

Project Name: Bone Morphogenetic Protein: The State of the Evidence for On-Label and Off-Label Use
Project ID: BMPE0109

Table 1: Invited Peer Reviewer Comments

Reviewer ¹	Section ²	Reviewer Comments	Author Response ³
1	Executive summary	The summary is long and hard to follow. It presents an overview of the methods rather than findings. The executive summary table lacks any synthesis of study size, source, quality, and outcomes. The text within the “conclusion column” re-iterates methods and states a general conclusion for each key question, without giving constructive information and basis for the conclusion.	The Executive Summary – like any summary or abstract – synthesizes methods to let the reader know how the assessment was done. The conclusions column in the summary table provides a high-level synthesis of the evidence review for each Key Question based on the AHRQ-modified GRADE framework. This takes into account study size, source, outcomes, and quality based on the USPSTF criteria.
	Introduction/Background	<ul style="list-style-type: none"> • The introduction provide good general context for BMP. The assertion that age influences fracture healing needs to be supported by citations (page 13, paragraph 5). • Animal studies supporting bone formation properties of BMP are not discussed. • The FDA approval studies and data would be helpful, such as FDA summary and effectiveness data and specific FDA prescription, training, and labeling information as an appendix. 	<ul style="list-style-type: none"> • A reference citation was provided on page 13, paragraph 5 regarding patient age as a factor in bone healing. • Animal studies are outside the scope of the assessment. • Published data describing results from the FDA pivotal trials are included and assessed in the report.
	Methods	<ul style="list-style-type: none"> • The search strategy was extensive. 	<ul style="list-style-type: none"> • The search was designed to

		<p>The selected time window of 1998-2009 is not explained but seems reasonable, given the FDA approval dates.</p> <ul style="list-style-type: none"> • The patient population description is vague. Again, reference to the populations studied in the FDA approval application may be helpful. • Separating fracture and spine studies is critical, but the methods do not describe this or explain why not. • “DDD” is not a skeletal bone defect, and neither is an arthrodesis procedure. • The discussion of radiographic outcomes for both fractures and spinal fusion needs more details and referencing. FDA definitions would be helpful, both for radiographic success and clinical success. • Neurological status outcomes description is not specific enough for application to literature review and it 	<p>take into account the FDA approval dates.</p> <ul style="list-style-type: none"> • The population description was broad, patients with a bony defect that requires repair. However, the KQs address specific indications for on-label uses, leaving the off-label uses less specific because it was unclear at the beginning what would be found in the literature. • Spine and fracture results are reported separately in the Results. • The reviewer is correct in that “DDD” is not a skeletal bone defect nor is an arthrodesis procedure. These errors are corrected in the text. • Radiographic outcomes reported from included articles conform to FDA definitions and are in accord with standard use. • The neurological status outcomes description is reported as it is used in the
--	--	---	---

		<p>contains no references.</p> <ul style="list-style-type: none"> Methods described for harms assessment (key questions 7 and 8) are particularly problematic. Assessment of harms is very limited. This aspect of the review is perhaps the most important from patient perspective, and the information presented does not synthesize well the safety characteristics of BMP. McMaster and AHRQ are cited as the source for the harms ascertainment, but even these modified questions are not addressed in the body of the report or in tables 36 and 37. How were harms defined and ascertained in the FDA studies? How were they defined and ascertained in the published studies? How do they compare across BMP-products? How do they compare in on-label vs. off-label applications? These questions are not answered by the report. Although Table 36 and Table 37 contain a lot of information, the information is not well-organized. The data are not synthesized in any structured way. The systematic harms ascertainment methods advocated by AHRQ would have been very helpful (AHRQ series paper 4: assessing harms when comparing medical interventions: AHRQ and the effective health-care program. Chou R, Aronson 	<p>Neck Disability Index instrument.</p> <ul style="list-style-type: none"> A major conclusion of this report is that the quality of reporting of harms in the literature is inconsistent. Our team systematically culled out harms data from every included comparative study and looked at noncomparative studies for those data (compiled in the Appendix). We included noncomparative studies because of the known limitations of harms reporting in RCTs. The results we compile represent what was actually reported (Table 36). The results of our modified McHarms survey highlight the limitations of the reporting (Table 37), which was the purpose of KQ8. Given the inconsistency and lack of comparability across studies, quantitative synthesis of the data is not valid. We state that the absence of harms reporting in a study does not necessarily provide evidence of the absence of harms. We agree about the importance of harms to patients, and believe our assessment underscores
--	--	---	--

