

RECEIVED  
2024 April 30 PM 2:13  
NAFTA Secretariat  
U.S. Section



UNITED STATES DEPARTMENT OF COMMERCE  
Office of the General Counsel  
OFFICE OF THE CHIEF COUNSEL FOR TRADE ENFORCEMENT & COMPLIANCE  
Washington, D.C. 20230

April 30, 2024

**VIA ELECTRONIC MAIL**  
**(tradeagreementssecretariat@trade.gov)**

Ms. Vidya Desai  
Secretary, United States Section  
NAFTA Binational Secretariat  
Suite 2061  
14<sup>th</sup> Street and Constitution Avenue, NW  
Washington, DC 20230

RE: Softwood Lumber from Canada: Final Affirmative Determination of Sales at Less Than Fair Value and Affirmative Final Determination of Critical Circumstances  
Secretariat File No. USA-CDA-2017-1904-03  
Final Results of Redetermination Pursuant to Panel Remand Order

Dear Ms. Desai:

Pursuant to the Panel's remand order of October 5, 2023, and its extension orders of December 22, 2023, February 28, 2023, and April 2, 2024, please find attached the Final Results of Redetermination in the above-captioned matter. The remand redetermination is a business proprietary document. We are also filing a public version of the remand redetermination.

If you have any questions regarding this matter, please do not hesitate to contact the undersigned at (202) 597-1011 or [jason.miller@trade.gov](mailto:jason.miller@trade.gov).

Respectfully submitted,

/s/ Jason Miller

Jason Miller  
Attorney  
Office of the Chief Counsel  
for Trade Enforcement and Compliance

Attachments

**Certain Softwood Lumber Products from Canada:  
Final Affirmative Determination of Sales at Less than Fair Value  
Secretariat File No. USA-CDA-2017-1904-03**

**FINAL RESULTS OF REDETERMINATION  
PURSUANT TO BINATIONAL PANEL ORDER**

**I. SUMMARY**

Pursuant to the Binational Panel Order<sup>1</sup> concerning the U.S. Department of Commerce’s (Commerce) final affirmative determination of sales at less-than-fair-value (LTFV) of certain softwood lumber products (lumber) from Canada,<sup>2</sup> we have prepared these final results of redetermination. In its decision, the Binational Panel (Panel) remanded, in part, the *Final Determination* to Commerce: (1) to reconsider the deduction of the 2006 SLA<sup>3</sup> export tax from gross U.S. prices in calculating a respondent’s export price or constructed export price; (2) for further explanation on Commerce’s analysis in determining whether to grant Resolute FP’s<sup>4</sup> request for a startup adjustment at its Atikokan Mill, particularly related to the second requirement for a startup adjustment, *i.e.*, that production levels were limited by technical factors

---

<sup>1</sup> See “Article 1904 Binational Panel Review, Pursuant to the North American Free Trade Agreement Panel Decision, Certain Softwood Lumber Products from Canada: Final Affirmative Determination of Sales at Less Than Fair Value, Decision and Order,” Secretariat File No. USA-CDA-2017-1904-03 (Binational Panel Order).

<sup>2</sup> See *Certain Softwood Lumber Products from Canada: Final Affirmative Determination of Sales at Less Than Fair Value and Affirmative Final Determination of Critical Circumstance*, 82 FR 51806 (November 8, 2017) (*Final Determination*), and accompanying Issues and Decision Memorandum (IDM).

<sup>3</sup> See “Softwood Lumber Agreement Between the Government of Canada and the Government of the United States of America,” Office of the United States Trade Representative, October 12, 2006 (2006 SLA).

<sup>4</sup> In the *Final Determination*, Commerce determined that Resolute FP Canada Inc. (Resolute FP), Resolute Growth Canada Inc., Abitibi-LP Engineered Wood Inc., Abitibi-LP Engineered Wood II Inc., Forest Products Mauricie LP, Produits Forestiers Petit-Paris Inc., and Société en commandite Scierie Opitciwan are a single entity (collectively, Resolute).

associated with the initial phase of commercial production; and (3) for explanation of “whether the limits {i.e., the three statistical criteria} on the use of the Cohen’s *d* test were met in this case, or whether those limits need not be followed by Commerce,”<sup>5</sup> and Commerce’s “choice for the Cohen’s *d* denominator, either simple averaging or an alternate choice {i.e., a weighted average or a single standard deviation}.”<sup>6</sup>

In accordance with the Binational Panel Order, in these final results of redetermination, we did not deduct the 2006 SLA export taxes from gross U.S. prices in the calculation of export price and constructed export price. Additionally, we further explain our reliance on log inputs to determine when the Atikokan Mill reached commercial production levels. Furthermore, we continue to use the Cohen’s *d* test as in the *Final Determination*, as we have further explained that the three statistical criteria need not be observed here, and that use of a simple average of the standard deviations of the test and comparison groups as the denominator of the Cohen’s *d* coefficient is reasonable. In accordance with these determinations, we have recalculated the estimated weighted-average dumping margin for each of the mandatory respondents and revised the estimated weighted-average dumping margin for all other producers and exporters.

## **II. BACKGROUND**

On October 5, 2023, the Panel issued its decision and order in which it remanded to Commerce, in part, the *Final Determination*.<sup>7</sup> Concerning the Panel’s findings related to Commerce’s Cohen’s *d* test, the Panel remanded this “issue to Commerce to address its consistency with the United States Court of Appeals for the Federal Circuit’s {(Federal Circuit)} determinations in *Stupp Corp. v. United States*, 5 F.4th 1341 (Fed. Cir. 2021) {(Stupp)}, *Mid*

---

<sup>5</sup> See Binational Panel Order at 30.

<sup>6</sup> *Id.* at 32.

<sup>7</sup> *Id.* at 1-3.

*Continent Steel & Wire Inc. v. United States*, 940 F.3d 662 (Fed. Cir. 2019) *{(Mid Continent I)}*, and *Mid Continent Steel & Wire Inc. v. United States*, 31 F.4th 1367 (Fed. Cir. 2022) *{(Mid Continent II)}*.”<sup>8</sup> Both *Stupp* and *Mid Continent II* reference certain academic literature not on the record of either the underlying investigation or this Binational Panel segment. Accordingly, on November 24, 2023, Commerce placed on the record the academic texts cited by the Federal Circuit in its opinions.<sup>9</sup> Commerce also provided interested parties an opportunity to comment further on these academic texts.<sup>10</sup>

On December 1, 2023, Resolute FP and the Government of Canada (GOC) submitted, in response to Commerce’s memorandum placing the academic texts from the Federal Circuit on the record, voluminous new factual information.<sup>11</sup> On December 21, 2023, Commerce rejected the Resolute FP and GOC 12/1 Submission because, while Resolute FP and the GOC provided some additional information that was on the record in the *Stupp*, *Mid Continent I*, and *Mid Continent II* ongoing litigation involving Commerce, they also included untimely new information given Commerce’s limited request for comments on the academic texts from the Federal Circuit:

Some of the factual information submitted by the Canadian parties is on the record of *Stupp*, *Mid Continent I*, and *Mid Continent II*. Accordingly, Commerce has decided to accept this information, as it is before the {Federal Circuit} in those cases. However, the Canadian parties also submitted additional factual information that is not on the record of those litigation proceedings. Therefore, given that the Canadian Parties submitted new factual information that was not requested by Commerce, and this factual information is not on the record of any of the underlying proceedings before the {Federal Circuit} and cited by the {Panel}, we are rejecting this information as untimely new factual information not previously contained on

---

<sup>8</sup> *Id.* at 2.

<sup>9</sup> See Memorandum, “Certain Softwood Lumber Products from Canada: Factual Information,” dated November 24, 2023 (Academic Texts from the Federal Circuit Memorandum).

<sup>10</sup> *Id.* at 2.

<sup>11</sup> See Resolute FP and GOC’s Letter, “Certain Softwood Lumber Products from Canada: Response to Commerce’s Factual Information,” dated December 1, 2023 (Resolute FP and GOC 12/1 Submission).

the record of this proceeding and not consistent with the {Panel's} remand determination.<sup>12</sup>

Commerce provided Resolute FP and the GOC an opportunity to resubmit their comments with the untimely new factual information not on the records of *Stupp, Mid Continent I*, or *Mid Continent II* redacted.<sup>13</sup>

On December 27, 2023, Resolute FP and the GOC submitted a redacted version of their December 1, 2023 submission, excluding the unsolicited new factual information as requested by Commerce.<sup>14</sup> Further, Resolute FP and the GOC requested that Commerce reconsider its rejection of the additional information which they had included in their December 1, 2023 submission.<sup>15</sup>

On February 20, 2024, Commerce released its Draft Redetermination to interested parties and provided them with an opportunity to comment on the Draft Redetermination.<sup>16</sup> On March 5, 2024, the Committee Overseeing Action for Lumber International Trade Investigations or Negotiations (COALITION, the petitioner in the LTFV investigation),<sup>17</sup> the Canadian Parties<sup>18</sup>

---

<sup>12</sup> See Memorandum “Certain Softwood Lumber Products from Canada: Rejection of Untimely Filed New Information,” dated December 21, 2023 (Rejection of December 1, 2023 Submission), at 2. On February 7, 2024, Commerce corrected its Rejection of December 21, 2023 Submission to reject the related subparts and public version of the Resolute FP and GOC December 1, 2023 Submission and not just the first part of the submission.

<sup>13</sup> *Id.*

<sup>14</sup> See Resolute FP and GOC’s Letter, “Certain Softwood Lumber from Canada: Resubmission of Response to Commerce’s Factual Information,” dated December 27, 2023 (Resolute FP and GOC 12/27 Submission).

<sup>15</sup> See Resolute FP and GOC’s Letter, “Certain Softwood Lumber from Canada: Request for Reconsideration,” dated December 27, 2023.

<sup>16</sup> See Draft Results of Redetermination Pursuant to Binational Panel Order, Certain Softwood Lumber Products from Canada: Final Affirmative Determination of Sales at Less than Fair Value, Secretariat File No. USA-CDA-2017-1904-03, dated February 20, 2024 (Draft Redetermination).

<sup>17</sup> See COALITION’s Letter, “Comments on Draft Results of Redetermination,” dated March 5, 2024 (COALITION Comments on the Draft Redetermination). The COALITION is an *ad hoc* association whose members include: U.S. Lumber Coalition, Inc.; Collum’s Lumber Products, L.L.C.; Fox Lumber Sales, Inc.; Hankins, Inc.; Pleasant River Lumber Company; PotlatchDeltic; Rex Lumber Company; S.I. Storey Lumber Co., Inc.; Stimson Lumber Company; Swanson Group; Weyerhaeuser Company; Carpenters Industrial Council; Giustina Land and Timber Company; and Sullivan Forestry Consultants, Inc.

<sup>18</sup> See Canadian Parties’ Letter, “Comments on the U.S. Department of Commerce’s Draft Redetermination Results,” dated March 5, 2024 (Canadian Parties Comments on the Draft Redetermination). The Canadian Parties include the GOC, the Government of Ontario, the Government of Québec, Canfor Corporation, Tolko Marketing and Sales Ltd. and Tolko Industries Ltd., and West Fraser Mills Ltd. (collectively, the Canadian Parties).

and Resolute FP<sup>19</sup> submitted comments on the Draft Redetermination. We have addressed the interested parties' comments below.

### III. REMANDED ISSUES

#### A. 2006 SLA Tax Deduction from Gross U.S. Price

In the *Final Determination*, Commerce deducted from gross U.S. price export taxes paid by the mandatory respondents to the Canadian government pursuant to the terms of the 2006 SLA on U.S. exports during a portion of the period of investigation (POI), October 1, 2015, through September 30, 2016.<sup>20</sup>

As an initial matter, in the Binational Panel Order, the Panel found that interested party arguments regarding the purpose and intent of the 2006 SLA on Canadian exports to the United States were properly raised before Commerce sufficiently to avoid the administrative law doctrine of exhaustion.<sup>21</sup> Regarding the deduction of the 2006 SLA export taxes, the Panel observed that under section 772(c)(2)(B) of the Tariff Act of 1930, as amended (the Act), Commerce is prohibited from deducting an export tax if it is “specifically intended to offset the countervailable subsidy received.”<sup>22</sup> The Panel disagreed that the Act supported Commerce’s requirement that in order to invoke section 772(c)(2)(B) of the Act, parties must present record evidence that an export tax has specifically offset certain countervailed subsidies.<sup>23</sup> The Panel held that section 772(c)(2)(B) of the Act prohibiting the deduction of an export tax “specifically intended to offset the countervailable subsidy received” is not the same as requiring that an export tax “specifically offset” a certain subsidy, and thus, the Panel concluded that the statutory

---

<sup>19</sup> See Resolute FP’s Letter, “Resolute’s Comments on Draft Results of Redetermination,” dated March 5, 2024 (Resolute FP Comments on the Draft Redetermination).

