Commentary

Reply to 'On the adequacy of scope test results: Comments on Desvousges, Mathews, and Train'

We did not misinterpret the scenarios described to respondents, as Bishop et al. (the authors) incorrectly claim. The authors provide a new interpretation based on information that was not given to respondents either graphically or verbally. However, the difference in interpretation is moot: as we show below, our criticisms of their study still hold, exactly the same, under the authors' new interpretation. Moreover, we have subsequently investigated the issue empirically – which the authors could have done themselves – and found that the hypothesis of adequate response to scope in their study is strongly rejected.

In their Comment, the authors provide new Figs. 1 and 2 as time-paths for recovery under their two scenarios, stating in a footnote that these graphs were not actually shown to the respondents.[1] However, these new graphs do not change anything: all of our criticisms hold exactly the same under the new graphs as for the original graphs.

- By geometry, areas a + b + c in their new Fig. 1a is four times larger than areas f + g in Fig. 2a, such that Scenario 1 provides four times as many lake service-years as Scenario 2. But Scenario 1 was estimated to be valued only a third more than Scenario 2. If river service-years have any value, the comparison is even more problematic, since Scenario 1 includes river service-years d + e that are not included in Scenario 2.
- The ratio of lake services obtained in Scenario 1 at any given time to those obtained in Scenario 2 decreases over time. So any positive discount rate over time makes the comparison even more problematic: the present value of the lake services under Scenario 1 is *more than* 4 times greater than under Scenario 2.
- The services under Scenario 1 were estimated to be worth $184 and those under Scenario 2 to be worth $138, which implies that the lake and river services obtained in Scenario 1 that were not obtained in Scenario 2 (areas d + e + a + b + c minus f + g) are estimated to be worth $46. By Figs. 1 and 2, these extra services are greater in all dimensions than the services under Scenario 2: more lake services (a + b + c–f–g > f + g), closer in time, and river services in addition to only lake service (d + e, which is not included in Scenario 2). As we stated in our article, this comparison indicates that the study's estimates violate the scope criterion: a greater scope of services is estimated to be valued less ($46 compared to $138).

Even though the authors provide Figs. 1–2 to further explain their scenarios, they disavow their use for evaluating the scenarios. They say that neither the graphs shown to respondents (Figs. A and B),

their new graphs, nor any other graphs should be used to represent their scenarios, because the time-paths of recovery are not known. Specifically, the authors state that "calculating present values of environmental benefits, as they [we] propose to do, would require detailed knowledge of how the environmental services and their values changes as the restoration proceeds over that 60 year time period."[2] However, this statement indicates that respondents were given insufficient information to evaluate the benefits of the program relative to the bid amount (the cost.) The authors argue that the value of the program depends on how the environmental services changed over time, and yet the survey did not provide this information to the respondent. So the authors violated a fundamental requirement of CV studies, namely, that the program must be described in sufficient detail to allow the respondent to evaluate its benefits relative to costs. The authors have jumped – to put it colloquially – from the frying pan into the fire: the argument that they use to deflect our criticism about inadequate response to scope creates an even larger problem for their study, that respondents were not given the information needed to evaluate the program.

The arguments so far have been made in logical grounds. Basically, we are arguing that the $46 difference is too small for the reasons we give, and the authors are arguing that the $46 might not be too small. But why argue about it theoretically? It can be measured empirically. And in fact we have done that, as described in Desvousges et al. (2015). To accomplish this, we modified the questionnaires as needed to represent the extra benefits that are included in Scenario 1 but not in Scenario 2 (accounting for the incremental nature of the benefits). We obtained an estimated value far in excess of $46. The hypothesis of adequate response to scope is rejected empirically at 99% confidence level.[3]

This testing required very little modification of the original survey, with the changes mirroring those already made by the authors for their differentiation of Scenarios 1 and 2. This brings us to the question that is most perplexing to us: why didn't the authors perform this test themselves? The reliability of their study hinges on whether households are actually willing to pay only $46 extra for the services in Scenario 1 that are not in Scenario 2. On its face, $46 seems very low relative to

---

[1] The authors seem to imply in the footnote that they provided verbal descriptions of paths to recovery: "Neither survey version used such graphs [as Figs. 1–2]. Rather they depended on verbal descriptions of recovery." In fact, the survey contained no verbal description of time paths to recovery, just the descriptions we reproduce in Figs. A and B about the resources returning to 1960 levels 10 or 40 years sooner.

[2] As well as the time-path of recovery being unknown, the authors say in this quotation that how households' values change over time in not known. As we stated in our paper and above, any positive discount rate applied to values over time exacerbates the problems with their estimates. The authors might be suggesting (though this suggestion is perhaps more specific than they were actually making) that services in the distant future are valued more highly than services in the near future, which is equivalent to applying a negative discount rate. We return to this issue in footnote 3.

