« Le droit de respirer un air de qualité non pollué par les émanations toxiques du tabac est aussi légitime que celui de disposer d'une eau potable. »

Prof. Dr. Michael Hengartner
Rector
University of Zürich
Künstlergasse 15
8001 Zürich

Geneva, 19 February 2015

Dear Prof. Dr Hengartner,

On behalf of OxyRomandie, I am writing to you to submit the attached comments. I should also like to offer a further explanation on a fundamental point and express one concern we have about the assessment process.

The reply by prof. Kaul and Wolf, which you forwarded to us in your email of 14 February, calls for clarifications and corrections. OxyRomandie's comments are attached to the present (see document entitled "Comments on Kaul and Wolf's reply to our Annex.").

Additionally, I should like to make the following statement, to allow you to fully understand our position. Apart from the "errors" and "issues" listed in our initial Annex, we consider the studies produced by prof. Kaul and Wolf biased on one further account, which is fundamental: they are sponsored by Philip Morris, a tobacco company. As you know, there is a growing body of literature documenting the evidence that studies funded by the drug industry are up to four times more likely to produce positive results than those with other sources of funding. This is called the "funding effect". In case of a company like Philip Morris, the funding effect takes an *extreme* form. Internal tobacco industry documents reveal that one the key criteria used by Philip Morris to fund external scientific research proposals is that they be "*consistent with Philip Morris business objectives*." Not surprisingly, research studies sponsored by the tobacco company *always* produce positive results for the sponsor.

I have personally reviewed tens of thousands of tobacco industry documents and have looked at hundreds of reports of external research projects funded by Philip Morris. I have not seen a single instance of them that produced a result made public which was *not* consistent with the company business interests. This feat is achieved using an array of manipulative and fraudulent techniques, which are well described by professor of history Robert Proctor at Stanford University in his recent book *Golden Holocaust*. The extent of the company's fraudulent behaviour, executed with the complicity of other tobacco companies, has been qualified as "racketeering" by a US federal court, a judgment upheld by the US Supreme Court in 2010.

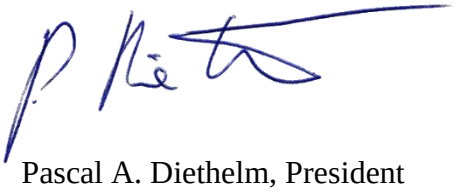
One of the methods used by the company is the careful selection of research topics so as to avoid results which could be detrimental to its business interests and to approach "sensitive" topics from an angle which insures that the result will not lead to a bad surprise. As D. Michael says in his article in the Washington Post, "*It's not the answers that are biased, it's the*

questions." [1] Generally, with Philip Morris, it's both. Research sponsored by the tobacco industry is inherently biased.

I should like also to express our concern about what you said in the interview you gave on RTS (19:30, 16 February). You stated that by requesting the opinion of an external expert, "everybody will see and will be convinced that there is nothing." ("En ayant une opinion claire, nous sommes convaincus que tout le monde va voir et va pouvoir être convaincu qu'il n'y a rien. ») This is a public declaration of what you are expecting from the external assessment: that OxyRomandie's critique be found unjustified. Our first reaction when we learnt that you had appointed an external expert was to applaud. We now fear that our satisfaction with this decision of yours might have been a bit premature. Indeed your public statement of what result you are expecting from the expert assessment – which the University of Zürich is commissioning – may now be considered as being part of the terms of reference of the "external" expertise.

We also insist that our request to you is not just based on narrow technical aspects of Kaul and Wolf's working papers. To be complete, the assessment needs to cover their ethical and deontological implications, which go beyond the contents of the two papers and notably extends to the way they were used, with the authors' explicit or tacit approval, by the tobacco industry to undermine an important public health policy decision by the UK government.

