectfully observe that prof. Jann may have fallen into the same trap, although his one-step approach based on logistic regression with a treatment indicator variable was a move in the right direction.

Comments on *4.8 Issue #1: Avoiding evidence by post-hoc change to the method*

[Page 42] *OxyRomandie wonders why December 2012 was excluded from the set of treatment-period months in the power simulations in Kaul and Wolf (2014a), while it was included in the power simulations in Kaul and Wolf (2014b). I agree that it is odd to exclude December 2012, as plain packaging came into effect in December 2012.*

We are pleased that prof. Jann agrees with us that it is "*odd to exclude December 2012*" especially considering that plain packaging was introduced in October 2012 and that in November, more than half of the packs sold by retailers were plain[21]. Quitline statistics indicate that the number of calls to the Quitline rapidly increased after 1 October to peak in November.[22] So if there was a "shock month", this was November 2012, not December.

We note that, for the data on adults, prof. Kaul and Wolf excluded December 2012 from the plain packaging period, reducing it down to 12 months, while other authors consider October as the start of the plain packaging period[23]. Not content with having amputated data on the pre-

---

Wakefield MA, Coomber K, Durkin SJ, et al. Time series analysis of the impact of tobacco control policies on smoking prevalence among Australian adults, 2001–2011. *Bull World Health Organ* 2014;92:413–422

[21] Scollo M, Lindorff K, Coomber K, et al. Standardised packaging and new enlarged graphic health warnings for tobacco products in Australia—legislative requirements and implementation of the Tobacco Plain Packaging Act 2011 and the Competition and Consumer (Tobacco) Information Standard, 2011 *Tob Control* 2015;24: ii9–ii16.

[22]

[23] Ibid.

treatment period by close to one third, they have also amputated 20% of the treatment data by excluding three critical months. All of this was done without truly convincing explanations. We consider this a major flaw.

[Page 42] *Excluding December 2012 from the power simulations, however, is not critical (as can be seen in Table 2 above, comparing Kaul and Wolf's results with my results that include December 2012).*

With respect, we do not agree with prof. Jann. We would note that, while excluding December or including December does not change much the power values, it substantially changes their *interpretation*. If December is included in the power simulations, the power values reported in the columns under the heading "*Any month effect*" in prof. Jann's Table 2 become pointless, as in fact we then *do have an effect*. The power figures in the table give the probability of observing an effect (in any month) given that there is actually an effect of the magnitudes listed in the leftmost column. As we have such an effect, what can we do with such power figures? Basically nothing, except perhaps for the first row, which gives the power figures for the null effect, indicating the associated level of Type I error, which is, unsurprisingly, very high. So the result is unusable one way or the other. The change is indeed critical, and the post-hoc extraction of December had the effect of avoiding this embarrassing situation, which would have made the paper pointless, leaving not much choice to the authors but to admit that, with the data they had and the method they used, they could not come to any conclusion, one way or the other.

Comments on *4.9 Issue #2: Unnecessary technicality of the method, hiding the methodological flaws of the papers*

[Page 42] *I do not consider the approach proposed by Kaul and Wolf (2014a) particularly complicated.*

We have witnessed intellectually sophisticated people having problem making sense of prof. Kaul and Wolf "*algorithmic approach*." Perhaps prof. Jann will agree with us that those who have difficulty understanding the difference between "absence of evidence" and "evidence of absence" might also have some difficulty with the intricacies of prof. Kaul and Wolf's idiosyncratic approach.

Comments on *4.10 Issue #3: Very ineffective and crude analytic method*

[Page 43] *Although I am puzzled about what OxyRomandie tries to illustrate here, I do agree that better test approaches exist than the one used by Kaul and Wolf. However, it is not an issue of lack of power, it is an issue of inflated type I error.*

We are glad prof. Jann states that better test approaches exist than the one used by prof. Kaul and Wolf. If our explanation was not very clear, this was partly due to the difficulty of understanding the use of tests with inflated Type I error, an issue which the explanations provided by prof. Kaul and Wolf and by prof. Jann have contributed to clarify. Our *t*-test is the same as the *t*-test used by prof. Kaul and Wolf, which is presented in column label *IM-1* in Table 2 of their second paper. The power values we present are smaller simply because we have calculated them using the gradual effect model in our simulation instead of the sudden effect model. When using the sudden effect model, we get the same figures as those in Table 2.