		<p>N, Atkins D, Ismaila AS, Santaguida P, Smith DH, Whitlock E, Wilt TJ, Moher D. J Clin Epidemiol. 2010 May;63(5):502-12. Epub 2008 Sep 26. PMID: 18823754).</p> <ul style="list-style-type: none"> • The methods for data abstraction, verification, and creation of evidence table should be more specific and clear. Exactly what constituted primary data for the different study designs? How were accuracy checks performed? • Use of USPSTF study quality criteria in Appendix 5 is very helpful; condensing the criteria into “good-fair-poor” categories makes it hard to evaluate the studies in the body of the report is less useful. The low rate reporting on how harms were identified needs to be highlighted in the body of the report; if this is not described, the remaining five characteristics are not really relevant. • Assessment of applicability is difficult to interpret based on the description in the methods. Rather than describing the EPICOT framework in a general sentence, it would be more useful to 	<p>the weakness of evidence in the literature and the need for better collection and analysis of such data using validated methods</p> <ul style="list-style-type: none"> • Data abstraction and analysis methods are presented on page 26 of the draft. Study selection criteria are described on page 25. Accuracy of abstracted data was verified by a second investigator with differences resolved by discussion or third party intervention as needed. • We systematically abstracted all BMP-related harms from all included articles, comparative and noncomparative. The quality of harms reporting was addressed by KQ8 using a modified McHarms survey. The USPSTF quality ratings are presented in the Appendix with annotation; overall ratings are carried over into the report tables and text. • It is unclear why comparing the included study populations to those in the FDA approval
--	--	---	---

		compare the study populations to the patients enrolled in the FDA approval studies. Also, separating spine and tibia studies would provide more clear applicability assessment. The report has minimal figures; the two on decision modeling are clear.	studies would make the report “more useful”. The results of FDA pivotal trials are included in the assessment. Spine and tibia studies are reported separately in the assessment.
Results	<ul style="list-style-type: none">Appendix 1 contains an amazing collation of very important information. The research team has done an outstanding job of assembling this information in Tables A through P. However, this information is not captured or synthesized well in the results section. The text is difficult to follow. Interventions, populations, benefits, and harms are interjected variably.	<ul style="list-style-type: none">We appreciate the reviewer’s kind words about the collation job we have performed. We disagree with the reviewer in that we do not believe synthesis of the data using quantitative measures is applicable, but rather would in most cases be inappropriate due to interstudy heterogeneity and size differential. The Appendix data are qualitatively synthesized according to KQ and indications in the body of the report.	
Key Questions	<ul style="list-style-type: none">The text does not directly answer the question. For example, the outcomes in Table 6 should be evaluated quantitatively, and the column headings defined for each of the three studies. No comparison is made to the	<ul style="list-style-type: none">Upon review of the draft, it became evident that the paper by Dawson et al, 2009 reported an off-label use. It was moved to the appropriate section and the text and tables were	

		<p>FDA premarket approval studies. The text addresses study methods, patient demographics, benefits, and harms, without directly comparing and contrasting these features in succinctly. In part, this is due to lack of explicit definition of success in the pre-specified work plan.</p>	<p>adjusted throughout the assessment to account for this. As a result, we were left with two studies, one much larger than the second, which would have overwhelmed if not negated the value of any qualitative analysis. The assessment compiles the data systematically, synthesizes it according to the AHRQ-modified GRADE convention, and reports it in that context.</p>
	Cost-Effectiveness Analysis	<p>This section of the report is the most detailed, well written, and clear. My only concern is the potential for bias in the source studies, charge/cost estimates, and poor quality of source data for transition probabilities</p>	<p>The following sentences have been added to the Discussion and Conclusion section:</p> <p>There was a limited evidence base for both open tibial fracture and spinal fusion, each consisting of a single randomized controlled trial. Biases may have existed in the source studies, for example possibly biased assessment of outcomes would result in inaccurate transition probabilities.</p>
	Summary and conclusions	<ul style="list-style-type: none"> Conclusions are justified by the data. However, the potential for sponsorship bias should be mentioned. Table 59 is again filled with large 	<ul style="list-style-type: none"> Sponsorship bias is not characterized. While the high proportion of industry sponsorship among the included studies suggests potential bias exists, it was not systematically investigated or quantified.

