

1. <http://www.councilforresponsiblegenetics.org/ViewPage.aspx?pageId=88>
2. Wojcicki, A. *Nat. Biotechnol.* **31**, 1075–1076 (2013).
3. <http://www.washingtonpost.com/wp-dyn/content/article/2010/05/12/AR2010051205156.html>
4. Vayena, E. & Prainsack, B. *Nat. Biotechnol.* **31**, 16–17 (2013).
5. Wray, N.R. *Nat. Rev. Genet.* **14**, 507–515 (2013).
6. Gibson, G. *Nat. Rev. Genet.* **13**, 135–145 (February 2012).
7. Do, C.B., Hinds, D.A., Francke, U. & Eriksson, N. *PLoS Genet.* **8**, e1002973 (2012).
8. Klein, R.J. *BMC Genet.* **8**, 58 (2007).
9. Spencer, C.C.A., Su, Z., Donnelly, P. & Marchini, J. *PLoS Genet* **5**, e1000477 (2009).
10. <http://www.genomicsengland.co.uk/100k-genome-project/>
11. <http://www.fastcompany.com/3018598/for-99-this-ceo-can-tell-you-what-might-kill-you-inside-23andme-founder-anne-wojcickis-dna-r>
12. http://gds.nih.gov/pdf/PTC_for_IRBs_and_Institutions_revised5-31-11.pdf
13. Dorfman, R. *Nat. Biotechnol.* **31**, 785–786 (2013).

Science-based risk assessment requires careful evaluation of all studies

To the Editor:

In a Correspondence in the May issue, Romeis *et al.*¹ argued that in the science for risk assessment of genetically modified (GM) crops, a “careful evaluation of test protocols is a critical step that is sometimes missed.”

We fully agree with this point. However, we completely disagree with their argument that this careful evaluation is only necessary or important “when a study’s conclusion challenges an accepted body of knowledge.” If science claims universality and impartiality, then the idea of selecting only certain studies for careful evaluation is not scientific.

The proposal from Romeis *et al.*¹ to select and evaluate only those studies with findings challenging the dominant wisdom seems even worse.

Surely test protocols for all studies should be subject to the same level of careful evaluation and respectful but critical scrutiny, no matter what their findings and conclusions. Without universal consistency in our critical evaluations, we cannot test but only presume on faith or prejudice that “an accepted body of knowledge” has been developed according to robust study designs. For the risks posed by crops expressing the insecticidal protein isolated from *Bacillus thuringiensis* (*Bt*) to nontarget organisms, careful evaluation suggests that this is arguably not the case^{2,3}. It seems particularly paradoxical to us that Romeis *et al.*¹ try to critique “selective reporting” of scientific literature by certain European countries and/

or the media, while arguing for a selective application of the careful scrutiny of study designs.

Selective reporting is also manifest in their own Correspondence in several ways.

In the first instance, and rather strangely, Romeis *et al.*¹ selectively report from their own work in a way that ultimately misrepresents their results. In the text of their Correspondence, the authors state that from their review of 36 peer-reviewed publications, four have reported putative toxic effects of the Cry1Ab insecticidal protein on nontarget organisms. In the table

summarizing these studies presented in their Supplementary Notes, however, at least three other publications also report putative toxic effects, whereas an additional four report finding significant negative effects but claim that these are not related to the Cry proteins. Why readers are given the false impression in the main text that only four publications report toxic effects when their own work as presented in the Supplementary Notes indicates otherwise is unclear. Furthermore, no justification is given for the selection of those four particular publications for a deeper level of critical scrutiny, instead of, for example, selecting those studies that report negative effects but claim that this is not related to Cry proteins.

Romeis *et al.*¹ claim that all studies showing negative effects of the *Bt* toxin Cry1Ab on nontarget organisms are of a “nonconclusive nature.” However,

according to the table they provide in their Supplementary Notes, only one study of the 55 reviewed was actually able to be confirmed as meeting all of the quality criteria they assessed. Additionally, in the Supplementary Notes, the authors state that the three quality criteria of test substance purity, homogeneity and equivalence between the purified and plant-expressed protein “are not addressed by the majority of the published studies and are thus not presented in detail.” Since few (if any) of the 55 published studies can therefore be said to meet the nine quality criteria Romeis *et al.* identify as important, all studies could, thus, arguably be presented as “nonconclusive.”