Further evidence of the ineffectiveness of their method is provided by Diethelm and Farley in their re-analysis (see attached paper), which arrives at results contradicting those of Kaul and Wolf, while using a standard, state-of-the-art, statistical approach (logistic regression).

Comments on *4.11 Issue #4: Non standard, ad-hoc method*

We respect prof. Jann's explanations, observing however that when we see his *Stata* code, we still get the impression that the calculations needed to repeat prof. Kaul and Wolf's analyses are not entirely performed using out of the box solutions and require some programming. Prof. Jann himself agrees that *a simpler and more straightforward approach would be to directly estimate the treatment effect by including additional parameters in the model.* This is a more standard way, which was adopted by Wakefield and her colleagues in the two papers mentioned above, and by Diethelm and Farley in their re-analysis (see attached paper).

We therefore maintain that Kaul and Wolf's approach is idiosyncratic and ad-hoc. Diethelm and Farley's re-analysis shows that, had they used a more standard, state-of-the-art approach (as D&F did), they would have obtained very different results from those contained in their papers, results which would probably not have pleased their financial sponsor.

Comments on *4.12 Issue #5: Contradiction and lack of transparency about the way data was obtained*

Prof. Kaul and Wolf's communication about the way they obtained their data is confusing. They have declared in the Lancet (referring to their data on adolescents): "*The data we have worked with are publicly available, and our analyses are described in detail and can be replicated.*" As we wanted to replicate the analysis, OxyRomandie has tried to obtain such "*public*" data. So far without success. We asked an Australian colleague who has published studies on tobacco using data from Roy Morgan to guide us. Her answer was "*The smoking prevalence data are not publicly available – the dataset is owned by the Roy Morgan Research (RMR) company.*" Then we asked her whether she could provide an order of magnitude on the price we would have to pay to obtain the data. The answer was that "*it would be somewhere from tens of thousands to hundreds of thousands of Australian dollars*".

As prof. Jann has pointed out, it is difficult to know how the data was aggregated and how were possible missing values treated. There is a kind of minimalism in which the data is described in prof. Kaul and Wolf papers that leaves the reader desiring to know more. This is our point and it still remains valid.

Comments on *4.13 Issue #6: Conflict of interest not fully declared*

[Page 44] *In general, the problem with industry sponsored research might not be so much that single studies are biased or flawed. A much bigger problem, in my opinion, is that industry funding biases the selection of studies that are conducted and that unfavorable results are often withheld, leading to publication bias. In the present case it does not seem that PMI could have withheld publication if results would have been unfavorable, but we do not really know.*

We agree. The funding bias has many manifestations, and publication bias is indeed an important one. Human bias is also important and cannot be excluded, even with people having the best intentions and whose good faith cannot be questioned. As has been demonstrated by social psychologists (see for example the excellent book *Influence*[24] by Robert Cialdini), the mechanisms of influence are multiple, complex, subtle and most often not perceptible by those who are being manipulated. The tobacco industry has developed highly sophisticated manipulation techniques over the last 50 years and it is hard for anyone dealing with it to escape from their influence – which most often is not conscious. The fact is that this industry has a record of having funded studies whose results are nearly 100% of the time favourable to their commercial interest (we are still looking for an exception). A landmark article[25] entitled "*Why review articles on the health effects of passive smoking reach different conclusions*" done in the late 90s concluded that "*the only factor associated with concluding that passive smoking is not harmful was whether an author was affiliated with the tobacco industry.*" This may well apply today to studies on the effect of plain packaging.

The only safe way to be protected against the influence of the tobacco industry is to refuse to collaborate with it and refuse any funding from it.