		sections of text rather than clear summary of numbers. More specific numbers rather than general estimates such as “low/moderate” would help better understand the answers to key questions. Study limitations are not addressed adequately in the summary.	<ul style="list-style-type: none"> Table 59 is the same as in the Executive Summary, and all comments on it were covered in the text above pertaining to that section of the report.
2	General	I found the extensive use of abbreviations to make reading difficult, as I was constantly looking back for definitions. I would recommend keeping the most familiar or obvious abbreviations (BMP, RCT, FDA, QALY), but simply writing out for each use the terms that are abbreviated with less familiar abbreviations (unless the meaning is reiterated in every section where they are used). Examples that threw me were AGB, ALGB, ICBG, FRA, DBM, HA-TCP, DSP, and the like.	Revisions will be made throughout the text to limit abbreviations.
	Executive Summary	Good	No response
	Introduction/Background	Good	No response
	Methods	Good	No response
	Results	<p>Generally good.</p> <ul style="list-style-type: none"> It would have helped me to identify exactly what made certain trials off-label use. The reason is that, at least theoretically, BMP might be effective for one off-label use, but not for others. Here, all off-label uses seem to be treated as one. For example, in off-label uses of BMP2 in the lumbar spine, were studies off-label because they were used for more than one spinal level, or because they were not for DDD, or for other reasons? Could 	<ul style="list-style-type: none"> Off-label uses in RCTs in the lumbar-sacral spine (table 23) were explicated in the table, and text was adjusted throughout to reflect changes in table. Trials were not otherwise separated. There are several different reasons the trials in table 23 were deemed off-label. These include use of a nonapproved formulation, or matrix, in

		<p>we separate trials of the different reasons they were off-label? Maybe BMP works for spinal stenosis, say, but not for multi-level fusions. We can't get a sense of that here.</p> <ul style="list-style-type: none"> • On page 33, there's a funny typo: below table 5, discussing lumbar fusions, there is a sentence that all patients had symptomatic single level DDD, but includes <i>arm</i> pain as a symptom. Not for the <i>lumbar</i> spine, I don't think • In the cost-effectiveness analysis, it took me a while to tumble to the baseline assumption that the costs of fusions with and without BMP were the same, thanks to DRG bundling. It would be nice to make this point more explicitly on page 81, since it is so 	<p>conjunction with the approved rhBMP2 (InFuse®); use of a non-anterior surgical approach with InFuse®; use of InFuse® with a nonapproved interbody entity; and, use in multi-level fusion. While the trials differed in rhBMP2 use, they were generally consistent in direction of effect, with statistically significant findings for radiographic success favoring BMP in three, including the two largest RCTs. This suggests the off-label factor(s) does not affect the result. BMP appeared to have benefit in this setting despite differences among the studies</p> <ul style="list-style-type: none"> • Reference to arm pain was deleted • The last sentence below was added to make this clearer: <p>Analyses included direct health care costs reported as Medicare payments from free publicly</p>
--	--	---	---

		counter-intuitive.	available sources, valued in 2007 U.S. dollars (Tables 44–49). Cost categories included initial hospitalization (hospital and physician costs) and secondary interventions (hospital/outpatient surgical center and physician costs). It was assumed that initial hospitalization was paid according to the diagnosis-related groups (DRG) system. Thus, base case analyses assume identical initial hospitalization costs whether BMP was used or not.
	Discussion/Conclusion	Discussion/Conclusion: Again, it would be helpful if conclusions regarding off-label use could be itemized by indication (the reason for being off-label)	See above
	Tables, Figures, Appendices	Good	No response
	References	I could easily have missed it, but I didn't see mention of the Cahill article on complications of BMP that appeared in JAMA during the search period (Cahill KS, et al. Prevalence, complications, and hospital charges associated with use of bone-morphogenetic proteins in spinal fusion procedures. JAMA 2009; July 1; 302: 58-66). Is there some reason?	Cahill et al presents a retrospective overview of complications associated with BMP use in spinal fusion, based on the Nationwide Inpatient Sample Database, a 20% sample of US community hospitals. It does not separate data according to BMP product, nor is it necessarily representative of BMP use in the US. It was excluded according to our predefined study inclusion criteria.