Sincerely yours,



Pascal A. Diethelm, President

Copy: Members of OxyRomandie (via OxyRomandie's website)

[1] David Michaels. "It's not the answers that are biased, it's the questions" The Washington Post, 15 July 2008. Available from http://www.washingtonpost.com/wp-dyn/content/article/2008/07/14/AR2008071402145.html

Comments on Kaul and Wolf's reply to our Annex

OxyRomandie, 19 February 2015

We address here some questions raised by prof. Kaul and Wolf's document entitled "Reply to the ANNEX "Errors and issues with Kaul and Wolf's two working papers on tobacco plain packaging in Australia". Having read this document, we maintain our critique in its entirety, leaving it to the expert designated by the University of Zürich to make his/her own assessment, reserving the right to seek a counter-expertise. However, some of the points raised by prof. Wolf and Kaul in their reply need further clarification and explanation, which we will treat below in the order posed by Wolf and Kaul

Re *1 Lack of authorship*

No comment.

Re *2.1 So-Called Error # 1: Erroneous and misleading reporting of study results*

We maintain our point. Prof. Wolf made the following statement 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."*** This is a highly misleading way of reporting the result of Kaul and Wolf's first study, before a group that advises the government of the United Kingdom on an important public health issue.

Prof. Wolf and Kaul claim that Philip Morris's response to the UK consultation "*was clear in saying that the study found no evidence for an effect.*" They however fail to recognize that the response went much beyond this statement. This is the full paragraph of Philip Morris's response:

> *In both studies, using standard techniques for statistical analysis and applying the standard statistical significance level of 5%, the experts found no evidence that "standardised packaging" had had an effect on smoking prevalence among Australians aged 14 to 17 years old (in the case of the March study) or Australians aged 14 and above (in the case of the June study).* ***Kaul and Wolf confirmed that if there had been an effect in reality (including of the magnitude predicted by Pechey and the DH), it would have been reflected in the data. According to the study, however, no effect was found.*** (emphasis ours)

The contrapositive of the statement in bold (which is logically equivalent to it) is the statement: "**No effect was reflected in the data, therefore there was no effect in reality**." Philip Morris indicates that Kaul and Wolf confirmed such a statement. In their response to our Annex, they say that PMI presented a fair characterization of their results. We conclude that they gave their consent to an **erroneous and highly misleading** reporting of their findings by their partner in a paper which is aimed at influencing an important public health decision by the UK government.

This is for us a crucial point of the case under consideration here. This type of misreporting of study results so as to mislead public opinion and decision makers reminds us of an emblematic precedent. In the June 1961 issue of the American Journal of Obstetrics and Gynaecology, a paper was published by a scientist who reported that he had found no deleterious effect of a drug known as Contergan when taken during the *last three months* of pregnancy[1]. The manufacturer of the drug used the result of this study, which it had commissioned, to publicly claim that its product was safe when taken *during* pregnancy. At a time when the world was becoming gradually worried that Contergan, also known as thalidomide, had severe teratogenic effects on the foetus, effect which happened during the *early months* of pregnancy. This shows how otherwise correct results can be used by a company to mislead the public. This is what PMI has done with Kaul and Wolf's findings, with their tacit agreement (and their findings were not even correct!).

Re. *2.2 So-Called Error #2: Power is obtained by sacrificing significance*

We maintain our point.

2.3 *So-Called Error #3: Inadequate model for calculating power which introduces a bias towards exceedingly large power values*

We repeat that the sudden-change model used by the two professors is not just a model among possible different ones. It is wrong and considerably exaggerates the true power of Kaul and Wolf's tests. The choice of such a model reveals some lack of knowledge regarding the nature of prevalence of tobacco use, which is affected by two components not prone to abrupt changes: uptake and quitting. The quotation of prof. David Hill is fallacious: Prof. Hill never said that "plain packaging will slash smoking rates" *in December 2012*!

[1] R. O. Nulsen, Trial of thalidomide in insomnia associated with the third trimester, *American Journal of Obstetrics and Gynecology* 81 (June 1961): 1245 8

Actually, the Australian legislation is far from intending to produce a sudden decline in smoking prevalence; it rather sees such decline as being a long-term consequence of the tobacco control measure. Indeed the rationale for plain packaging stated in the Tobacco Plain Packaging Bill 2011 is the following:

- *reduce the attractiveness and appeal of tobacco products to consumers, particularly young people;*
- *increase the noticeability and effectiveness of mandated health warnings;*
- *reduce the ability of the tobacco product and its packaging to mislead consumers about the harms of smoking; and*
- *through the achievement of these aims **in the long term**, as part of a comprehensive suite of tobacco control measures, contribute to efforts to reduce smoking rates.* (emphasis ours)

The gradual model (with a very small **Δ** value) is thus a faithful representation of such expectations, not the sudden-change model.