<sup>20</sup> See *Final Determination* IDM at Comment 20.

<sup>21</sup> See Binational Panel Order at 34-35.

<sup>22</sup> *Id.* at 36.

<sup>23</sup> *Id.* at 36-37.

language does not support a requirement that a respondent link an export tax to a specific subsidy or that the export tax actually offset a specific subsidy received.<sup>24</sup> Rather, the Panel held that the decision of whether to deduct the export tax must focus on the intent of the export tax.<sup>25</sup> Thus, the Panel found that Commerce erroneously removed the words “intended to” and thereby imposed a burden not found in the Act.<sup>26</sup> The Panel further found no support for the proposition that “specifically intended” must mean a singular intent. Thus, while the 2006 SLA may be intended to address events outside of the imposition of countervailing duties (CVD), any additional intent is not a justification for the deduction of an export tax.<sup>27</sup>

The Panel determined that Commerce should further explain its determination with respect to whether an export tax could be specifically intended to do more than one thing or that the 2006 SLA can simultaneously have more than one specific intent.<sup>28</sup> In so doing, the Panel held that Commerce should also consider what authority it has for the proposition that an export tax must completely offset the amount of a net subsidy in order to satisfy the statutory language of “specifically intended to offset a countervailing subsidy.”<sup>29</sup> The Panel also noted that parties’ numerous comments and claims regarding the purpose of the 2006 SLA only confirmed that the purpose of the 2006 SLA was to broadly resolve all pending lumber disputes between the United States and Canada.<sup>30</sup> The Panel concluded that it would need to consider further explanation before it could conclude that the export tax paid under the 2006 SLA was not intended to offset a specific subsidy, and thus, the Panel remanded to Commerce to reconsider the treatment of the export tax and specifically consider what authority Commerce has to read sections 772(c)(2)(B)

---

<sup>24</sup> *Id.* at 37.

<sup>25</sup> *Id.*

<sup>26</sup> *Id.*

<sup>27</sup> *Id.*

<sup>28</sup> *See* Binational Panel Order at 271

<sup>29</sup> *Id.*

<sup>30</sup> *Id.*

and 771(6) of the Act as requiring a demonstration of a specific offset of, rather than a specific intent to offset, countervailed subsidies.<sup>31</sup>

## **Analysis**

Based on our reexamination of the issue, we have determined not to deduct the 2006 SLA export tax from gross U.S. price. As we acknowledged in the *Final Determination*, under sections 772(c)(2)(B) and 771(6)(C) of the Act, export taxes, duties, or other charges levied on the export of merchandise to the United States, specifically intended to offset the countervailed subsidies, are not deductible from export price or constructed export price.<sup>32</sup> We further emphasized that the Act is clear that although U.S. price may be reduced by an export tax, it cannot be reduced by an export tax which is “levied on the export of merchandise to the United States specifically intended to offset the countervailable subsidy received.”<sup>33</sup> We noted that the reason for this exception to the allowance for export tax adjustments is clear – if the trade distorting effects of the countervailed subsidies on the U.S. price can be shown to be reduced by a “specific” amount as a result of the export tax, it would be illogical for Commerce to negate that reduction in a potential distortion by nonetheless lowering the U.S. price by deducting that particular export tax in its calculations.<sup>34</sup>

We specified in the *Final Determination* that to satisfy sections 772(c)(2)(B) and 771(6)(C) of the Act, a party must prove, by citing record evidence, that an export tax has specifically offset a countervailed subsidy received in order for no deduction to be applied, and we noted that no party had done so.<sup>35</sup> However, as concluded by the Panel, the intent of the 2006

---

<sup>31</sup> *Id.* at 38-39.

<sup>32</sup> See *Final Determination* IDM at Comment 20.

<sup>33</sup> *Id.*

<sup>34</sup> *Id.*

<sup>35</sup> *Id.*

SLA was to resolve all pending lumber disputes between the United States and Canada,<sup>36</sup> and these disputes included countervailed subsidies. Article III of the 2006 SLA specifies that it revoked the existing CVD order without the possibility of reinstatement, while Article V specifies that during the period that the 2006 SLA is in effect, Commerce will effectively be prohibited from imposing new CVDs on lumber exports to the United States by Canada.<sup>37</sup> Thus, regardless of whether the 2006 SLA specified its intent regarding the export tax, the Panel, supported by Articles III and V of the 2006 SLA, concluded that the 2006 SLA was intended to resolve, in part, Commerce’s previous imposition of CVDs. Section 772(c)(2)(B) of the Act prohibits deducting an export tax from U.S. price if it is “specifically intended to offset the countervailable subsidy received.” Thus, no further evidence is necessary to conclude that section 772(c)(2)(B) of the Act prohibits deducting the 2006 SLA export tax from U.S. price. Furthermore, neither section 772(c)(2)(B) of the Act nor the regulations specify that an export tax must solely be intended to offset a countervailed subsidy. Therefore, the 2006 SLA export tax that is meant to resolve, in part, the imposition of CVDs, meets the criteria of a prohibited deduction under the Act. Accordingly, for these final results of redetermination, we have not deducted the 2006 SLA export tax from the reported U.S. price.

### **Interested Parties’ Comments**

#### *COALITION’s Comments*<sup>38</sup>

According to the COALITION, the Panel instructed Commerce to reconsider whether an export tax “intended to do more than one thing” could satisfy the requirements of sections 772(c)(2)(B) and 771(6)(C) of the Act, which prohibit the deduction of export taxes “specifically

---

<sup>36</sup> See Binational Panel Order at 37.

<sup>37</sup> See 2006 SLA at Articles III and V.

<sup>38</sup> See COALITION Comments on the Draft Redetermination at 3-4.

intended to offset the countervailable subsidy received” from U.S. price.<sup>39</sup> However, while Commerce has determined that an export tax with more than one specific intent may satisfy the statute, Commerce has not adequately addressed the threshold question of whether export taxes imposed under the SLA were specifically intended to offset countervailable subsidies in the first place. Commerce appears to infer such intent from Articles III and V of the 2006 SLA; however, the Panel has already agreed that “the SLA does not expressly state its purpose,”<sup>40</sup> and Commerce has not explained how Articles III and V demonstrate any specific intent to offset countervailable subsidies received with respect to the imposition of export taxes under Article VII. Absent a demonstration of such specific intent, there is no lawful basis for finding that the 2006 SLA export taxes are prohibited deductions under the statute.

*Resolute FP’s Comments*<sup>41</sup>

Resolute FP argues that the Draft Redetermination correctly does not deduct 2006 SLA export taxes from the U.S. price in accordance with the law and the decision of the Panel.<sup>42</sup> The finding that such taxes were intended to offset countervailable subsidies is supported by substantial evidence.

**Commerce’s Position:**

The COALITION and the Panel are both correct that “the SLA does not expressly state its purpose.”<sup>43</sup> The 2006 SLA’s purpose was not to specify why the SLA export taxes and volume export restraints were imposed but rather to effectuate an agreement between the United States and Canada settling multiple disputes and ending both the existing antidumping duty (AD)

---

<sup>39</sup> *Id.* at 4 (citing Binational Panel Order at 37).

<sup>40</sup> *Id.*

<sup>41</sup> See Resolute FP Comments on the Draft Redetermination at 6.

<sup>42</sup> *Id.* (citing Draft Redetermination at 6-8).

<sup>43</sup> See COALITION Comments on the Draft Redetermination at 4 (citing Binational Panel Order at 37).

and CVD Canadian lumber proceedings. Thus, the intent of the SLA export taxes and volume export restraints must be inferred. We have reasonably inferred, based on the text of the Agreement, that the SLA export taxes were specifically intended to offset the countervailable subsidies that Commerce found to exist in the Canadian lumber industry.

Article III of the SLA (United States' commitment to revoke CVD and AD orders on lumber from Canada), read in conjunction with Article VII (Canada's commitment to apply export charges and volume export restraints on lumber exports to the United States), make clear that Canada's application of these export taxes was conditioned on the United States' revocation of its lumber CVD and AD orders, and thus, the elimination of countervailing and antidumping duties on lumber from Canada.<sup>44</sup> Consequently, the 2006 SLA export taxes, along with the described volume export restraints, replaced the offset of the CVDs on countervailable subsidies received. As these taxes replaced, in part, duties specifically intended to offset countervailable subsidies received, it can reasonably be inferred that these taxes were likewise specifically intended to offset, in part, the countervailable subsidies received. Moreover, the fact that these taxes and volume restraints might have also been intended to offset dumping (and to address other outstanding issues covered by the 2006 SLA) does not detract from the reasonableness of our inference that they were, in part, specifically intended to offset the countervailable subsidies received. As the panel stated, “{w}e see no authority for the proposition that ‘specifically intended’ must mean a singular intent.”<sup>45</sup>

The COALITION has accused Commerce of myopically focusing on Articles III and V of the 2006 SLA.<sup>46</sup> As explained, we reference Articles III and V because they identify the

---

<sup>44</sup> See 2006 SLA at Article III.

<sup>45</sup> See Binational Panel Order at 37.

<sup>46</sup> See COALITION Comments on the Draft Redetermination at 4.

United States' revocation of the AD and CVD orders on lumber from Canada, and a promise not to reintroduce these offsets through further trade remedy investigations, as conditions for Canada's application of the SLA export taxes. Thus, these Articles indicate that the 2006 SLA export taxes and volume export restraints were specifically intended, in part, to replace the offset of the CVD order.

The COALITION claims that, absent express language in the Agreement to this effect, there is no "legal basis" for Commerce's conclusion that the SLA export taxes were specifically intended to offset countervailable subsidies received.<sup>47</sup> However, the COALITION fails to cite any support for the proposition that sections 772(c)(2)(B) and 771(6)(C) of the Act impose an express statement requirement. The 2006 SLA specified that one of the conditions for Canada's imposition of the export taxes and voluntary export restraints was the resolution of the United States' CVD order through its revocation, the refunding of CVD cash deposits, and a commitment not to collect any new CVDs as an offset on lumber exports to the United States from Canada.<sup>48</sup> Because these taxes (along with the voluntary export restraints) resolved and replaced, in part, the CVD order, Commerce's determination that they were specifically intended to offset countervailable subsidies received is reasonable and lawful.

The COALITION further notes that the SLA export tax rates are based on U.S. lumber prices, provincial export volume, and the company exporters themselves,<sup>49</sup> which have no correlation to countervailability determinations or subsidy calculations. That the setting of the offsets does not relate to how Commerce measures countervailable subsidies is not evidence that contradicts that one of the intents of the SLA export tax was to offset Canada's countervailable

---

<sup>47</sup> See COALITION Comments on the Draft Redetermination at 11.

<sup>48</sup> See 2006 SLA at Articles III, IV, and V.

<sup>49</sup> See COALITION Comments on the Draft Redetermination (citing 2006 SLA at Articles VII:2, VII:6, VII:7, and VIII:1).

subsidies to its lumber producers. As we state and demonstrate above, a required condition for the imposition of the SLA export taxes was the ending of measures countervailing Canadian lumber subsidies. Clearly, one of the effects and intents of the 2006 SLA was to replace CVDs with export taxes and voluntary export restraints.

In light of the above, we continue to find that the SLA export taxes are “specifically intended to offset the countervailable subsidy received.” Consequently, consistent with sections 772(c)(2)(B) and 771(6)(C) of the Act, Commerce cannot, and has not, deducted these SLA export taxes from U.S. price.

### **B. Resolute FP Start-Up Costs**

In the *Final Determination*, Commerce denied Resolute FP’s request for a startup adjustment at its Atikokan Mill, stating it had not met the second of the two statutory criteria.<sup>50</sup> For a startup adjustment, section 773(f)(1)(C) of the Act requires that: (1) a producer is using a new production facility or producing a new product that requires substantial additional investment; and (2) production levels are limited by technical factors associated with the initial phase of commercial production. The SAA directs Commerce to measure production based on the “units processed” when determining whether commercial production levels have been reached.<sup>51</sup> While Commerce agreed that the Atikokan Mill was a new production facility, Commerce determined, based largely on an analysis of the monthly log inputs (*i.e.*, production starts), that the mill had emerged from the startup phase prior to the POI.<sup>52</sup> Accordingly, Commerce denied the startup adjustment for the Atikokan Mill.