¹ Peer reviewers are not listed in alphabetical order.

² If listed, page number, line number, or section refers to the draft report.

³ If listed, page number, line number, or section refers to the final report.

Project Name: Bone Morphogenetic Protein: The State of the Evidence for On-Label and Off-Label Use
Project ID: BMPE0109

Table 2: Public Review Comments

Reviewer Name ¹⁷	Reviewer Affiliation ²	Section ³	Reviewer Comments	Author Response ⁴
Anonymous Reviewer 1	NA	General	This is a comprehensive review of the state of the evidence of on and off-label use of bone morphogenetic protein. This report is well written, to the point and well documented scientifically. I have been working in this field for over twenty five years and there are no grievous omissions in their background section or references. The report is well organized and broken into easily discernible sections and the Result section focuses on 10 Key Questions which have been identified by the group and pertinent to this body of knowledge. I concur with the assessment and the Key Questions. These have been carefully formulated and documented. The summaries and conclusions seem to be well supported and I believe the report to be objective. I did not see any sections that indicated investigator bias on the part of the team members. Therefore, in general I would rate the overall report as outstanding, objective and technically accurate.	We appreciate the comments. No further response.
		Executive Summary	The executive summary is precise, clearly written and outlines the medical aspects of the search and the methods in which the Key Questions were formulated and answered which is clearly illustrated in a summary table which evaluates the Key Questions and Conclusions. This is very easily readable and the conclusions are based on sound scientific evidence.	
		Introduction/Background	The Introduction/Background is to the point covering specific areas in which the “products” can applied as a	

	<p>substitute for bone graft. It is well referenced. Additionally, the indications for the FDA approvals are clearly outlined and referenced. I do not see any glaring omissions in this introduction and background section.</p>
Methods	<p>The Methods Section was clearly delineated, understandable and thoroughly covered all the areas of interest. Additionally, the group outlined specific questions, and gave appropriate references, and provided data analysis mechanisms of rating the body of evidence. I found this approach to be clearly stated and provided objectivity.</p>
Results	<p>Search results were clearly documented in the body of the text, as well as in numerous tables which were easy to read and well referenced. The statistical analysis and powers, emphasizes are clearly stated in the evidence that is presented according to the various key questions that were posed by the group. I found these questions, tables, and information to be accurate and clearly written. I thought this was an outstanding section and the approach allowed for logical conclusions and well documented judgments.</p>
Discussion/Conclusion	<p>Summary and Discussion/Conclusion section is precise to the point and easy to interpret and verify. I found this easy to follow and based on my own knowledge of it by the work and their methods believe that they came to logical rational conclusions based on scientific evidence.</p>
Tables	<p>Tables were detailed, many in number, however they were easy to read and referred back to specific points in the text. I believe these were supportive, although tedious to read.</p>

		Figures	N/A	
		Appendices	Complete, extensive and pertinent to the text of the body. The references were complete and no obvious omissions were made in the reference list.	
Anonymous Reviewer 2	NA	General	I and each of the five other spine surgeons in my hospital routinely use BMP for interbody and posterior lateral onlay fusions. It works. The goal of the surgery is fusion.	No response
Baker, Ray MD	NASS	Results	<ul style="list-style-type: none"> • Our main critique focuses on the need for clarification about the surgical approach implied by on-label and off-label use, as detailed in the executive summary. For example, in the response to question number 6, the authors cite studies that have demonstrated "cervical swelling" with use of BMP in the cervical spine. This, according to the studies cited, is quite specific to the anterior cervical approach. This should be made clearer in the executive summary. • The same critique applies to the statements about off-label use in the lumbar spine. Presumably, this is BMP for posterolateral fusion or posterior lumbar interbody fusion. This distinction should be made clearer. 	<ul style="list-style-type: none"> • So noted, with text and tables revised as suggested. • So noted, text and tables were revised to reflect these in the RCTs. Summary conclusions and GRADE tables were revised to reflect the