The choice of model plays a crucial role in the calculation of power values. The power to detect a decrease of magnitude **Δ** = 0.25 goes from 0.67 with Kaul and Wolf's sudden-change (inadequate) model to 0.58 with the gradual (realistic) model, and for **Δ** = 0.5, from 0.85 to 0.71. This means that the probability of Type II errors in those two cases, best estimated with simulation using the realistic gradual model, are respectively 0.42 and 0.29, i.e. quite large. On the other hand, the two decreases corresponding to **Δ** = 0.25 and **Δ** = 0.5 are very substantial and probably exceeds what the Australian authorities would expect. Indeed, the first decrease correspond to a relative reduction of the smoking prevalence by 4.2%, while the second correspond to a relative decrease by 5.5%, instead of the 2.9% "pre-existing" decrease. The tests used by Kaul and Wolf have substantial probabilities of not detecting such significant changes. Therefore some words of caution about this limitation of their findings should have been expressed in their papers. Instead, the two authors leaned on the other side, saying that if there had been an effect, they would have found it.

Re 2.4 *So-Called 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 maintain our point. We acknowledge our error and thank prof. Wolf and Kaul for pointing it out: our *corrigendum* was indeed not justified.

Re *2.5 So-Called Error #5: Failure to take into account the difference between pointwise and uniform confidence intervals*

We maintain our point. Kaul and Wolf oscillate opportunistically between pointwise and uniform approaches in a way that totally confuses the reader.

Re *2.6 So-Called Error #6: Invalid significance level due to confusion about one-tail vs. two- tail test*

We maintain our point. We concede that the theoretical explanations provided by Kaul and Wolf, with Type III errors, are technically correct. However, this supposes that the two sides of hypothesis H1 : $\mu \neq 0$ ($\mu < 0$ and $\mu > 0$) are a priori reasonably likely. They are not: plain packaging has been intensively studied by the Australian government and it is with a great degree of confidence in its future effectiveness that the measure has been adopted. Furthermore, tobacco companies would not fight the measure so fiercely if they did not fear this would be detrimental to the sales of their product. While not completely excluded, the possibility that plain packaging would lead to an increase of prevalence, for instance through the rebellious effect among youth, is *extremely* unlikely. Sir Cyril Chandler arrived at the following conclusion in his report to the UK government:

> *Having reviewed the evidence it is in my view highly likely that standardised packaging would serve to reduce the rate of children taking up smoking and implausible that it would increase the consumption of tobacco. I am persuaded that branded packaging plays an important role in encouraging young people to smoke and in consolidating the habit irrespective of the intentions of the industry. Although I have not seen evidence that allows me to quantify the size of the likely impact of standardised packaging, I am satisfied that the body of evidence shows that standardised packaging, in conjunction with the current tobacco control regime, is very likely to lead to a modest but important reduction over time on the uptake and prevalence of smoking and thus have a positive impact on public health*

In such a situation, the two-tailed approach used by Kaul and Wolf, giving equal consideration to the very likely and the implausible greatly increases the probability of Type II errors (false negatives) for an extremely marginal reduction of Type I errors. In other words, the method is biased in favour of a “no effect” result, i.e. in favour of the result expected by their financial partner. The correct hypothesis in our view is: H0 : $\mu \geq 0$ vs. H1 : $\mu < 0$.

We are here in a similar situation faced by the US Environment Protection Agency in the early 1990s when they used a one-tail test to establish a link between exposure to secondhand smoke and lung cancer: the tobacco industry insisted that a two-tailed test was the correct approach, as, they claimed, it could not be excluded that being exposed to secondhand smoke would have a *protective* effect against lung cancer. We know the rest of the story: this tactic was only meant to delay the inevitable - secondhand smoke was officially recognized as a human carcinogen by the International Agency for Research on Cancer in.2004.