In drawing the conclusion that only studies finding negative effects are “nonconclusive,” Romeis *et al.*¹ also selectively report information by failing to mention the debates surrounding these so-called “outlier” studies that they deem to be generating “false-positive effects,” according to their standards. These debates have taken place in peer-reviewed journals^{4–6} and have involved the scientists responsible for those studies responding to the claims and accusations that their results are invalid due to faulty study designs. For Romeis *et al.*¹ to hide such scientific debate from readers and simply assert that the research is unreliable and nonconclusive following their selective evaluation is not only misrepresentative, it falsely claims a scientific consensus when the issues in fact remain open. In these published debates, it is particularly interesting to note that whereas the scientists finding possible harm recognize and explicitly acknowledge the inherent openness of science and thus the need for continual collective exchange and further research, Romeis *et al.*¹ and other critics of the work seem to present their own science as conclusive and final, with no need for further investigation required (of either body of work). Surely a diversity of approaches, methods and possibly findings is not only normal, but quintessentially what constitutes good science, and also crucial for its public trustworthiness.

An equally problematic form of selective reporting lies in the way Romeis *et al.*¹ present the results from the meta-analyses. They state that numerous experimental studies and field observations have provided conclusive evidence that Cry1Ab protein expressed in *Bt* maize does not cause adverse effects on arthropods, and that supporting data for this have been analyzed in reviews and meta-analyses. What they fail to report, however, is that in at least one of the meta-analyses they cite⁷, the results did in fact



demonstrate the opposite—a significant reduction of nontarget invertebrates in Cry1Ab maize generally, and for MON810 maize specifically, when the comparators were fields with no insecticide applications. This was based on analyses of 24 studies and 415 tests in the former, and 11 studies and 235 tests in the latter. When non-GM fields were sprayed with insecticides, there was a higher invertebrate abundance in Cry1Ab maize generally, but not in MON810 (based on 8 studies/90 tests and 4 studies/39 tests, respectively). This means that for GM crops generally and for the European controversy surrounding the approval of MON810 specifically, depending on what is considered to be the appropriate comparator for defining adverse effects, the body of available scientific knowledge leads to alternative conclusions. Romeis *et al.*¹ have inserted a (legitimately debatable) policy norm into their ‘science’ and then tried to suggest that their conclusions are final.

What is a relevant comparator, and other questions raised by Romeis *et al.*¹, including what are our environmental protection goals and therefore what do we deem relevant risk-based science for policy, are not scientific questions alone but rather questions for policy authorities accountable to democratic processes. These types of questions are part of what has been termed ‘risk assessment policy’ (RAP)⁸. RAP is a normative process that both precedes and iterates with scientific risk assessment, shaping its orientation, boundaries, mandate and inferences, which makes risk assessment not exclusively a scientific process, as claimed by Kuntz *et al.*⁹ in another recent Correspondence in *Nature Biotechnology*. The significance of this has been recognized by organizations, such as the United Nations Food and Agriculture Organization (Rome), the World Health Organization (Geneva), the US National Research Council (Washington, DC) and the European Commission (Brussels)¹⁰. The question of how to judge the relevance of a scientific study for risk assessment should therefore not be restricted to the scientific community (or worse to a small set of self-appointed researchers), but rather is an issue that requires broader-based and accountable consideration.

In addition to the question of relevance, the question of quality in science for policy also requires broader-based consideration. This is because what we are dealing with here is not simply a romanticized pure science with no social context or consequences, but rather a science specifically conducted to inform policy-making in a controversial

arena of great public interest (i.e., the future of food production and biodiversity). Given this, the task of defining the criteria for evaluating quality cannot be legitimately nor sufficiently decided by Romeis *et al.*¹ alone, nor even the world of their peers. It is not for an esoteric group of researchers or experts to pronounce what constitutes quality in science for policy and then dismiss as irrelevant and unreliable any other science not corresponding with how their own particular interests and values have defined what is important. This issue also deserves (and in a democratic society, demands) deliberation across a wide range of stakeholders and implicated actors.