			<ul style="list-style-type: none"> Reasonable interpretation and extrapolation of the data supporting BMP-2 use inside an LT cage would support that use of BMP-2 in other cages or interbody implants has similar efficacy and results and should therefore not be categorized as a similar off-label indication as posterior lumbar fusion. 	<p>changes.</p> <ul style="list-style-type: none"> We acknowledge Dr. Baker may be correct in his assertion, but the assessment was based on strict adherence to the FDA-approved marketing label for each BMP product.
Callaghan, John MD, et al.	AAOS	Conclusions	The key questions were adequately developed and the summaries were consistent with study data, however, the conclusions presented are vague and are inadequate to support clinical decision-making. The lack of specificity may be attributable to the need for more research describing outcomes and opportunities for BMP usage.	The conclusions were based on analysis of the body of evidence for each use according to the AHRQ-modified GRADE convention. They reflect the quality and extent of published literature at the time the assessment was prepared.
		Results	We would like to note that packaging problems occurred during initial shipments of OP-1 and this quality control issue may have affected the efficacy of the product. Further, there are current concerns over the percentage content of BMP in comparable commercially prepared dosages. The technology assessment does not discuss variations in BMP dosages, which may generate bias in the literature.	Agreed. However, we are not aware of any controlled studies that were designed to investigate the effect of dose on clinical outcomes. We recorded doses used as available, but synthesis of this information is

				complicated by variability in study design and quality, patient characteristics, and actual use of BMP (i.e., with bone graft extenders).
		Cost-Effectiveness Analysis	We acknowledge the difficulty to assess cost since all applications may not be specifically coded as BMP. However, cost-effectiveness studies used in this TA may not have taken into consideration the costs of rehabilitation, amputation, repeated surgery and prosthetic fittings. Evaluations of alternative therapies demand such factors should be considered to provide a balanced assessment of the options.	We were limited by the available data sources on the occurrence of secondary interventions. No data sources addressed the occurrence of rehabilitation, amputation and prosthetic fittings. We chose to model outcomes for which we had evidence.
Kemner, Jason, Medtronic, Inc.	Introduction, Background, Methods	No comment	No response	
	Results	<ul style="list-style-type: none"> • Clarifications on BMP formulations, dose, and FDA status were provided. • Similarly, we would like to note a discrepancy regarding the 	<ul style="list-style-type: none"> • Clarifications were noted and text was revised to reflect this input. • Text was revised to reflect this comment. 	

		<p>notation used for Stryker's OP-1 formulations. It is noted in several locations that this product is rhBMP7/ACS (for example, page App1-127). This product uses a different carrier from that used in INFUSE Bone Graft. It is a granular collagen carrier that is derived from bovine bone as compared to the ACS, which is derived from bovine tendon and is a contiguous sheet.</p> <ul style="list-style-type: none"> Also, in Table 36, the first column is mislabeled for Jones et al. (Ref # 90) and Boraiah et al. (Ref # 108). These should be 		
		<ul style="list-style-type: none"> Table was revised to address this comment. 		

		<p>labeled as BMP2 Studies.</p> <ul style="list-style-type: none"> On page 29, and in other areas of the report ?Reference 73? (Dawson 2009) is categorized as an on-label application of rhBMP-2. This is an important piece of evidence and should be included in the assessment. However, this particular study evaluated rhBMP-2 in an application that has not been approved by the FDA and should be included in the off-label category. The review of clinical literature in the report does not include the long-term follow up data of those included 	<ul style="list-style-type: none"> Text, tables, and conclusions were revised to reflect this discrepancy. This did not alter conclusions. We became aware of this paper after the draft was prepared. Upon examination, we determined its results do not change the assessment conclusions but do footnote it in the Results chapter. 	
--	--	--	--	--

	<p>in the ALIF IDE trial. The following citation provides important data regarding the long-term results of those treated with INFUSE Bone Graft.</p> <p>(Burkus JK, Gornet MF, Schuler TC, Kleeman TJ, Zdeblick TA, Six-Year Outcomes of Anterior Lumbar Interbody Arthrodesis with Use of Interbody Fusion Cages and Recombinant Human Bone Morphogenetic Protein-2. J Bone Joint Surg Am., 91:1181-1189, 2009.)</p> <ul style="list-style-type: none"> On pages 16 and 17, the report identifies an INFUSE Bone Graft MasterGraft 2008 HDE device approval for symptomatic, posterolateral lumbar spine pseudoarthrosis 	<ul style="list-style-type: none"> This comment was addressed in revised text and tables. 	
--	---	--	--

	among patients for whom autologous bone and/or bone marrow harvest are not feasible or are not expected to promote fusion, such as diabetics and smokers. This HDE approval was voluntarily withdrawn by Medtronic in early 2010. This action was not the result of any quality or safety concerns identified by Medtronic or the Agency. Please update the assessment regarding the voluntary HDE approval withdrawal.		
Cost-Effectiveness Analysis	The base case cost-effectiveness analyses, which are conducted from the perspective of Medicare, are the	The opening paragraph of the executive summary section on the cost-effectiveness analyses has been revised as follows: When base case analyses assume identical initial	