---

<sup>50</sup> See *Final Determination* IDM at Comment 45.

<sup>51</sup> See Statement of Administrative Action Accompanying the Uruguay Round Agreements Act, H.R. Doc. 103-316, Vol. 1 (1994) (SAA), at 836.

<sup>52</sup> See *Final Determination* IDM at Comment 45.

The Panel held that Commerce must explain why measuring log inputs satisfies the language of the SAA. The Panel noted that while the statute does not define the phrase “production levels,” the SAA directs Commerce to consider “the actual production experience of the merchandise in question,” and states that “{p}roduction levels will be measured based on units processed.”<sup>53</sup> Although Commerce has the discretion to interpret and apply the statute, Commerce must provide a reasonable explanation for its position.<sup>54</sup> In this regard, the Panel found that Commerce failed to adequately explain why “log inputs” should be equated with “units processed” or “production levels” as required by the SAA and failed to provide any persuasive or binding legal precedent for this approach.<sup>55</sup>

The Panel observed that there are many steps in the processing of lumber, however, Commerce considered only one static factor, log inputs. For example, the Panel noted that it is possible more logs were needed at the start of the operation to fine tune the process or that many logs were not fully processed, or that the lumber products were not commercially marketable due to some defect in the production process.<sup>56</sup> In fact, Commerce acknowledges that Resolute FP had issues with a planer that produced several broken, unmarketable products. Therefore, Commerce must explain why such technical issues were not sufficient to limit production.<sup>57</sup>

Next, the Panel found that the SAA’s use of the past tense phrase “units processed” suggests that the processing should be complete, yet Commerce appears to treat the phrase as “units *to be* processed.” According to the Panel, the statute directs Commerce to measure the quantity of the subject merchandise produced or “commercial production levels,” which may not

---

<sup>53</sup> See Binational Panel Order at 47.

<sup>54</sup> *Id.*

<sup>55</sup> *Id.* at 47-48.

<sup>56</sup> See Binational Panel Order at 48.

<sup>57</sup> *Id.*

be the same as raw material inputs.<sup>58</sup> Thus, the Panel concludes that if Commerce is unable to measure the amount of the subject merchandise processed as directed by the statute, Commerce must explain why that is the case and why using log inputs is the best substitute.<sup>59</sup>

Finally, the Panel dismissed Commerce's citation to *Agro Dutch*,<sup>60</sup> which the Panel noted was a U.S. Court of International Trade (CIT) decision that was not binding on the Panel, as distinguishable from the present case and not sufficiently persuasive authority either.<sup>61</sup> First, the Panel explained that in *Agro Dutch*, Commerce interpreted the term "units processed" to be "how many units Agro Dutch set out to produce," and not the number of inputs, which was the measure used in the present case. Second, the Panel noted that the producer in *Agro Dutch* provided insufficient evidence to support its claim that technical factors limited production (providing only total output and yield rates, but not the inputs that would allow Commerce to evaluate the evidence). Third, the Panel pointed out that *Agro Dutch* contended that Commerce should have considered the output yields between the new and preexisting growing rooms. According to the Panel, this argument is the most similar to the current case but is still inapposite because *Agro Dutch* relied on improvements in efficiency in making its argument, which the SAA indicates are not to be considered. In the current case, Resolute FP is not comparing new, more efficient equipment to older equipment; rather, Commerce compared the production starts at all Resolute FP mills.<sup>62</sup> Thus, the Panel concluded that although Commerce has previously stated that production starts are the best measure of a facility's capacity to produce at commercial production levels, Commerce has not yet adequately explained in prior cases or here

---

<sup>58</sup> *Id.*

<sup>59</sup> *Id.*

<sup>60</sup> See *Agro Dutch Foods Ltd., v. United States*, 110 F. Supp. 2d 950, 956 (CIT 2000) (*Agro Dutch*).

<sup>61</sup> See Binational Panel Order at 49-50.

<sup>62</sup> *Id.*

why it believes that to be true or how this conclusion is consistent with the language of the statute and the SAA.<sup>63</sup>

## Analysis

We continue to find that an analysis of Resolute FP's log inputs (production starts) at the Atikokan Mill was the best measure of determining when the mill achieved commercial production levels. However, in accordance with the Panel's remand request, we have provided additional explanation and case precedent to remedy the deficiencies in our support of this approach. These citations demonstrate that Commerce has a long-standing practice of relying on production inputs or starts as the key indicator of when a company has exited the startup phase of production. *diffe*

First, we respectfully disagree with some of the Panel's interpretation of *Agro Dutch*. The Panel stated that “*i*n *Agro Dutch*, Commerce explained that it interpreted the term ‘units processed’ to be ‘how many units *Agro Dutch* set out to produce’” and the Panel explained that “*a*s an initial matter, the number of units a producer intends to produce is different from the number of inputs, which was the measure Commerce used in the present case.”<sup>64</sup> However, Commerce's interpretation of the term “units processed” in *Agro Dutch*, in full, was as follows:

At oral argument, the Government clarified the term “units processed.” Counsel explained that Commerce meant information on how many units *Agro Dutch* set out to produce. In other words, *how much input was used* during the period of investigation (POI).<sup>65</sup>

Thus, in *Agro Dutch*, Commerce did equate the inputs into production to the “units processed.” Additionally, this explanation recognizes that a company's intended production output dictates the quantity of inputs required (similar to a recipe for baking a cake, if one wants

---

<sup>63</sup> *Id.* at 50.

<sup>64</sup> *Id.* at 49.

<sup>65</sup> See *Agro Dutch*, 110 F. Supp. 2d 950, 955 (emphasis added).

one cake, one needs x amount of flour, eggs, *etc.*). Thus, under Commerce’s interpretation of units processed, we consider a company’s inputs to be indicative of its intended production levels. Where a company significantly increases the inputs or production starts at a new facility, which in turn reflects its intended output, Commerce considers this an indication of the company’s confidence in the new equipment and confidence that significant technical issues have been resolved and will no longer hamper production levels.

Even putting aside *Agro Dutch*, there are still numerous Commerce determinations that support its long-standing and consistent interpretation of “units processed” as production starts or inputs. For example, in *SRAMs from Taiwan*, Commerce stated that “the SAA at 836 (166) directs us to examine the units processed in determining the claimed startup period.

Accordingly, our determination of the startup period was based, in large part, on a review of the wafer *starts* at the new facility during the POI, which represents the best measure of the facility’s ability to produce at commercial production levels.”<sup>66</sup> Furthermore, in *CTL Plate from Canada*, Commerce explained that “{i}n order to determine the duration of the initial phase of commercial production, we examined {the respondent’s} reported production starts at the {electric arc furnace}. Our determination of an appropriate startup period was based, in large part, on a review of scrap *starts* at the new facility during the {period of review (POR)}, which represents the best measure of the facility’s ability to produce at commercial production levels.”<sup>67</sup> In *Brass SS from the Netherlands*, Commerce remarked that “{t}he decision to significantly increase the number of caster *starts* is indicative of {the respondent’s} resolution of

---

<sup>66</sup> See Notice of Final Determination of Sales at Less Than Fair Value: *Static Random Access Memory Semiconductors from Taiwan*, 63 FR 8909, 8930 (February 23, 1998) (*SRAMs from Taiwan*) (emphasis added).

<sup>67</sup> See *Certain Corrosion-Resistant Carbon Steel Flat Products and Certain Cut-to-Length Carbon Steel Plate from Canada: Preliminary Results of Antidumping Duty Administrative Reviews and Intent To Revoke in-Part*, 63 FR 37320, 37324 (July 10, 1998), unchanged in *Certain Corrosion-Resistant Carbon Steel Flat Products and Certain Cut-to-Length Carbon Steel Plate From Canada: Final Results of Antidumping Duty Administrative Reviews and Determination To Revoke in Part*, 64 FR 2173 (January 13, 1999) (*CTL Plate from Canada*) (emphasis added).

technical problems that had initially restricted production.”<sup>68</sup> Commerce likewise relied on production starts or inputs in *Chlorinated Isos from Spain* and *Concrete Rebar from Türkiye* to determine when commercial production levels had been reached.<sup>69</sup> In *Concrete Rebar from Türkiye*, Commerce opined that the level of production starts “clearly exhibited confidence in the new production facilities. Had {the respondent} had not been confident in its plant’s ability to produce, we would have expected to see a significantly lower number of production starts.”<sup>70</sup>

Indeed, in a prior Canadian lumber proceeding, Commerce explicitly defined “processed” lumber as lumber “inputs.”<sup>71</sup> Commerce stated that, “{i}n evaluating {the respondent’s} start-up adjustment, we focused on when the new planing and MSR facility achieved commercial production levels. We reviewed the volume of lumber processed (*i.e.*, inputs) at the planing and MSR facility from July 2000 through July 2001 and the volume of lumber produced (*i.e.*, outputs) at the planing and MSR facility from July 2000 through November 2001.”<sup>72</sup> Thus, in that Canadian lumber case, Commerce concluded that,

{c}onsistent with the SAA and {Commerce’s} practice, we continue to apply production starts as the best measure of a facility’s capability to produce at commercial production levels. From analyzing {the respondent’s} data, we noted that the volume of lumber input into the planing facility significantly changed in the fourth quarter of 2000 . . . . This change in volume, consistent with the SAA, indicated that the planing process reached commercial production levels at the end of the third quarter 2000.<sup>73</sup>

---

<sup>68</sup> See *Notice of Final Results of Antidumping Duty Administrative Review and Determination Not to Revoke the Antidumping Duty Order: Brass Sheet and Strip from the Netherlands*, 65 FR 742, 743 (January 6, 2000) (*Brass SS from the Netherlands*) (emphasis added).

<sup>69</sup> See *Chlorinated Isocyanurates from Spain: Notice of Final Determination of Sales at Less Than Fair Value*, 70 FR 24506 (May 10, 2005) (*Chlorinated Isos from Spain*), and accompanying IDM at Comment 9; and *Certain Steel Concrete Reinforcing Bars from Turkey; Final Results, Rescission of Antidumping Duty Administrative Review in Part, and Determination To Revoke in Part*, 70 FR 67665 (November 8, 2005) (*Concrete Rebar from Türkiye*), and accompanying IDM at Comment 24.

<sup>70</sup> See *Concrete Rebar from Türkiye* IDM at Comment 24.

<sup>71</sup> See *Notice of Final Determination of Sales at Less Than Fair Value: Softwood Lumber from Canada*, 67 FR 15539 (April 2, 2002) (*Lumber from Canada*), and accompanying IDM at Comment 32.

<sup>72</sup> *Id.*

<sup>73</sup> *Id.*

We note that the planing process is a later stage in lumber production; thus, the input at this stage is lumber and not logs, but the focus remains the same—the inputs into production were most probative in Commerce’s decision as to when commercial production levels were reached.

The Panel also stated that the number of inputs is only one static factor in a process, and concluded that focusing on the number of inputs fails to account for the entire series of actions taken to produce the resulting product.<sup>74</sup> As such, the Panel suggested that more logs may have been needed at the start of the operation to fine tune the manufacturing process and that many logs were not fully processed or that the resulting lumber products produced in the early stages were not commercially marketable.<sup>75</sup>

However, in *DRAMs from Taiwan*, Commerce explained that “units going into finished goods inventory are not a good measure of the achievement of commercial levels of production, given that they are more a reflection of the quality of the product produced and the yields achieved in the production process.”<sup>76</sup> In fact, Commerce addressed the very issue of startup adjustments for commercially unviable production in comments to the final antidumping regulations. One commenter argued that where merchandise, although in production, is not yet of a quality sufficient for sale, some startup adjustment would be appropriate. However, Commerce declined to make product quality a criterion in determining the length of the startup period.<sup>77</sup>

we believe that this suggestion is inconsistent with the statute and the SAA ... . The SAA at 836 states that in making a determination as to when a producer reaches commercial production levels, {Commerce} will measure the producer’s actual production levels based on the number of units processed ... the SAA does not refer

---

<sup>74</sup> See Binational Panel Order at 48.