Re 2.7 So-Called Error #7: Invalid assumption of long term linearity

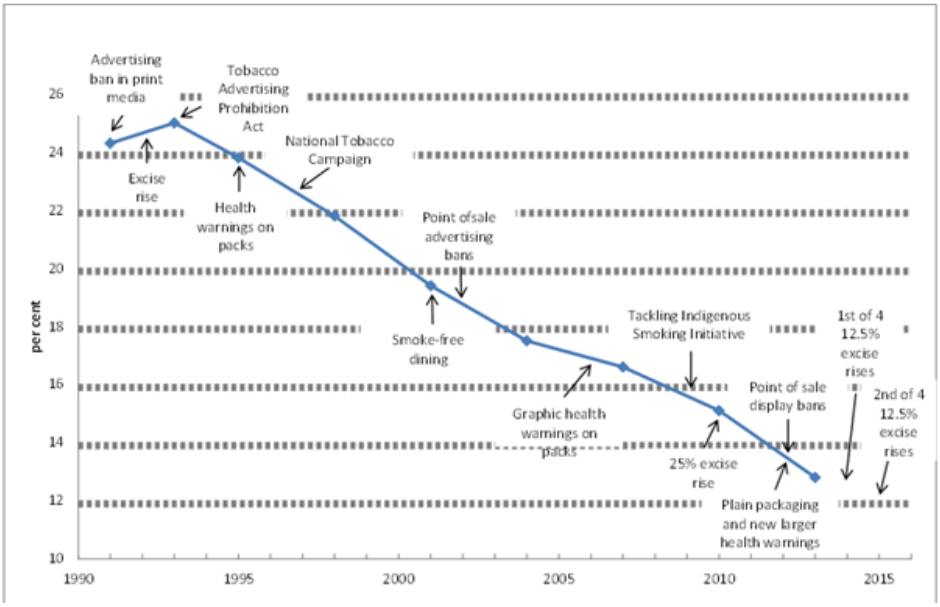
We maintain our point: the assumption of a long-term "pre-existing" linear decline of prevalence is unwarranted.

Prof. Kaul and Wolf's approach is fundamentally circular: they base their inference method on the assumption that tobacco control measures produce no effect whatsoever – the decrease of smoking prevalence observed in Australia over the last 15 years is the result of a "*pre-existing*" trend on which the country's tobacco control measures – regarded by the whole world as exemplary – have no impact, to arrive at the conclusion that one of Australia's most recent tobacco control measure has no effect. Nothing guarantees, and no explanation is provided, that the evolution of smoking prevalence in Australia will inevitably follow a "*pre-existing*" declining straight line.

While one does indeed observe in many developed countries a long term downward trend of smoking prevalence, such trend is not generally linear. At any given moment in time, the state of smoking prevalence can be seen as the result of a balance between two opposing forces: on one side, the tobacco industry constantly uses all its marketing strength to stimulate consumption, create new smokers and dissuade current smokers from quitting while on the other side tobacco control policy pushes to achieve lower levels of prevalence and consumption, inflecting up or down any underlying effect (notably cohort effect). When tobacco control relaxes its efforts, the balance shifts and we observe shortly afterwards that smoking prevalence stops declining and may even start increasing again, notably among the youth. When plotting prevalence over the long term (or total consumption when prevalence figures are not available) one often see an irregular curve reflecting the combined effects of these factors. For instance, the Centers for Disease Control summarizes the evidence in the United States as follows: "States that have made larger investments in comprehensive tobacco control programs have seen larger declines in cigarettes sales than the United States as a

whole, and the prevalence of smoking among adults and youth has declined faster as spending for tobacco control programs has increased."[2]

As concerns Australia, the decline in smoking prevalence observed since mid-1990s is the result of a string of tobacco control measures taken at fairly regular intervals, as shown in the graph below:



Smoking prevalence rates for 14 years or older and key tobacco control measures implemented in Australia since 1990 Source: website of the Australian government

A "*pre-existing"* linear trend that extends over a very long period does not make sense given the nature of the data under consideration: percent prevalence that varies between 0 and 1: at some stage in the future, smoking prevalence in Australia would become negative.

Furthermore, the simple continuation of the decline of prevalence at the same rate is in itself a tobacco control success, as is observed in [ref. 20 in Annex],

> *Such a continued decline is likely a positive sign given that, as prevalence reaches very low levels, the proportion of more addicted smokers who find it more difficult to quit increases. This therefore means that a continuation of the same rate of decline in this context, rather than a levelling off, in fact likely represents an increase in effectiveness of tobacco control measures.*

[2] Centers for Disease Control and Prevention. Best Practices for Comprehensive Tobacco Control Programs—2014. Atlanta (GA): U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease and Health Promotion, Office on Smoking and Health, 2014.