	<p>primary analyses. As reported in Tables 50 and 53, the base case cost-effectiveness analyses of spinal fusion and open tibial repair find BMP to be the dominant strategy compared to the standard of care, thus yielding lower costs and higher quality-adjusted life years. The results of the base case analyses are not included in the executive summary table page 9. While a discussion of the sensitivity analyses may not be inappropriate in the executive summary, the primary focus should be the results of the base case analyses. The base case is consistent with the necessary assumptions of a Medicare perspective cost-effectiveness analysis which is that spine fusion cases performed with or without BMP are assigned to the same DRGs and thus generally receive the identical payment</p>	<p>hospitalization costs within the Medicare diagnosis-related group payment system, use of rhBMP-2 dominates the alternative strategy for both open tibial fracture and spinal fusion. In sensitivity analyses, the incremental cost-effectiveness ratios (ICERs) for both open tibial fracture and spinal fusion are highly influenced by the assumed added cost of rhBMP2</p> <p>Evidence was lacking on whether secondary interventions were performed in outpatient or inpatient settings. We decided to assume secondary interventions were performed as outpatient procedures as a conservative approach.</p>	
--	--	--	--

		<p>amount. The same is true for tibial repair cases performed with or without BMP. In the general context of the Medicare payment system, the dominant findings from the base-case cost-effectiveness analyses should be noted in the executive summary.</p> <p>Additionally, in Table 44 and thus within the CEA, invasive secondary interventions of bone graft, exchange nail or plate fixation may more likely be inpatient encounters with costs reflective of a DRG payment. This may better reflect clinical practice and associated costs and could influence results of the base case as well as sensitivity analyses.</p>		
	Discussion/C onclusion, Tables, Figures, Appendices, References	No comment	No response	

Rutka, James, MD, PhD	AANS	Results	<ul style="list-style-type: none"> The assessment qualified the FDA HDE approval for rhBMP7 as follows: “the use of OP-1 Putty will not expose patients to an unreasonable or significant risk of illness or injury and the probable benefit to health from using the device outweighs the risk of illness or injury”. Major issues dealing with the use of BMP as an adjunct to spinal fusion, however, remain unaddressed by this assessment and the current literature. Identified risk factors for failed fusion surgery include: Cigarette smoking, diabetes, osteoporosis, dialysis dependent renal disease, etc. Individuals with these characteristics are typically excluded from the majority of clinical trials because of their propensity to develop a non-union. Nonetheless, these patients, often because of these risk factors, require spinal fusion surgery due to disabling symptoms. The potential for BMP to enhance fusion rates, as demonstrated in many studies and reported in this assessment, may prove to be a significant clinical benefit to these patients and likely result in a reduced need for revision surgeries. Also not addressed in this assessment are patients who have had bone graft harvested previously and therefore have limited availability of autograft bone. Under these circumstances, allograft bone offers insufficient fusion potential and the compassionate use of BMP is appropriate. Another group not discussed in this review are 	<ul style="list-style-type: none"> The language was taken from the FDA Approval Summary We agree with Dr. Rutka, but identified no study that specifically addressed these patient factors. Again, we agree with Dr. Rutka. BMP would seem to provide a good alternative for patients with these
--------------------------	------	---------	---	---

			<p>patients who for religious or cultural reasons or for concerns over the risk of transmission of infectious agents refuse cadaveric allograft yet still have a need for bone graft during surgery. Unfortunately, many of these clinical situations arise with such a low frequency that generating valid medical evidence may prove difficult if not impossible.</p>	<p>characteristics, but studies addressing these issues were not identified.</p>
--	--	--	---	--

¹ Names are alphabetized by last name. Those who did not disclose name are labeled "Anonymous Reviewer 1," "Anonymous Reviewer 2," etc.

² Affiliation is labeled "NA" for those who did not disclose affiliation.

³ If listed, page number, line number, or section refers to the draft report.

⁴ If listed, page number, line number, or section refers to the final report.