Such broad-based deliberation on quality criteria is crucial, not because science “is simply a matter of opinion, no better than any other opinion,”⁹ but because, as Romeis *et al.*¹ correctly highlight, scientific research can be conducted under a range of test protocols and study designs, with the potential for each of these to lead to contrasting results. Which of the range of possible research approaches is best, with the most relevance for risk assessment and the highest quality as fit for purpose, will inherently depend on what questions are deemed most salient and what is most valued by a broad section of society. That is, it is both a scientific and a social matter.

What is at stake here is not only the future of GM crops but also the role and scope of scientific inputs to governance processes; thus, it is crucial that we maintain and enhance the credibility and legitimacy of those inputs. This requires, *inter alia*, impartial and consistent quality evaluation of all scientific studies, no matter what their conclusions.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Fern Wickson¹, Thomas Bøhn^{1,2},
Brian Wynne³, Angelika Hilbeck^{1,4,5} &
Silvio Funtowicz⁶

¹GenØk Centre for Biosafety Society, ²UiT The Arctic University of Norway, Tromsø, Norway.

³Lancaster University, Lancaster, UK. ⁴Centre

for Biosafety, Tromsø, Norway. ⁵Swiss Federal Institute of Technology, Zurich, Switzerland.

⁶University of Bergen, Centre for the Study of the Sciences and Humanities, Bergen, Norway.

e-mail: silvio.funtowicz@svt.uib.no

1. Romeis, J. *et al. Nat. Biotechnol.* **31**, 386–387 (2013).
2. Wickson, F.J. *J. Risk Res.* **12**, 749–770 (2009).
3. Dolezel, M. *et al. Environ. Sci. Europe* **23**, 33–48 (2011).
4. Bøhn, T. *et al. Environ. Sci. Europe* **24**, 22 (2012).
5. Hilbeck, A. *et al. Environ. Sci. Europe* **24**, 10 (2012).
6. Wickson, F. & Wynne, B. *Ethics Policy Environ.* **15**, 321–340 (2012).

7. Marvier, M. *et al. Science* **316**, 1475–1477 (2007).
8. Millstone, E. *et al. Risk Assessment Policies: Differences Across Jurisdictions* <http://ftp.jrc.es/EURdoc/JRC37719.pdf> (Joint Research Centre, European Commission, 2008).
9. Kuntz, M. *et al. Nat. Biotechnol.* **31**, 498–500 (2013).
10. Food and Agriculture Organisation of the United Nations and World Health Organisation. *Food Safety Risk Analysis: A Guide for National Food Safety Authorities* <ftp://ftp.fao.org/docrep/fao/009/a0822e/a0822e00.pdf> (FAO and WHO, 2006).

Romeis *et al.* reply:

We believe that Wickson *et al.*¹ have confused different elements of an environmental risk assessment (ERA), and did not understand or chose to ignore the purpose of our original Correspondence². In addition, we strongly disagree with their contention that we believe that a “careful evaluation is only necessary or important” for “studies with findings challenging the dominant wisdom” when our Correspondence clearly describes the importance of ensuring the quality of all data that are used to inform environmental risk assessments.

Various stakeholders, including society writ large, should be involved in defining which elements of the environment are valued and should be protected from harm (i.e., protection goals). Even so, in the context of risk assessment and regulation of transgenic plants, it is the responsibility of both the scientists conducting the work and the regulators who interpret their results and conclusions to ensure that experimental studies are conducted properly. A concern about transgenic maize events that produce the Cry1Ab protein from *Bacillus thuringiensis* (*Bt*) is their potential to harm arthropods outside the order of the target pests, that is, Lepidoptera. This concern is derived from the value that society places on many of these nontarget arthropods (NTAs) because of the important ecosystem services they provide, such as biological pest control, pollination and decomposition. For insecticidal *Bt* maize to cause harm to a nontarget species, the species must be sensitive to the insecticidal protein; the associated risk hypothesis is “The *Bt* protein is not toxic to valued nontarget species at concentrations present in the field.”³ To test this hypothesis, purified Cry1Ab protein or *Bt* maize tissue that contains Cry1Ab is provided to nontarget species in controlled laboratory studies. As we state in our Correspondence, such NTA studies need to be carefully designed to obtain high confidence in the results and to minimize the possibility for false positives and false negatives.