Re *3.1 Issue #1: Avoiding evidence by post-hoc change to the method*

We maintain our point. By treating December as a special month in their second paper and not in their first, Kaul and Wolf altered their methodology in a *post-hoc* fashion. They said they wanted to a take into account a "*shock effect*" that took place in December 2012. But there was no such effect in December 2012. Plain packaging came into force between 1st October and 1st December 2012, with more than 50% of packs used by consumers known to be "plain" by 1st November 2012[3]. The "shock" was spread over three months and thus much reduced, with October being the month with the greatest "shock."

Re *3.2 Issue #2: Unnecessary technicality of the method, hiding the methodological flaws of the papers*

We maintain our point.

Re *3.3 Issue #3: Very ineffective and crude analytic method*

We maintain our point. We believe that we use the same *t*-test as the authors. We simply applied it in the simulation using the gradual model, rather than Kaul and Wolf's sudden change model. When applying the *t*-test with the latter model, we obtain the following results.

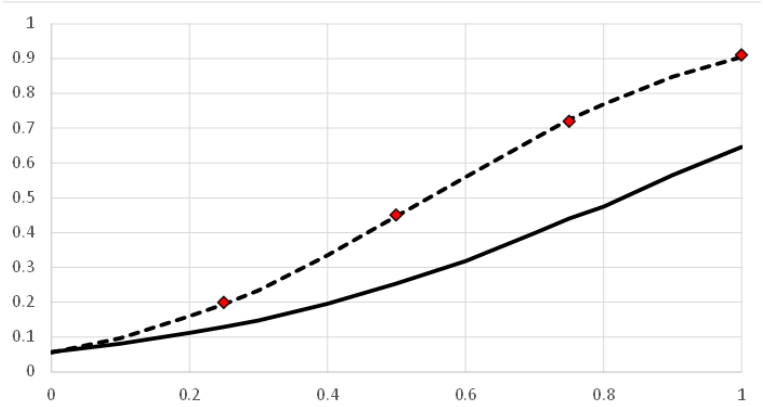


Figure 6 bis. Comparison of the power of Kaul and Wolf's confidence interval method (continuous line) with a t-test (dashed line), at equal levels of significance, obtained by Monte Carlo simulation with 50,000 iterations (using Kaul and Wolf's sudden change model). The red dots represent the values of column IM-1 of Table 2 in [10].

[3] Wakefield MA, Hayes L, Durkin S, et al. Introduction effects of the Australian plain packaging policy on adult smokers: a cross-sectional study. *BMJ Open* 2013;3:e003175. doi:10.1136/bmjopen-2013-003175

We see that we are getting the same power values for the *t*-test as those listed by Kaul and Wolf in Table 2 of their second paper. Our *t*-test must therefore be very similar to the one they use, let alone being exactly the same.

Re *3.4 Issue # 4: Non standard, ad-hoc method*

We maintain our point.

Re *3.5 Issue #5: Contradiction and lack of transparency about the way the data was obtained*

We maintain our point. We think that the last statement made by the prof. Kaul and Wolf is incorrect. They say that Roy Morgan data are *not* proprietary. They are. Single Source is a registered trademark of Roy Morgan Research Ltd. The company owns the data it sells. The data are thus proprietary, not public.

Re *3.6 Issue #6: Conflict of interest not fully declared?*

We maintain our point. We should like to add that conflict of interests is something that needs to be declared specifically and not something that should be left to the reader to guess. The role of IPE in the project is not declared. If this consulting firm also receives funding from PMI, this should be stated in the papers, as the authors seem to have a vested interest in the company. This needs further clarification.

Re *3.7 Issue #7: Lack of Peer Review*

We maintain our point. We should like to add that the debate has indeed taken place, but only after the political use of the authors' papers by the tobacco companies, which were immediately publicized on a wide scale, using in the case of the second paper the Reuters news agency as relay to send the message to the public and to political decision makers across the world. The imprint this has caused in their mind is not likely to be changed by the methodological debates occurring in scholarly journals.