<sup>75</sup> *Id.*

<sup>76</sup> See *Notice of Final Determination of Sales at Less than Fair Value: Dynamic Random Access Memory Semiconductors of One Megabit and Above (“DRAMs”) from Taiwan*, 64 FR 56308, 56321 (October 19, 1999) (*DRAMs from Taiwan*).

<sup>77</sup> See *Antidumping Duties; Countervailing Duties*, 62 FR 27296, 27364 (May 19, 1997) (*Preamble*).

to quality of merchandise as a criterion for measuring the length of the startup period, but instead relies strictly on the number of units processed as a primary indicator of the end of the startup period. In fact, the SAA at 836 states that {Commerce} will not extend the startup period in a manner that would cover product improvements and cost reductions that may occur over the life cycle of the product. {Commerce} believes this to be a clear reference to product quality and yield improvements that may continue to exist long after startup has ended.

Accordingly, Commerce has consistently interpreted the SAA to mean that it is a company's increase in production starts (*i.e.*, the number of inputs that are processed) that indicates whether the company has reached commercial production levels. The level of investment that a company will dedicate (*i.e.*, greater raw material costs) to a new production process demonstrates that a company is confident enough in the new equipment to put more of its resources into the process. However, an examination of output, as the Panel suggests, would mistakenly rely on the quality of the merchandise produced, or yield of the output product, factors that the SAA specifically states should not be used to extend the startup period.

Further, Commerce examines a variety of products with varying degrees of quality differences in its investigations, and a reliance on the increase in inputs provides a more uniform and predictable proxy for whether commercial production levels have been reached than output. Thus, both in the *Final Determination* and on remand, Commerce has evaluated the Atikokan Mill's production levels by measuring the number of logs that were input, and therefore, "processed" at the Atikokan facility.

Moreover, had the intention been for Commerce to look at subject merchandise, the SAA could have used the terms "units produced" instead of "units processed."<sup>78</sup> Even so, if we were to consider the production yields that the Atikokan Mill achieved versus the other Resolute FP mills, we find they also do not support a conclusion that the Atikokan Mill was in a startup phase

---

<sup>78</sup> See SAA at 836.

during the POI. During the POI, the Atikokan Mill posted saleable production yields that were comparable to the other Resolute FP mills, *i.e.*, [ ] percent saleable production from the log inputs for the Atikokan Mill versus a range of [ ] percent to [ ] percent for the mills not claiming a startup adjustment.<sup>79</sup>

We also respectfully disagree with the Panel’s assertion that Commerce’s evaluation of log inputs reflects a review of the “units to be processed” rather than the “units processed.”<sup>80</sup> The monthly log inputs were not logs sitting available in inventory, but rather the logs that were actually processed in the mill during the month.<sup>81</sup> The same exhibit showing the monthly log inputs (the line item on the worksheet is actually described as “M3 consumed”), also shows the millions of board feet processed from the logs at the sawmill and the hours of processing time, demonstrating that the logs were processed, not “to be processed.”<sup>82</sup>

Hence, Commerce has a long-standing practice of relying on inputs or production starts when complying with the SAA’s directive to measure production levels based on “units processed.” We submit that this is a reasonable and consistent methodology for assessing whether a company’s new facilities have exited the startup phase of production. Where a company significantly ramps up production inputs or starts, it is demonstrating through its commitment of significant resources (*i.e.*, the raw materials input into production) that it believes the new facility is no longer hampered by the technical difficulties that are indicative of a startup phase. Consequently, on remand, we have continued to deny Resolute FP’s startup adjustment based largely on our analysis of its log inputs (*i.e.*, log consumption).

---

<sup>79</sup> See Resolute’s Letter, “Response of Resolute FP Canada Inc. to Section D Questionnaire,” dated March 20, 2017 (Resolute DIQR), at Exhibit D-18.

<sup>80</sup> See Binational Panel Order at 48.

<sup>81</sup> See Resolute’s Letter, “Softwood Lumber from Canada: Resolute’s Response to Supplemental Section D Questionnaire,” dated May 12, 2017 (Resolute SDQR), at Exhibit SD-30.

<sup>82</sup> *Id.*

## Interested Parties' Comments

### *COALITION's Comments*<sup>83</sup>

The COALITION states that Commerce's explanation for its interpretation of section 773(f)(1)(C)(ii)(II) of the Act and its resulting reliance on log inputs for determining whether Resolute FP's Atikokan Mill reached commercial production levels during the POI is reasonable.<sup>84</sup> The Draft Redetermination address the Panel's order that the agency further explain why an analysis of log inputs fulfills the statutory requirement that Commerce consider whether "production levels were limited by technical factors associated with the initial phase of commercial production."<sup>85</sup> Because Commerce complied with the Panel's remand order and explained why its analysis of log inputs is reasonable and in accordance with law, Commerce should continue to determine that Resolute FP has not established that it is entitled to a startup adjustment for the Atikokan Mill in the final results of redetermination.

### *Resolute FP's Comments*<sup>86</sup>

According to Resolute FP, the Panel remanded Commerce's denial of a start-up adjustment for Resolute FP's new Atikokan sawmill and planer on the grounds that it had already reached commercial levels of production by the POI.<sup>87</sup> Commerce's finding in the Draft Redetermination is based, as before, solely on the number of inputs (*i.e.*, logs) going into the production process.<sup>88</sup>

---

<sup>83</sup> See COALITION Comments on the Draft Redetermination at 3.

<sup>84</sup> *Id.* at 3 (citing Draft Redetermination at 10-16).

<sup>85</sup> *Id.* (citing section 773(f)(1)(C)(ii)(II) of the Act; and Binational Panel Order at 47-48 ("Commerce has not adequately explained why 'log inputs' should be equated with 'units processed' or 'production levels' as required by the SAA nor has Commerce provided any persuasive or binding legal precedent for the approach it took in this case.")).

<sup>86</sup> See Resolute FP Comments on the Draft Redetermination at 2-3.

<sup>87</sup> *Id.* at 2 (citing Binational Panel Order at 41-50).

<sup>88</sup> *Id.* (citing Draft Redetermination at 16).

The SAA, the authoritative expression of congressional intent on the dumping law, does not limit Commerce to looking at production inputs when considering commercial levels of production. Instead, the inquiry for “commercial production” is to be done on a “case-by-case basis,” taking into account the nature of the industry among “other factors.”<sup>89</sup> Here, the record demonstrates that Atikokan at the time of the investigation was a mill on the uneven path to commercial production.<sup>90</sup> Equipment failures, for example, hampered production of lumber until the end of the POI.<sup>91</sup> Capacity utilization also lagged behind other mills.<sup>92</sup> Commerce misconstrues the SAA’s guidance to examine “units processed,” which the Panel correctly interpreted as not the equivalent of inputs or production starts *per se*.<sup>93</sup>

Commerce attempts to support its findings based on a flawed interpretation of *Agro Dutch*,<sup>94</sup> even though the Panel found that the case was irrelevant based on factual differences with this proceeding.<sup>95</sup> Commerce’s citations to its own administrative decisions hold no persuasive value given underlying factual differences with the present case and the clear errors in its statutory interpretation concerning commercial levels of production. Commerce should correct the final remand redetermination to find that Atikokan had not reached commercial levels of production and that, therefore, a startup adjustment is warranted.

---

<sup>89</sup> *Id.* (citing SAA at 4040, 4173-74; and Binational Panel Order at 47).

<sup>90</sup> *Id.* (citing, *e.g.*, Resolute’s Letter, “Resolute Canada’s Response to Initial Section A Questionnaire,” dated March 1, 2017 (Resolute AIQR), at Exhibit A-19 (Pt. 1 at 5); Resolute DIQR at Exhibit D-13 at 5; Resolute SDQR at Exhibit SD-2; and Memorandum, “Verification of the Cost Response of Resolute FP Canada Inc.,” dated July 27, 2017 (Resolute Cost Verification Report), at 32).

<sup>91</sup> *Id.* (citing Resolute Cost Verification Report at 32).

<sup>92</sup> *Id.* (citing Resolute AIQR at Exhibit A-19 (Pt. 1 at 5); Resolute DIQR at Exhibit D-13 (page 5); and Resolute SDQR at Exhibit SD-2).

<sup>93</sup> *Id.* (citing Binational Panel Order at 48-50).

<sup>94</sup> *Id.* (citing *Agro Dutch*, 110 F. Supp. 2d 950, 955-57).

<sup>95</sup> *Id.* (citing Binational Panel Order at 49).

### **Commerce's Position:**

We continue to disagree that a startup adjustment is appropriate in this case. As an initial matter, we note that Resolute FP did not defer costs associated with the startup of its Atikokan facility in its normal books and records.<sup>96</sup> Rather, in its generally accepted accounting principles (GAAP) based accounting records, the company recognized the cost of manufacturing lumber at its Atikokan facilities as current expenses (*i.e.*, the startup costs were not capitalized and allocated to a post-startup period).

Section 773(f)(1)(A) of the Act establishes that costs are to be calculated based on a company's normal books and records where they are in accordance with home country GAAP and reasonably reflect the costs associated with production and sale of the merchandise. In fact, as Resolute FP acknowledges, generally accepted accounting principles do not afford special treatment to producers engaged in startup operations.<sup>97</sup> Regardless of any inefficiencies and excessive costs resulting from a new production process or equipment being used for the first time, GAAP does not allow the capitalization of production costs after the first sellable product has been produced (also defined as the beginning of commercial production according to GAAP).<sup>98</sup> However, to mitigate the possibility of creating dumping margins due to startup operations, U.S. law provides an exception, at section 773(f)(1)(C)(ii) of the Act, to the requirement that reported costs reflect the producer's normal books and records. Even so, Commerce rigorously scrutinizes respondent requests for startup adjustments and only affords

---

<sup>96</sup> See Resolute SDQR at D-36 to D-37 (explaining that Resolute FP did not capitalize any production costs once the first sellable product was produced at the Atikokan Mill).

<sup>97</sup> See, *e.g.*, Resolute SDQR at D-36 to D-37.

<sup>98</sup> *Id.* at D-36 to D-37.

companies a very narrow window in very limited situations to set aside the use of their GAAP-based records and submit lower production costs for reporting purposes.<sup>99</sup>

In this case, we found that Resolute FP met the first requirement for a startup adjustment, *i.e.*, operating a new production facility, but that the company had reached commercial production levels and exited the initial phase of production, or startup, prior to the POI.<sup>100</sup>

Resolute FP claims that Commerce has a flawed interpretation of the statute and misunderstands the meaning of commercial production levels. If so, this is an interpretation that Commerce has consistently applied since initially implementing this part of the Act.<sup>101</sup> To address this claim, we have first reviewed the specific language of the law, the SAA, and Commerce’s regulations regarding commercial production levels and the end of the startup period.

Section 773(f)(1)(C)(ii) of the Act states that where a respondent is operating a new facility or product, “adjustments shall be made ... where production levels are limited by technical factors associated with the initial phase of commercial production.” According to section 773(f)(1)(C) of the Act, the startup period ends “at the point at which the level of commercial production that is characteristic of the merchandise, producer, or industry concerned

---

<sup>99</sup> See, e.g., *Notice of Final Results of Sales At Less Than Fair Value: Collated Roofing Nails from Korea*, 62 FR 51420, 51426 (October 1, 1997) (where Commerce rejected the startup adjustment request for a relocation of the company’s manufacturing facilities as not being a new or substantially retooled facility); *Certain Cold-Rolled and Corrosion-Resistant Carbon Steel Flat Products from Korea: Final Results of Antidumping Duty Administrative Review*, 63 FR 13170, 13199 (March 18, 1998) (where Commerce rejected the startup adjustment request for a new production facility for a new product stating that the product was not “new” since it was comparable with other product lines); *DRAMs from Taiwan*, 64 FR 56308, 56321 (where Commerce determined that commercial production levels were reached prior to the POI based on the respondent’s increase in wafer starts during the pre-POI period); and *CTL Plate from Canada*, 63 FR 37320, 37324 (where Commerce concluded that while the first two months of production at a new electric arc furnace facility were not at commercial production levels, the chronic production problems experienced by the respondent did not constitute “technical factors” that are unique to a startup operation).

<sup>100</sup> See *Final Determination IDM* at Comment 45.