What we do not understand, and what also leads us to think that the declaration of conflict of interest is not complete, is the role played by a third party, German-based IPE Institute for Policy Evaluation. This consulting firm is not mentioned in the contract between the University of Zürich and PMI but, curiously, it is this company which seems in charge of the communication on Kaul and Wolf's studies – not the University of Zürich. On the day the professors' second paper appeared on the website of the university, the company posted on its site a media release signed by them, coordinating this communication with the Reuters news agency. This is intriguing. We re-iterate our point of view here: if this organization is paid to do this work (by whom?) and if prof. Kaul and/or Wolf have a commercial interest in the company, we think that for the sake of transparency, they should have declared it in both papers, as this constitutes a potential conflict of interest.

---

[24] Cialdini RB. Influence: The Psychlogy of Persuasion. Harper Business; Revised edition, 26 December 2006

[25] Barnes EB and Bero LA. Why review articles on the health effects of passive smoking reach different conclusions. *JAMA*, May 20, 1998—Vol 279, No. 19

Comments on *4.14 Issue #7: Lack of peer review*

[Page 45] *OxyRomandie criticize that the working papers were published without having gone through peer review. All I can say is that it is standard practice in economics to make results available as a working paper before submission to a peer reviewed journal. Whether Kaul and Wolf assisted in an orchestrated campaign of tobacco industry to promote the papers I cannot judge.*

We have no problem with the practice of making results available as a working paper before submission to a peer reviewed journal: We note however that, to this day, such submission has not taken place, contrary to what Kaul and Wolf said when they were interviewed by Sir Chantler's team on 26 March 2014.[26]

What OxyRomandie criticises is – and the expression was well chosen by prof. Jann – the "o*rchestrated campaign of the tobacco industry to promote the papers*". We do not know and therefore (like prof. Jann) cannot judge whether and to what extent prof. Kaul and Wolf directly assisted in such campaigns, while they certainly contributed to them via the media releases issued by marketing consulting *IPE Institute for Policy Evaluation*, in which they both appear to have a vested interest.

These media campaigns were orchestrated in close coordination with the publication of the papers on the website of the University of Zürich, leaving no time for the scientific debate to take place, but placing immediately the two papers as key pieces of evidence of the ineffectiveness of plain packaging in the political debate and in litigation.

We therefore consider that the issue has not been satisfactorily addressed by prof. Jann.

Comments on *5 Conclusions*

We accept prof. Jann's recommendations. As long as the publication of prof. Kaul and Wolf's papers on the website of the University of Zürich are accompanied by a statement that draws the readers' attention to their defective nature and to the misleading character of their conclusions, we do not ask that they be *removed* from the website. On the contrary, we think it important that future researchers, who may want to investigate this matter, have access to the papers, even if simply as material of historical interest on how the tobacco industry was able to corrupt science with the help of a prestigious academic institution.

---

[26] *"We will be submitting our study to a peer-reviewed outlet in due time. Given the straightforward nature of the data and the statistical methodology, we do not expect changes to the basic findings during the reviewing process."* in Meeting to discuss "The (Possible) Effect of Plain Packaging on the Smoking Prevalence of Minors in Australia: A Trend Analysis" working paper. Attendees: Kaul A, Wolf M, Cox C, Collis J and Edwards L. King College London, 20 March 2014. Available from: https://www.kcl.ac.uk/health/Packaging-review/packaging-review-docs/meetingsandbriefings/Professors-Kaul--Wolf-%28University-of-Zurich%29-20-March-2014.pdf

It is of crucial importance that the University of Zürich, together with prof. Kaul and Wolf, take their distance with respect to the use made by the tobacco sponsor of the two studies and publicly denounce without ambiguity the further misrepresentation and exploitation of their already defective results, notably in the tobacco multinational's submission to the UK government in response to the 2014 consultation on plain packaging.

The contract that links the University to Philip Morris International gives the University the right to do so. The University's may also make use of its moral rights over the works delivered to Philip Morris, to protect them from being distorted by the tobacco company.

2015.12.14/pad