[3] We also estimated the value that respondents placed on services in the more distant future relative to the less distant future, by comparing the authors' Scenario 2 with a similar scenario that is specified the same except less distant in time. Respondents' values were found to be higher for services that are closer in time, consistent with a positive discount rate.
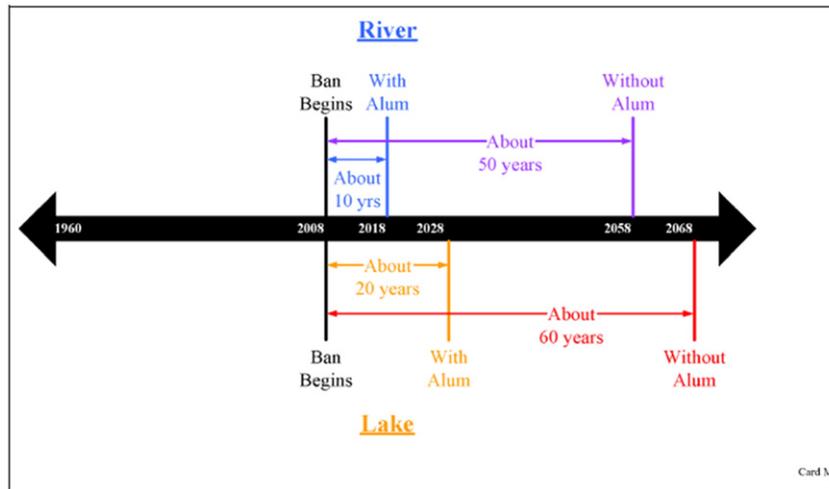
Fig. A. Information Given to Respondents in Bishop et al.'s Survey for Scenario 1. The interviewer says the following while showing Card M: "*Without alum treatments, it will take about 50 years* (*POINT*) *for the river to get back to what it was like in around 1960* (*POINT to 1960*). *That is about 10 years* (*POINT*). *That is about 40 years longer. It will take the lake about 60 years* (*POINT*) *to get back to what it was like in around 1960* (*POINT TO 1960*) *instead of about 20 years. That is also about 40 years longer.*" The cost of the program was introduced by the interviewer saying: "…*the river and lake will naturally return to what they were like in around 1960. If the people of Oklahoma want this to happen 40 years sooner, there will be an additional cost for the alum treatments.*" Immediately before voting, the respondent is told: "*Voting for the program means* (*PAUSE*) *that it is worth it to you* (*PAUSE*) *for your household to pay the additional one-time tax of* $(*BIDAMT*) (*PAUSE*) *to return the Illinois River, Flint Creek, Barren Fork Creek, the smaller creeks flowing into them, and Tenkiller Lake to what they were like in around 1960 40 years sooner.*"
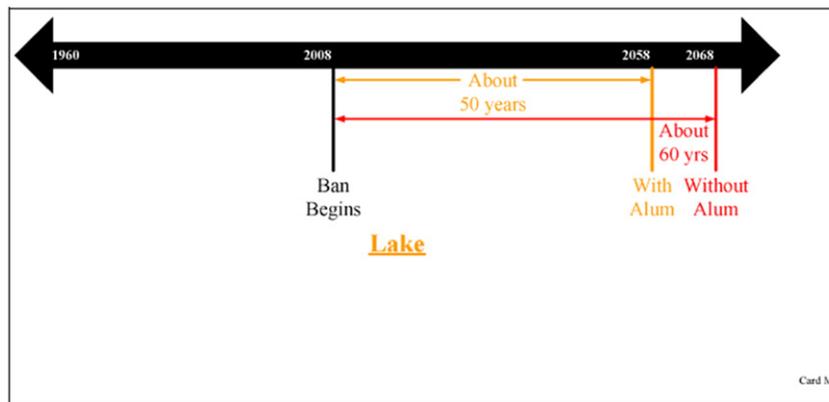

Fig. B. Information Given to Respondents in Bishop et al.'s Survey for Scenario 2. The interviewer says the following while showing Card M: "*Without alum treatments, it will take the lake about 60 years* (*POINT*) *to get back to what it was like in around 1960* (*POINT TO 1960*) *instead of about 50 years. That is about 10 years longer.*" The cost of the program was introduced by the interviewer saying: "…*the lake will naturally return to what it was like in around 1960 in 60 years. If the people of Oklahoma want the lake to return to what it was like in around 1960 in 50 years rather than 60 years, there will be an additional cost for the alum treatments.*" Immediately before voting, the respondent is told: "*Voting for the program means* (*PAUSE*) *that it is worth it to you* (*PAUSE*) *for your household to pay the additional one-time tax of* $(*BIDAMT*) (*PAUSE*) *to return Tenkiller Lake to what it was like in around 1960 in 50 years rather than 60 years.*"

the other estimates, and the authors' only argument is that $46 can't be ruled out. Instead of arguing about it, why not test it empirically?

Potential difficulties in conducting an adding-up test have been given as a reason not to estimate the willingness to pay (WTP) for a scenario that represents the extra benefits in one scenario that are not included in another. These arguments are addressed in Desvousges et al. (2015). However, these arguments reflect a strange reversal of position by CV proponents: It is often said, in justifying CV estimates in the face of criticism, that "some number is better than no number"; but when it comes to estimating the WTP for an extra scenario that actually tests for adequate response to scope, the sentiment switches to "no number is better than some number."

In conclusion: The NOAA panel stated that the burden of proof rests on the CV researchers to show that their study does not evidence inadequate response to scope. By the authors' own statements, they have not met this burden.[4] We have argued that their results, by analysis of the original graphs for respondents and the authors' new graphs of

time-paths of recovery, evidence inadequate response to scope and even a scope reversal. The authors claim that our analysis is invalid because the time-paths of recovery for their scenarios are not necessarily those shown on any graphs. This claim creates an even greater problem: that the survey did not provide sufficient information for respondents to be able to evaluate the program. Rather than arguing theoretical possibilities, we performed an empirical test, which the authors could have performed themselves, and found that the hypothesis that the study's estimates evidence adequate response to scope is strongly rejected.

References

Arrow, K., Leamer, E.E., Schuman, H., Solow, R., 1994. Comments of proposed NOAA scope test. Appendix D of Comments of Proposed NOAA/DOI Regulations on Natural Resource Damage Assessment. Environmental Protection Agency, U.S.

Desvousges, W., Mathews, K., Train, K., 2015. An adding-up test on contingent valuations of river and lake quality. Land Econ. 91 (3), 556–571.

---

[4] NOAA panel members (Arrow et al., 1994) explicitly said that the type of scope test that the authors performed does not meet their criterion for demonstrating adequate response.