<sup>101</sup> See, e.g., *SRAMs from Taiwan*, 63 FR 8909, 8930; *CTL Plate from Canada*, 63 FR 37320, 37325; *Brass SS from the Netherlands*, 65 FR 742, 746; *Lumber from Canada IDM* at Comment 32; *Chlorinated Isos from Spain IDM* at Comment 9; and *Concrete Rebar from Türkiye IDM* at Comment 24.

is achieved.” Commerce’s regulations state that in determining when a producer reaches commercial production levels Commerce will “consider the actual production experience of the merchandise in question, measuring production on the basis of units processed” and “will examine factors in addition to those specified in section 773(f)(1)(C)(ii) of the Act, including historical data reflecting the same producer’s or other producers’ experiences in producing the same or similar products. A producer’s projections of future volume or cost will be accorded little weight.”<sup>102</sup>

Further, Commerce’s regulations state that “the attainment of peak production levels will not be the standard for identifying the end of the startup period, because the startup period may end well before a company achieves optimum capacity utilization” and “will not be extended to cover improvements and cost reduction that may occur over the entire life cycle of a product.”<sup>103</sup> The SAA adds that “the nature and timing of startup operations will vary from industry to industry and from product to product, and that any determination of the appropriate startup period involves a fact-intensive inquiry. In some industries, the startup period could be as short as one to two months; in others it could be much longer. For this reason, the Administration intends that Commerce determine the duration of the startup period on a case-by-case basis.”<sup>104</sup> Moreover, the SAA clarifies that “companies must demonstrate that, for the period under investigation or review, production levels were limited by technical factors associated with the *initial phase* of commercial production and not by factors unrelated to startup, such as marketing

---

<sup>102</sup> See 19 CFR 351.407(d)(3).

<sup>103</sup> See 19 CFR 351.407(d)(4).

<sup>104</sup> See SAA at 838. For example, Commerce found startup periods of two months for a new planer mill, two months for a new electric arc furnace, five months for a new strip caster, and nine months for a new chemical plant. See *Lumber from Canada* IDM at Comment 32; *CTL Plate from Canada*, 63 FR 37320, 37325; *Brass SS from the Netherlands*, 65 FR 742, 746; and *Chlorinated Isos from Spain* IDM at Comment 9, respectively.

difficulties or chronic production problems.”<sup>105</sup> Finally, in responding to comments on the final antidumping rule, Commerce provided the following clarification in the *Preamble*:<sup>106</sup>

{Commerce} has not adopted the suggestion to make product quality a criterion in determining the length of the startup period, because we believe that this suggestion is inconsistent with the statute and the SAA. Section 773(f)(f)(C)(ii) of the Act provides that {Commerce} will consider startup as having ended as of the time the producer achieves a level of commercial production that is characteristic of the merchandise, producer, or industry concerned. The SAA at 836 states that in making a determination as to when a producer reaches commercial production levels, {Commerce} will measure the producer’s actual production levels based on the number of units processed. The SAA also provides that, to the extent necessary, {Commerce} will examine other factors (such as historical data reflecting the same producer’s or other producer’s experience in producing the same or similar products) in determining the end of the startup period.

We note also that the SAA does not refer to quality of merchandise as a criterion for measuring the length of the startup period, but instead relies strictly on the number of units processed as a primary indicator of the end of the startup period. In fact, the SAA at 836 states that {Commerce} will not extend the startup period in a manner that would cover product improvements and cost reductions that may occur over the life cycle of a product. {Commerce} believes this to be a clear reference to product quality and yield improvements that may continue to exist long after startup has ended and, if taken into consideration, could result in extending the startup period beyond the point at which commercial production is achieved.

If the statute is ambiguous in its reference to the commercial production levels that signify the end of the startup period, Commerce clarifies the question when it states that the “SAA does not refer to quality of merchandise as a criterion for measuring the length of the startup period, but instead relies strictly on the number of units processed as a primary indicator of the end of the startup period.”<sup>107</sup> Thus, if Commerce is not considering production quality, *i.e.*, the end result of production, when measuring commercial production levels to determine the end of the startup period, then “units processed” must reference the quantities of raw materials that were started into production and processed. Based on the above guidance, Commerce

---

<sup>105</sup> See SAA at 838.

<sup>106</sup> See *Preamble*.

<sup>107</sup> *Id.*, 62 FR at 27364 (referencing SAA at 836).

implemented from the inception of the provision a startup analysis that is heavily reliant on a respondent's production starts and not on its finished production. In doing so, Commerce has consistently examined a company's monthly production starts in determining when commercial production levels were reached. Where Commerce noted a significant ramping up of the monthly production starts, Commerce concluded that the initial phase of production and the startup period were complete.

For example, *SRAMs from Taiwan* was one of the first cases where Commerce had to address the issue of a startup period under the new law. Commerce found that the respondent met the first criterion for a new product or facility and, in evaluating the startup period, Commerce stated "our determination of the startup period was based, in large part, on a review of the wafer starts at the new facility during the POI, which represents the best measure of the facility's ability to produce at commercial production levels."<sup>108</sup> Similarly, in *CTL Plate from Canada* Commerce relied "on a review of scrap starts at the new facility"<sup>109</sup> and in *DRAMs from Taiwan* Commerce stated that the "increase in wafer starts is indicative of ProMos's resolution of technical problems that had initially restricted production."<sup>110</sup> Commerce also reviewed the production inputs in *Brass SS from the Netherlands*, *Lumber from Canada*, *Chlorinated Isos from Spain*, and *Concrete Rebar from Türkiye*.<sup>111</sup> Thus, in analyzing respondent requests for a startup adjustment, Commerce has consistently analyzed production starts as a key predictor of when the startup period has ended. In implementing this analysis, Commerce relied on guidance from the statute, the SAA, and Commerce's final antidumping regulations. In particular, we find

---

<sup>108</sup> See *SRAMs from Taiwan*, 63 FR 8909, 8930.

<sup>109</sup> See *CTL Plate from Canada*, 63 FR 37320, 37325.

<sup>110</sup> See *DRAMs from Taiwan*, 64 FR 56308, 56322.

<sup>111</sup> See *Brass SS from the Netherlands*, 65 FR 742, 746; *Lumber from Canada* IDM at Comment 32; *Chlorinated Isos from Spain* IDM at Comment 9; and *Concrete Rebar from Türkiye* IDM at Comment 24.

that Commerce’s original interpretation of “units processed” is rooted in the following explanation: “{t}he SAA does not refer to quality of merchandise as a criterion for measuring the length of the startup period, but instead relies strictly on the number of units processed as a primary indicator of the end of the startup period.”<sup>112</sup> Thus, we posit that determining the end of Resolute FP’s startup period through an analysis of units processed (*i.e.*, log inputs that entered production and were processed) is consistent with statutory intent and prior practice.

Resolute FP also argues that the commercial level of production analysis is not limited to inputs and that the SAA instructs Commerce to consider “other factors.” According to Resolute, Commerce considered only logs inputs in its Draft Redetermination, however, “{a}n examination of logs in isolation tells nothing about what a mill should be consuming compared to budgeted projections, nor how well the equipment is functioning in producing the subject merchandise.”<sup>113</sup> Further, Resolute FP alleges that “Atikokan was in the midst of what can only be called an uneven transition to commercial levels of production – *i.e.*, ‘units processed’ – during the POI, even when the overall number of logs utilized increased or sometimes matched the number of consumed logs at other plants.”<sup>114</sup> Consequently, Resolute FP contends Commerce “must consider a number of other metrics to get a full view of the level of production at Atikokan during the startup period.”<sup>115</sup> As such, Resolute FP proffers that the Atikokan Mill lumber recovery factor (LRF), an expression of processing yields, fluctuated between [ ] and [ ] throughout the POI, and production was running at [ ] thousand board feet per hour (mbf/hr), which is behind the targeted speed of [ ] mbf/hr.<sup>116</sup> Resolute FP states that even using

---

<sup>112</sup> See *Preamble* (referencing SAA at 836).

<sup>113</sup> See Resolute FP Comments on the Draft Redetermination at 10.

<sup>114</sup> *Id.*

<sup>115</sup> *Id.* at 12.

<sup>116</sup> *Id.*

the data cited by Commerce, *i.e.*, saleable lumber yields, there was a [ ] between the Atikokan Mill and Resolute FP’s higher producing mills since the Atikokan Mill’s [ ] percent yield represented the [ ] yield.<sup>117</sup> Further, Resolute FP claims that over 2015-16, Atikokan stood around [ ] percent capacity while the average of Resolute FP’s mills was around [ ] percent.<sup>118</sup>

While we agree with Resolute FP that Commerce “will examine factors in addition to those specified in section 773(f)(1)(C)(ii) of the Act, including historical data reflecting the same producer’s or other producers’ experiences in producing the same or similar products,”<sup>119</sup> we disagree that we failed to consider any of the other factors listed in the statute. In fact, in our *Final Determination*, we compared the POI log inputs at 12 other Resolute FP mills to the those at the Atikokan Mill.<sup>120</sup> Further, in our Draft Redetermination, we also considered the production yields that the Atikokan Mill achieved versus the other Resolute FP mills.<sup>121</sup> Specifically, the Atikokan Mill posted POI saleable production yields that were comparable to the other Resolute FP mills, *i.e.*, [ ] percent saleable production from the log inputs for the Atikokan Mill versus a range of [ ] percent to [ ] percent for the mills not claiming a startup adjustment.<sup>122</sup>

Even so, although yields may indicate the existence of technical factors that limited production output, the SAA directs us to examine the units processed in determining the claimed startup period.<sup>123</sup> Furthermore, in responding to comments on the final AD rules Commerce

---

<sup>117</sup> *Id.* at 13.

<sup>118</sup> *Id.*

<sup>119</sup> See CFR 351.407(d)(3).

<sup>120</sup> See Memorandum, “Cost of Production and Constructed Value Calculation Adjustments for the Final Determination – Resolute FP Canada Inc.,” dated November 1, 2017 (Resolute Final Cost Memorandum), at 3-5.

<sup>121</sup> See Draft Redetermination at 15.

<sup>122</sup> See Resolute DIQR at Exhibit D-18.

<sup>123</sup> See SAA at 836.

noted that “the SAA does not refer to quality of merchandise” and inferred that “yield improvements that may continue to exist long after startup has ended and, if taken into consideration, could extend the startup period beyond the point at which commercial production is achieved.”<sup>124</sup> Nevertheless, we have considered the other metrics that Resolute FP proffered as evidence that the Atikokan Mill was not out of startup until well after the POI. Resolute FP points to Atikokan’s fluctuating LRF; however, Resolute FP fails to compare these recovery factors to its own experience at its other mills. Based on record evidence, Resolute FP achieved a company-wide LRF of [ ] for the year 2016 and [ ] for the first quarter of 2015 (Q4 2015), *i.e.*, the first three months of the POI.<sup>125</sup> During these same periods, the Atikokan Mill posted LRFs of [ ] for 2016 and [ ] for Q4 2015.<sup>126</sup> Furthermore, while Resolute FP argues that the Atikokan Mill’s saleable yield at [ ] percent is the [ ] of all the mills, this yield is certainly within the [ ] percent to [ ] percent range of the other Resolute FP mills which were producing at commercial levels.<sup>127</sup>

Resolute FP also claims the Atikokan Mill was not operating at its targeted production speeds, nor operating at full capacity. However, the SAA states that “{a} producer’s projections of future volume or cost will be accorded little weight, as actual data regarding production are much more reliable than a producer’s expectations.”<sup>128</sup> “The attainment of peak production levels will not be the standard for identifying the end of the startup period, because the startup period may end well before a company achieves optimum capacity utilization.”<sup>129</sup> Even so, we note that during Q4 2015, the first three months of the POI, the Atikokan Mill consumed [ ]

---

<sup>124</sup> *See Preamble.*

<sup>125</sup> *See* Resolute DIQR at Exhibit D-29.

<sup>126</sup> *Id.*

<sup>127</sup> *Id.* at Exhibit D-18.

<sup>128</sup> *See* SAA at 838.

<sup>129</sup> *Id.*

m3 of logs and achieved a LRF of [ ].<sup>130</sup> By comparison, the consumption at all Resolute FP mills during this period ranged from a [ ] m3 for the St Hilarion Mill to a [ ] m3 for the Thunder Bay Mill and the company as a whole achieve an average LRF of [ ].<sup>131</sup> Thus, we counter that while the Atikokan Mill may have not been operating at its peak or optimum capacity, it was operating well within the range of other Resolute FP mills producing the same products at commercial production levels.

