

**Web-only supplement:**

**The use of the checklist of validity of predictions in HIA – the example of the EU withdrawal subsidies for fruits and vegetables**

The use of the checklist is illustrated with the example of an HIA we conducted on the EU Common Agricultural Policy to withdraw fruits and vegetables (FV) from the market when prices drop below an intervention-threshold. The withdrawn products are mostly composted. The assumption underlying the assessment was that ending withdrawal support would maximally lead to a proportional increase in consumption equal to the increase in availability of FV. The health gain for the Dutch population in this scenario was estimated at 1930 DALY per year or an increase in life expectancy by 3.8 days for men and 2.6 days for women.[6] Below, we highlight some of the points an independent assessment of the validity of our study could focus on, without of course intending to relieve the assessors from their responsibility to make their own judgements.

**Plausibility**

The plausibility of this study is best assessed by a team that includes epidemiologists and economists specialised in international agricultural trade and econometrics.

*Initial conditions*

Is the policy plan / project described accurately ?	Leaving aside difficult issues of political feasibility, the brief descriptions of the current EU policy and the proposed intervention should be judged. One question would be how soon such policy change could enter into force.
Is the description of the baseline situation accurate?	The data on amounts withdrawn and FV consumption were relatively old. The amounts may change rapidly while consumption patterns are likely to remain stable.
Has uncertainty in the initial conditions been assessed?	Only with respect to the amount of FV withdrawn from the market, not consumption or health outcomes. Is this justified by the assumption that uncertainty in those factors is relatively minor?
How robust is the model to (foreseeable) changes in the initial conditions?	Maximum amounts and compensation for withdrawals were to be lowered for a number of products over the years; this was not taken into account. Would this lead to overestimation of the potential effects? We ignored trends in FV consumption and disease occurrence. Is this justified?

*Theoretical framework*

Is the causal web underlying the analysis valid according to the state of the pertaining scientific field?	The association between FV consumption and health is generally accepted. Validity assessment could focus on the plausibility of the CAP withdrawal policy influencing FV consumption.
Is the order of magnitude of the causal relations in concurrence with current scientific knowledge?	The effects of changes in FV consumption on health as these were taken from recent reviews. RRs were not age-adjusted and applied uniformly to all ages. Would this bias

	results? The effect of policy change on FV consumption was only explored in a 'maximum effect' scenario. A more realistic scenario would require an econometric equilibrium model. We did not find one, and we also found no similar analysis in the literature.
Has the degree of certainty of the causal relations been described?	Two sources of uncertainty have been taken into account: the amount withdrawn and the relative risks of disease for changes in FV consumption. However, additional uncertainty remains, especially in the effect of policy change on FV consumption. Was the present analysis sufficient?
Are all exposures to determinants of health that are likely to result from the intended policy/project included in the analysis?	Other effects (e.g. on FV producers) are conceivable, but were estimated to be negligible compared to the effects on FV consumption. Is this justified?
Of the exposures included, have all plausible health outcomes been included?	We included CVD and cancer at a number of sites. For other diseases the evidence was deemed insufficient by the authors of the reviews we based the analysis on.
Have all populations likely to be affected by the policy been included in the analysis?	The analysis was restricted to the general Dutch population, with a qualitative comment that reform is likely to benefit low SES groups more than proportionally.
If available, how do the results of similar exercises compare with the predicted effects in this HIA? Can any differences be satisfactorily explained by differences in the initial conditions (including intervening events during the period of analysis) or lack of formal validity of the previous analyses?	To our knowledge no similar exercise has been conducted, but an independent assessor might know of similar work.

### Formal validity (verification)

Formal validity could be assessed by the same team that assessed plausibility. This team would probably want to have the spreadsheets used for the analysis.

#### *Initial conditions*

Have the right methods been applied and have these methods been applied correctly?

	<i>Correct method</i>	<i>Correct application</i>
Description of policy proposal	We based our brief description mainly on a previous health-focused analysis.[19] It might have been more elegant to refer to EU documents and reports.	Assessors could check for inaccuracies in the description.

Description baseline situation	See above. Would an independent assessor agree with our choice of the source of data on FV consumption and disease occurrence?	Assessors could check the spreadsheet and the paper for inaccuracies in the numbers.
--------------------------------	--	--

*Theoretical framework*

Have the right methods been applied and have these methods been applied correctly?

	<i>Correct method</i>	<i>Correct application</i>
Construction of causal framework	Lacking an econometric model, the method to reason from policy change to FV consumption was rather simple. A life table approach was used to model the health effects of changes in FV consumption.	Though it had limitations, the analysis was straightforward.
Estimation of magnitude of causal relations	For the relation policy - consumption simple assumptions were made. We conducted a PubMed search for recent meta-analyses or reviews on the relative risks of FV to disease, and contacted authors of older work if no result.	Perhaps other estimates for the RRs for cardiovascular disease would have been found if other authors had been approached.
Estimation of degree of certainty of causal relations	Uncertainty in the effect of policy on consumption could only partly be included; for the RRs confidence intervals in reviews were used. Bootstrapping was used to assess the overall uncertainty.	Check the procedures used and re-run a bootstrap procedure.
Search for significant determinants of health of which exposure changes as a result of the proposed policy	No formal search was conducted because substantial other effects were not deemed plausible.	Not applicable.
Search for health outcomes that result from changes in exposure	Relied on reviews of the effect of FV consumption on CVD and cancer. Restricting inclusion to diseases with statistically significant relationship with FV may lead to underestimation of effect.	Was the search strategy correct?

Search for populations likely to be affected by the policy	No search; restricted to general Dutch population.	Not applicable.
--	--	-----------------

### **Predictive validity**

#### *Historical predictive validity*

Are historical data on initial conditions and subsequent outcomes available on which the model underlying the HIA can be tested?	Such data would need to link changes in agricultural policy to FV consumption and health. We do not know of any.
If testing has been performed, how well does the model 'postdict' these outcomes, and can any differences between model and empirical data be explained satisfactorily by differences in the initial conditions or uncertainty in initial conditions (including intervening events during the period of analysis) and/or outcomes?	Not applicable.

#### *In retrospect*

To what extent did the predictions materialise?	Even if the proposed policy change would be effectuated, it would be impossible to measure any effect on population health. It might be possible to measure changes in FV consumption but even this would require a large sample size to detect the modest changes that are expected.
---	---