Resolute FP also claims that Commerce has failed to address the SAA's instructions to "identify those technical difficulties associated with startup that resulted in the underutilization of facilities," yet in the very next paragraph Resolute FP states that "{d}uring the cost verification, {Commerce} itself made several key observations about these {technical} difficulties impeding commercial production."<sup>132</sup> We do not disagree that the production levels in the Atikokan Mill were limited by technical factors associated with the initial phase of commercial production. Rather, we determined based on the significant increase in production starts in conjunction with comparisons to the production experience of other Resolute FP mills, that the initial phase of production at the Atikokan Mill ended prior to the POI.<sup>133</sup> Hence, our determination was consistent with using "the number of units processed as a primary indicator of the end of the startup period."<sup>134</sup> Moreover, our review of the technical factors was to confirm that the underutilization of the mill was related to the initial phase of commercial production and not related to other non-startup issues. The mere presence of technical issues at a factory does not convey to the respondent the right to a startup adjustment. In fact, the SAA clarifies that

---

<sup>130</sup> See Resolute DIQR at Exhibit D-29.

<sup>131</sup> *Id.*

<sup>132</sup> See Resolute FP Comments on the Draft Redetermination at 10.

<sup>133</sup> See Resolute Final Cost Memorandum at 3-5.

<sup>134</sup> See *Preamble*.

“companies must demonstrate that, for the period under investigation or review, production levels were limited by technical factors associated with the *initial phase* of commercial production and not by factors unrelated to startup, such as marketing difficulties or *chronic production problems*.”<sup>135</sup> Resolute FP argues that Commerce during verification saw the defective nature of the planer that caused the underutilization of the mill. We do not disagree that there were issues with the Atikokan planer; however, we note that even at the July 2017 cost verification, Resolute FP claimed to be just emerging from the startup period that commenced in April 2015.<sup>136</sup> Thus, Resolute FP claimed that the Atikokan Mill had not yet “met full commercial production” after over two years of production.<sup>137</sup>

A startup adjustment is not intended to fully cover a company’s “uneven transition” to the desired production levels and efficiencies.<sup>138</sup> Rather, it is to address the initial phase of commercial production such as the development of process parameters, installation, adjustment, calibration, and testing of new equipment, *etc.*<sup>139</sup> Commerce has previously pointed out that problems chronic in nature (*i.e.*, marked by long occurrence or frequent in recurrence), do not reflect the types of issues surrounding the initial phase of production that the startup was intended to address.<sup>140</sup>

Resolute FP also criticizes our citation to *Agro Dutch*; however, in the *Final Determination*, Commerce merely referenced the court’s statement that “Congress

---

<sup>135</sup> See SAA at 838 (emphasis added).

<sup>136</sup> See Memorandum, “Verification of the Cost Response of Resolute FP Canada Inc. in the Antidumping Duty Investigation of Softwood Lumber from Canada,” dated July 27, 2017, at 33.

<sup>137</sup> *Id.*; see also Resolute SDQR at D-36.

<sup>138</sup> See 19 CFR 351.407(d)(4) stating that “the startup period may end well before a company achieves optimum capacity utilization.”

<sup>139</sup> See, e.g., *SRAMs from Taiwan*, 63 FR 8909, 8930.

<sup>140</sup> See, e.g., *Chlorinated Isos from Spain* IDM at Comment 9; *CTL Plate from Canada*, 63 FR 37320, 37324; and *Stainless Steel Bar from India*; *Final Results of Antidumping Duty Administrative Review*, 68 FR 47543 (August 11, 2003), and accompanying IDM at Comment 2.

unambiguously expressed its intent in the SAA where it stated that ‘{p}roduction levels will be measured based on units processed.’<sup>141</sup> In response to the Panel’s query of what “units processed” means, we have provided additional cases where we defined the term as production starts.<sup>142</sup> Resolute FP now contends that Commerce’s reliance on its own determinations “do not carry the same weight as ‘authoritative judicial opinions’” nor are they dispositive.<sup>143</sup> However, Resolute FP does not provide any “authoritative judicial opinions” of its own to support its claim for a startup adjustment.<sup>144</sup> Rather, Resolute FP merely contends that Commerce’s reliance on *Lumber from Canada* is “disingenuous.”<sup>145</sup> According to Resolute FP, in its discussions of *Lumber from Canada*, where Commerce granted a startup adjustment for a new planing facility, Commerce admits that “the planing process is a later stage in lumber production” and, hence, “the input at this stage is lumber not logs.”<sup>146</sup> Resolute FP argues that the Atikokan Mill similarly experienced “technical difficulties {with its new planer} that resulted in the underutilization” of the mill.<sup>147</sup> Consequently, Resolute FP concludes that *Lumber from Canada* actually supports granting Resolute FP a startup adjustment.

In its submissions to Commerce, Resolute FP did not provide the lumber inputs to the Atikokan planer mill. Rather, the company only provided the inputs to the initial stage of production, the sawmill.<sup>148</sup> Thus, we cannot consider what is not on the record. The SAA clearly states, “the general rule in antidumping practice {is} that a party seeking an adjustment has the burden of establishing the entitlement to that adjustment both as a legal and factual

---

<sup>141</sup> See *Final Determination* IDM at Comment 45.

<sup>142</sup> See *Draft Redetermination* at 10-16.

<sup>143</sup> See Resolute FP Comments on the *Draft Redetermination* at 17.

<sup>144</sup> *Id.* at 15-18.

<sup>145</sup> *Id.* at 17.

<sup>146</sup> *Id.* (citing *Draft Redetermination* at 13).

<sup>147</sup> *Id.* at 18 (citing SAA at 836).

<sup>148</sup> See Resolute DIQR at Exhibit D-18.

matter.”<sup>149</sup> Nevertheless, if we consider the information that Resolute FP did provide for the planer alone, *i.e.*, planer mill output and planer mill efficiencies, we find the data do not support a startup adjustment. From April 2015 to September 2015, the planer at the Atikokan Mill increased its output from [ ] MBF to [ ] MBF, a [ ]-fold increase in production, while it increased its production speeds from [ ] MBF/hr to [ ] MBF/hr.<sup>150</sup> Although these figures might not necessarily meet the optimal levels of the planer’s capabilities, they indicate a significant improvement and ramping up of production on the planer at the Atikokan Mill. As noted previously, Resolute FP claimed that the Atikokan Mill had not yet “met full commercial production” in over two years of production.<sup>151</sup> By comparison, in *Lumber from Canada*, the respondent lumber producer was granted a startup period for a new planer mill that extended three months, from July 2000 to September 2000.<sup>152</sup> If we assumed the same initial phase of commercial production for Resolute, the startup period for the Atikokan planer would extend from April 2015 to June 2015, *i.e.*, a startup period that ended three months prior to the start of the POI.

In sum, we continue to find that the Atikokan Mill reached the end of the initial phase of commercial production prior to the POI. While we do not agree that significant weight should be afforded to yield improvements, we find that even these additional metrics proffered by Resolute FP fail to support a startup adjustment for the Atikokan Mill. As noted above, both the Atikokan Mill log inputs and its LRFs were comparable with other Resolute FP mills. Although the planer at the Atikokan Mill may have been producing below its capacity and at lower than optimum efficiencies, we noted a significant ramp up in these statistics between April 2015 and September

---

<sup>149</sup> See SAA at 836.

<sup>150</sup> See Resolute DIQR at Exhibit D-18; and Resolute SDQR at Exhibit SD-30.

<sup>151</sup> See Resolute SDQR at D-36.

<sup>152</sup> See *Lumber from Canada* IDM at Comment 32.

2015.<sup>153</sup> Thus, although Resolute FP considered the planer mill to be producing below-optimum levels two years after it was installed, we find this is indicative of chronic production problems rather than technical problems surrounding the initial phase of commercial production on new equipment.<sup>154</sup> “Where technical problems persist over a long period, they may in fact signal abnormal situations that the statutory startup provisions are not intended to cover.”<sup>155</sup> While the question of how and when technical problems change from “normal” to “chronic” must be examined in light of the facts of each case, we find the *Lumber from Canada* decision is instructive of what may be typical in the industry. Consequently, we have continued to deny Resolute FP’s startup adjustment for the Atikokan Mill in these final results of redetermination.

### C. The Differential Pricing Analysis

In its decision, the Panel states that *Stupp* is a Federal Circuit decision and that the Panel is bound by its decision:

We therefore follow *Stupp* and remand to Commerce for an explanation of whether the limits on the use of the Cohen’s *d* test were met in this case or whether those limits need not be followed by Commerce. In this regard, Commerce is invited to clarify its argument concerning availability of the full universe of sales data.<sup>156</sup>

The issue before the Federal Circuit questioned the reasonableness of Commerce’s use of the Cohen’s *d* test when the test and comparison groups did not meet the three statistical criteria of “sufficient size, normal distribution and roughly equal variances.”<sup>157</sup> According to the Panel, the Federal Circuit “found that applying the test to groups having very few data points was

---

<sup>153</sup> See Resolute SDQR at Exhibit SD-30.

<sup>154</sup> *Id.* at D-36 and SD-30.

<sup>155</sup> See *SRAMs from Taiwan*, 63 FR 8909, 8930.

<sup>156</sup> See Binational Panel Order at 30.

<sup>157</sup> *Id.* at 28 (citing *Stupp*).

particularly problematic” and “could lead to upward bias and a greater number of results that pass the Cohen’s *d* test.”<sup>158</sup>

Further, the Panel states that *Mid Continent I* and *Mid Continent II* are Federal Circuit decisions and that it is bound by those decisions: “{w}e therefore follow *Mid Continent I & II* and remand to Commerce for an explanation of its choice for the Cohen’s *d* denominator, either simple averaging or an alternate choice.”<sup>159</sup> The Panel states that, in *Mid Continent I*, the Federal Circuit held “that Commerce had not adequately explained why it used simple averaging, since the statistical literature calls for weighted averaging when groups are of different sizes.”<sup>160</sup> Further, the Panel summarized that in *Mid Continent II*, the Federal Circuit understood that a weighted average would be used when the size of the test and comparison groups differed, and a simple average when the two groups are of equal size.<sup>161</sup> The Federal Circuit also suggested that because Commerce has the “full information on prices” (*i.e.*, the full populations in the two groups), that it could use the single standard deviation of all prices in both groups.<sup>162</sup>

We have provided the analysis below in accordance with the Binational Panel Order, which remanded the *Final Determination* to Commerce to address the issues raised in *Stupp*, *Mid Continent I*, and *Mid Continent II*. We note that the CIT has sustained on several occasions that Commerce need not observe the three statistical criteria of sufficient size, normal distribution, and roughly equal variances.<sup>163</sup> Further, the CIT has sustained Commerce’s use of a simple

---

<sup>158</sup> *Id.*

<sup>159</sup> *Id.* at 31-32.

<sup>160</sup> *Id.* at 30 (citing *Mid Continent I*, 940 F.3d at 674-75).

<sup>161</sup> *Id.* at 31.

<sup>162</sup> *Id.* (quoting *Mid Continent II*, 31 F.4th at 1377 (“seemingly the preferred way if the full set of population data is available”)).

<sup>163</sup> See *Stupp Corp. v. United States*, 619 F. Supp. 3d 1314, 1321 (CIT 2023) (holding that “Commerce has adequately addressed the {Federal Circuit’s} concerns”), appeal docketed Federal Circuit No. 2023-1663 (March 27, 2023); *SeAH Steel Corp. v. United States*, 619 F. Supp. 3d 1309, 1313 (CIT 2023) (*SeAH*) (denying motion for reconsideration and holding that “Commerce’s use of a population, rather than a sample, in the application of the

average to calculate the denominator of the Cohen's *d* coefficient.<sup>164</sup> In general, although CIT decisions are not binding,<sup>165</sup> they are persuasive authority and provide further support that Commerce's differential pricing analysis as a whole is reasonable.

### 1. A Summary of The Statistical Criteria

In an investigation, section 777A(d)(1)(A) of the Act provides that Commerce will determine whether subject merchandise is sold at LTFV using either an average-to-average (A-to-A) comparison method or a transaction-to-transaction method (T-to-T) comparison method.<sup>166</sup> Alternatively, section 777A(d)(1)(B) of the Act provides that Commerce may use an average-to-transaction (A-to-T) comparison method if two requirements are satisfied: (1) there is a pattern of prices that differ significantly for comparable merchandise among purchasers, regions, or time periods (the pattern requirement); and (2) either the A-to-A method or the T-to-T method cannot account for such differences (the meaningful difference requirement). The SAA states that section 777A(d)(1)(B) of the Act is to address the concern that the A-to-A method "could conceal 'targeted dumping.'"<sup>167</sup> "Targeted dumping" represents "such situations, {where} an

---

Cohen's *d* test sufficiently negates the questionable assumptions of thresholds that were raised in *Stupp* . . . . The concerns described in *Stupp* that might be raised when the Cohen's *d* test is applied to samples are inapplicable because in this case, Commerce applied the Cohen's *d* test to a population."); *Marmen Inc. v. United States*, 627 F. Supp. 3d 1312, 1322 (CIT 2023) (holding that "{b}ecause Commerce adequately explained how its methodology is reasonable, the Court holds that Commerce's use of the Cohen's *d* test applied as a component of its differential pricing analysis is in accordance with law."); appeal docketed Federal Circuit No. 2023-1877 (May 11, 2023); and *NEXTEEL Co., Ltd. v. United States*, 676 F. Supp. 3d 1345, 1357 (CIT 2023) (*NEXTEEL*) ("Commerce explained that its analysis in the Cohen's *d* test is to determine whether prices differ significantly between sales to a specific purchaser, region, or time period (*i.e.*, the test group) and all other comparable sales (*i.e.*, the comparison group), and these sales prices include all of the sales prices that are used to calculate each respondent's weighted-average dumping margin and represent the full population of sales prices to each test and comparison group . . . . The Court holds that Commerce has adequately explained how its methodology is reasonable").

<sup>164</sup> See *Mid Continent Steel & Wire, Inc. v. United States*, 680 F. Supp. 3d 1346 (CIT 2024) (*Mid Continent IV*), appeal docketed Court No. 2024-1556 (March 11, 2024).

<sup>165</sup> See *Algoma Steel Corp.*, 865 F.2d 240, 243 (Fed. Cir. 1989).

<sup>166</sup> See 19 CFR 351.414(b)(1) and (b)(2); see also SAA at 842-43.

<sup>167</sup> See 19 CFR 351.414(b)(3); see also SAA at 842.

exporter may sell at a dumped price to particular customers or regions, while selling at higher prices to other customers or regions.”<sup>168</sup>

Section 777A(d)(1)(B) of the Act does not specify a particular methodological approach to determine whether the pattern or meaningful difference requirements have been satisfied to apply an alternative comparison methodology based on the A-to-T method. In exercising its discretion under the statute, Commerce developed a differential pricing analysis, which has been generally sustained by the courts.

For the reasons explained below, Commerce finds that the three statistical criteria identified in *Stupp* (*i.e.*, normality of distribution, number of observations, and homogeneity of the variances)<sup>169</sup> are not relevant to Commerce’s Cohen’s *d* test. These statistical criteria do not serve as the basis for Dr. Cohen’s thresholds, which are used to determine whether the price differences are significant. Further, when applying the Cohen’s *d* test, the three statistical criteria are relevant when using sampled data to ensure that the sample results reliably represent the full population of data, but they are not relevant here because the results of the Cohen’s *d* test are based on the full universe of prices in each test and comparison group and are not based on sampled data.

Below, Commerce provides further explanation regarding the application of the Cohen’s *d* test in determining whether the prices among purchasers, regions, or time periods differ significantly. The Cohen’s *d* test is the first step, along with the ratio test and the meaningful difference test, in Commerce’s differential pricing analysis which is used to examine whether the A-to-A method is appropriate to calculate a respondent’s weighted-average dumping margin. First, Commerce presents the definition of effect size as a measure of significance. Second,

---

<sup>168</sup> See SAA at 842.

<sup>169</sup> See *Stupp*, 5 F.4th at 1360.

Commerce examines the role of the U.S. price data and the importance of the fact that these data encompass the entire universe, the full population, of price data in each test group and in each comparison group and not sampled data. Third, Commerce addresses the relationship between the three statistical criteria and Dr. Cohen's proposed thresholds. Fourth, Commerce addresses the references to the academic literature by the Federal Circuit. Fifth, Commerce discusses the Federal Circuit's hypothetical examples and the distortions alleged to exist when prices are within a narrow range or when there are a small number of prices in a test or comparison group. Sixth, and lastly, Commerce discusses the overview of its use of an alternative comparison methodology as a result of its application of the differential pricing analysis.

a. Effect Size as a Measure of Significance

The purpose of the Cohen's *d* test is to evaluate the extent to which the prices to a particular purchaser, region, or time period differ significantly from the prices of all other sales of comparable merchandise. Dr. Cohen's *d* coefficient is a recognized measure of effect size which gauges the extent of the difference between the means of two groups.<sup>170</sup> The Cohen's *d* coefficient, as a measure of effect size, provides "a simple way of quantifying the difference between two groups and has many advantages over the use of tests of statistical significance alone."<sup>171</sup> Further, the Cohen's *d* coefficient "quantifies the size of the difference between two groups, and may therefore be said to be a true measure of the significance of the difference."<sup>172</sup>

---

<sup>170</sup> See generally Cohen, Jacob, *Statistical Power Analysis for the Behavior Sciences*, Second Edition, Lawrence Erlbaum Associates (1988) (*Cohen*) (included in Academic Texts from the Federal Circuit Memorandum at Attachment 1).

<sup>171</sup> See Coe, Robert, "It's the Effect Size Stupid: What Effect Size Is and Why It Is Important," paper presented at the Annual Conference of the British Educational Research Association (September 2002) (*Coe*), at 1 (included in Academic Texts from the Federal Circuit Memorandum at Attachment 3).

<sup>172</sup> *Id.* at 7.

The precise purpose for which Commerce relies on the Cohen’s *d* test is to satisfy the statutory language to measure whether a difference in prices is significant.

In his presentation of “power analysis,” Dr. Cohen states that the “power of a statistical test is the probability that it will yield statistically significant results.”<sup>173</sup> Further, he writes that the “power of a statistical test depends upon three parameters: the significance criterion, the reliability of the sample results, and the ‘effect size,’ that is, the degree to which the phenomenon exists.”<sup>174</sup> Any such “statistical test” is to “provide a basis for rejection of the *null hypothesis* and hence for the proof of the existence of the phenomenon under test.”<sup>175</sup> The *null hypothesis* is academic-speak to describe a question which the test is meant to “reject” to demonstrate that a phenomenon exists.

In Commerce’s Cohen’s *d* test, the question is whether prices between the test group and the comparison group differ significantly. Thus, the *null hypothesis* is that the prices do not differ significantly—that the prices are equal between the two groups, and, thus, the difference is zero. Rejection of the *null hypothesis* indicates that the prices do differ. In the Cohen’s *d* test, when the difference is equal to or exceeds Dr. Cohen’s “large” threshold, then the difference is significant, which in Commerce’s analysis establishes that the prices in the test group pass the Cohen’s *d* test.<sup>176</sup> To understand the context of the effect size as presented by Dr. Cohen, we examine each of these parameters.

The “significance criterion” represents “the standard of proof that the phenomenon exists, or, equivalently, the standard of disproof . . . that the phenomenon does not exist.”<sup>177</sup> For

---

<sup>173</sup> See *Cohen* at 1.

<sup>174</sup> *Id.* at 4.

<sup>175</sup> *Id.* at 2 (emphasis added).

<sup>176</sup> See *Certain Softwood Lumber Products from Canada: Preliminary Affirmative Determination of Sales at Less Than Fair Value*, 82 FR 29833 (June 30, 2017), and accompanying Preliminary Decision Memorandum (PDM) at 14.

<sup>177</sup> See *Cohen* at 2.

example, a significance level, *a*, defines the chance of rejecting a true null hypothesis, *i.e.*, a Type I error. The complement of this, *b*, represents the risk of accepting a false null hypothesis, *i.e.*, a Type II error. Each of these error types represents potential sampling error where the sampled data does not represent the full population.

For the second parameter, Dr. Cohen states that the “reliability (or precision) of a sample value is the closeness with which it can be expected to approximate the relevant population value. It is necessarily an estimated value in practice, as the population value is generally unknown. Depending upon the statistic in question, and the specific statistical model on which the test is based, reliability may or may not be directly dependent upon the unit of measurement, the population value, and the shape of the population distribution. However, it is always dependent upon the size of the sample.<sup>178</sup> For the difference in the means test, the reliability of the sample results, as measured by the *t*-test, is dependent upon the normality of the distribution, the variance, and the sample size.<sup>179</sup>

For the third parameter, “the phrase ‘effect size’ {is} to mean ‘the degree to which the phenomenon is present *in the population*,’ or ‘the degree to which the null hypothesis is false.’”<sup>180</sup> In the Cohen’s *d* test, Commerce compares the prices in the test group with the prices in the comparison group: “where two populations are being compared, the null hypothesis usually takes the form ‘the difference in the value of the relevant parameters is zero.’”<sup>181</sup> Thus, “when the null hypothesis is false, it is false to some specific degree, *i.e.*, *the effect size (ES) is some specific nonzero value in the population*. The larger this value, the greater the *degree* to

---

<sup>178</sup> See Cohen at 6.

<sup>179</sup> *Id.* at 19.

<sup>180</sup> *Id.* at 9-10 (emphasis added); see also Ellis, Paul D., *The Essential Guide to Effect Sizes: Statistical Power, Meta-Analysis, and the Interpretation of Research Results*, Cambridge University Press, 2010 (*Ellis*) (included in Academic Texts from the Federal Circuit Memorandum at Attachment 2), at 5 (“The best way to measure an effect is to conduct a census of an entire population but this is seldom feasible in practice.”).

<sup>181</sup> See Cohen at 9.

which the phenomenon under study is manifested.”<sup>182</sup> Thus, the effect size is a parameter, a characteristic, of the full population, which exists independent of an analysis based on data sampled from that full population. Further, when the null hypothesis posits that the difference in the means is zero, a non-zero effect size quantifies the degree to which the means differ.

b. Full Population or Sampled Data

When calculating the effect size, or any parameter (*e.g.*, the mean, standard deviation), the parameter’s calculated value may be based on a full population or on a sample of that population. When the value is based on sampled data, the calculated value is an estimate of the actual (unknown) value of the population. As such, the reliability of the estimated value to represent the unknown value of the population may be evaluated using statistical inference, *i.e.*, the statistical significance of the estimated value. The role of statistical significance in research and data analysis is the overall subject of Dr. Cohen’s text along with many other academic books and papers. As discussed above, the statistical significance is dependent upon various criteria, which for parameters such as the mean and standard deviation include the three statistical criteria enumerated by the Federal Circuit in *Stupp*.

Commerce’s dumping analysis assesses the pricing behavior of the respondent in the U.S. market. The U.S. sale prices on which this analysis is based constitute the full population of sales data and are not a sample of a respondent’s sales data (*i.e.*, the data are for *all* sales in the United States of subject merchandise by a company during the POI or POR). The basis for this analysis is the respondent’s U.S. sales of the subject merchandise for a given period of time. By definition, these U.S. sales comprise the complete universe of sales on which the respondent’s

---

<sup>182</sup> *Id.* at 10 (emphasis in original); see also *Ellis* at 4-5 (“An effect size refers to the magnitude of the results as it occurs, or would be found, *in the population*. Although effects can be observed in the artificial setting of a laboratory or sample, effect sizes exist in the real world” (emphasis added)).

weighted-average dumping margin depends. The differential pricing analysis examines all such sales to determine whether the A-to-A method is the appropriate comparison methodology on which to base this calculation. Therefore, in the context of the calculation of the weighted-average dumping margin, the data used are not a sample, but rather constitute the entire population of a respondent's sales of subject merchandise during the period under examination for the calculation of the weighted-average dumping margin.

The Cohen's  $d$  test evaluates the extent to which the U.S. prices to a particular purchaser, region, or time period differ from the U.S. prices of all other sales of comparable merchandise. In the pattern requirement, the Act requires Commerce to consider whether U.S. prices for comparable merchandise to a particular purchaser, region, or time period (*i.e.*, the test group) differs significantly from the U.S. prices to other purchasers, regions, or time periods (*i.e.*, the comparison group). As such, the Act refocuses Commerce's analysis to calculate the respondent's weighted-average dumping margin from the pricing behavior of the respondent in the U.S. market to consider the pricing behavior to the test group separate from the pricing behavior to the comparison group. Accordingly, the sales to the test group and the sales to the comparison group are not sampled, but each constitutes a separate population of sale prices, each of which represents all the sales of the comparable merchandise to each group. Thus, the sales to each of these two groups, the test and comparison groups, themselves constitute the full population of data in the context of the calculation of the mean, standard deviation, and Cohen's  $d$  coefficient for the purpose of the pattern requirement.

When Commerce applies the Cohen's  $d$  test, which compares the mean (*i.e.*, weighted-average) prices in the test group with the mean prices in the comparison group, the prices in each group include all prices of comparable merchandise to the test group and to the comparison

group. The calculated mean and standard deviation of the test and comparison groups are not estimates but are the actual values for the mean and standard deviation of each group. Accordingly, the reliability or the statistical significance of these values is inapposite. Unlike with sampled data where the estimated parameters will change with each sample selected from a population, each time these parameters would be calculated for a given pair of test and comparison groups, the exact same results will be found because the calculated parameters are the parameters of the entire population and not an estimate of the parameters based on a sample. Therefore, because Commerce's Cohen's *d* test is based on the full population of prices to each of the two groups, the statistical criteria need not be observed as part of Commerce's analysis to establish that the results reliably reflect all the respondent's U.S. price data.

c. Dr. Cohen's Proposed Thresholds Are Not Dependent on the Statistical Criteria

The Federal Circuit has previously affirmed the use of Dr. Cohen's large, 0.8, threshold as a measure of significance in the difference in prices.<sup>183</sup> In *Stupp*, however, the Federal Circuit expressed concern that diverging from the statistical criteria, may "undermine the usefulness of the interpretive cutoffs,"<sup>184</sup> *i.e.*, the large 0.8 threshold used in the Cohen's *d* test to determine that the price difference is significant.

Once the size of the effect, *i.e.*, the Cohen's *d* coefficient, has been calculated, such measurements "must be interpreted to extract meaning."<sup>185</sup> Dr. Ellis provides three avenues by which one may interpret the measurements of effect size: context, contribution to knowledge,

---

<sup>183</sup> See *Mid Continent I*, 940 F.3d at 673 ("Commerce reasoned that even a small absolute difference in the means of the two groups can be significant (for the present statutory purpose) if there is a small enough dispersion of prices within the overall pool as measured by a proper pooled variance or standard deviation; the 0.8 standard is "widely adopted" as part of a "commonly used measure" of the difference relative to such overall price dispersion; and it is reasonable to adopt that measure where there is no better, objective measure of effect size. We agree with the Trade Court that this rationale adequately supports Commerce's exercise of the wide discretion left to it under {section 777A(d)(1)(B) of the Act}" (citation omitted)).

<sup>184</sup> See *Stupp*, 5 F.4th at 1357.

<sup>185</sup> See *Ellis* at 32.

and Dr. Cohen's thresholds.<sup>186</sup> Dr. Ellis recognized that Dr. Cohen's established thresholds for evaluating the magnitude of the effect size which are "easy to grasp" and "are sufficiently grounded in logic for Cohen to hope that his cut-offs 'will be found to be reasonable by reasonable people.'"<sup>187</sup> Despite some criticism of Dr. Cohen's thresholds, they are nevertheless, widely accepted.<sup>188</sup>

Dr. Cohen proposed, for each type of effect size, "as a convention, {effect size} values to serve as operations definitions of the qualitative adjectives 'small,' 'medium,' and 'large.'"<sup>189</sup> Dr. Cohen recognized the possible "dangers" of his proposal: "{t}he definitions are arbitrary, such qualitative concepts as 'large' are sometimes understood as absolute, sometimes as relative; and thus they run a risk of being misunderstood."<sup>190</sup> Nonetheless, Dr. Cohen expected, that "{a}lthough arbitrary, the proposed conventions will be found to be reasonable by reasonable people."<sup>191</sup> As attested to by Dr. Ellis in general, and found by the Federal Circuit with respect to Commerce's Cohen's *d* test, "the 0.8 standard is 'widely adopted' as part of a 'commonly used measure' of the difference relative to such overall price dispersion; and it is reasonable to adopt that measure where there is no better, objective measure of effect size."<sup>192</sup>

Dr. Cohen established numerical values to represent operational definitions of a small, medium, and large effect for numerous types of statistical tests.<sup>193</sup> For the difference in the means, Dr. Cohen established numerical thresholds of 0.2, 0.5, and 0.8 to describe, respectively,

---

<sup>186</sup> *Id.* at 35.

<sup>187</sup> *Id.* at 41 (citation omitted).

<sup>188</sup> *Id.* at 40 ("Cohen's cut-offs provide a good basis for interpreting effect size and for resolving disputes about the importance of one's results.")

<sup>189</sup> *See Cohen* at 12.

<sup>190</sup> *Id.*

<sup>191</sup> *Id.* at 13.

<sup>192</sup> *See Mid Continent I*, 940 F.3d at 673 (internal citation omitted).

<sup>193</sup> *See Ellis* at 41 (which lists various measures of effect size and the corresponding numerical thresholds for small, medium, and large effects).

a small, medium, and large effect size.<sup>194</sup> These values are not dependent on the statistical criteria cited by the Federal Circuit. Dr. Cohen presented different approaches to illustrate these qualitative thresholds. First, Dr. Cohen presented various “U Measures” where he calculated various measure of non-overlap, which are dependent on certain assumptions, *i.e.*, normality of the data and equal variances, to permit such calculations.<sup>195</sup> Further, Dr. Cohen presented real-world observations to demonstrate each of the effect-size threshold levels.<sup>196</sup> These initial illustrations, as well as researchers’ experience in applying these thresholds, as Dr. Cohen expected, have confirmed that “the proposed conventions will be found to be reasonable by reasonable people.”<sup>197</sup>

For the “large” 0.8 threshold, Dr. Cohen described the effect as the difference in IQ of a PhD graduate and a college freshman, the difference in IQ between a college graduate and a student with only a 50-50 chance of passing high school, or the difference in height between 13 and 18 year-old girls.<sup>198</sup> This level of difference was selected by Commerce as a conservative standard to determine that the observed price differences are significant since this threshold is “grossly perceptible and therefore {represents} large differences.”<sup>199</sup> Commerce could have also used the medium 0.5 threshold as it “is conceived as one large enough to be visible to the naked eye.”<sup>200</sup> However, Commerce elected to use the most conservative, large threshold to provide the strongest evidence that the observed prices differed significantly.

As discussed above, given that Dr. Cohen’s thresholds are not based on a statistical analysis, the concerns about the statistical criteria do not impact the usefulness of Dr. Cohen's

---

<sup>194</sup> See *Cohen* at 24-27.

<sup>195</sup> *Id.* at 21-23.

<sup>196</sup> *Id.* at 24-27.

<sup>197</sup> See *Ellis* at 32.

<sup>198</sup> See *Cohen* at 27; see also *Ellis* at 41.

<sup>199</sup> See *Cohen* at 27.

<sup>200</sup> *Id.* at 26.

thresholds. These thresholds are illustrated by real-world observations and experience, and, thus, are not derived from any particular statistical criterion such as normality of distribution, equal variances, or number of observations.

d. The Federal Circuit's Citations of the Academic Literature

In *Stupp*, the Federal Circuit ordered Commerce to provide further explanation regarding three statistical criteria, which the plaintiff in that case had argued must be met for the application of the Cohen's  $d$  test. In its decision, the Federal Circuit included several quotations from the academic literature which the Federal Circuit found raised questions concerning the need for Commerce to address the statistical criteria in its Cohen's  $d$  test. Commerce finds that each of the concerns do not impact the Cohen's  $d$  test as applied as part of the differential pricing analysis. We address each in turn.

The Federal Circuit's first concern, based on a quote from Dr. Cohen, was that "we maintain the assumption that the populations being compared are normal and with equal variability, and conceive them further as equally numerous."<sup>201</sup> However, this quotation must be understood in context when the entire passage is viewed:

{Section} 2.2.1  $d$  as Percent Nonoverlap: The U Measures. If we maintain the assumption that the populations being compared are normal and with equal variability, and conceive them further as equally numerous, it is possible to define measures of nonoverlap (U) associated with  $d$  which are intuitively compelling and meaningful.<sup>202</sup>

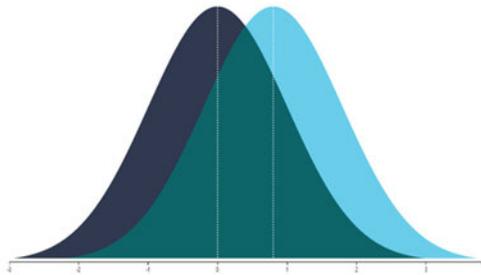
As discussed above, Dr. Cohen proposed small, medium, and large thresholds to be used as a convention to interpret a measure of effect size which he believed would be found to be reasonable. For the difference of the means, Dr. Cohen selected three numeric values to serve as these thresholds, *i.e.*, 0.2, 0.5, and 0.8 for small, medium, and large effects, respectively. As one

---

<sup>201</sup> See *Stupp*, 5 F.4th at 1357 (citation omitted).

<sup>202</sup> See *Cohen* at 21.

approach to illustrate these different thresholds of effect size, Dr. Cohen provided the “U Measures.”<sup>203</sup> In the measure of “non-overlap,” Dr. Cohen considered the extent that two compared sets of data do not overlap one another, *i.e.*,  $U_1$ . In the illustration below, the percent non-overlap is the total area which is under only one curve; or, alternatively, the area that is not under both curves.



The closer together the two bell curves, the smaller the difference in the means, the smaller associated effect size, and the smaller the non-overlap area (*i.e.*, the area under each of the curves and not under the other).<sup>204</sup> Conversely, the farther apart the two bell curves, the greater the difference in the means, the larger the associated effect size, and the larger the non-overlap area. To calculate the amount of non-overlap, one must know the areas under each bell curve, which requires the statistical criteria cited by Dr. Cohen (*i.e.*, normality and equal variances). These are the statistical criteria questioned by the Federal Circuit. Without those assumptions, the  $U_1$  measures (or the  $U_2$  or  $U_3$  measures) could not be calculated as presented in Table 2.2.1.<sup>205</sup> Thus, the assumptions of normality and equal variances are required for Dr. Cohen to calculate

---

<sup>203</sup> *Id.* at 21-23.

<sup>204</sup> In the extreme, when effect size,  $d$ , is zero, then the area of non-overlap is zero. *See Cohen* at 21 (“When  $d = 0$ , and therefore either population distribution is perfectly superimposed on the other, there is 100% overlap or 0% nonoverlap, hence  $U_1 = 0$ .”).

<sup>205</sup> *See Cohen* at 22.

the various U measures included in Table 2.2.1,<sup>206</sup> and do not relate to Dr. Cohen’s proposed thresholds. As discussed above, Dr. Cohen selected the values for his proposed thresholds for which he was confident that reasonable people would find reasonable.

Similarly, the Federal Circuit’s first citation of *Grissom*<sup>207</sup> that the “usual interpretation ... of estimating the percentile standing ... with the supposed normal distribution ... would be invalid”<sup>208</sup> also involved a similar analysis concerning the overlap of the two compared sets of sampled data. Figure 3.1 graphically demonstrates the percent of the comparison group whose values are less than the mean of the test group ( $\mu_e$ ).<sup>209</sup> Similar to Dr. Cohen’s calculation of non-overlap of two sets of data, the calculation of the “percentile standing” of 84 percent requires the assumptions that the two sets of data be normally distributed and have equal variances.<sup>210</sup> As explained above, without the assumptions of normality and equal variances, the area beneath the curve of the control group that is less than the mean of the experimental group could not be quantified (*i.e.*, the “density function” that permits the calculation of 84 percent of the control group (the area under the curve) is less than the mean of the experiment group). As with Dr. Cohen’s percent non-overlap, this does not impact Commerce’s application of the Cohen’s *d* test or the use of Dr. Cohen’s thresholds.

---

<sup>206</sup> Indeed, different assumptions could have been made by Dr. Cohen, but what is required is that the algebraic formula for each curve to be known so that the area under each curve can be calculated to derive the various U measures. For example, a normal distribution is represented by a known mathematical equation which is dependent on the variance of the distribution.

<sup>207</sup> See Grissom, Robert J. and John J. Kim, *Effect Size for Research, Univariate and Multivariate Applications*, Second Edition, Routledge (2012) (*Grissom*) (included in Academic Texts from the Federal Circuit Memorandum at Attachment 5).

<sup>208</sup> See *Stupp*, 5 F.4th at 1358 (quoting *Grissom* at 66).

<sup>209</sup> See *Grissom* at 62.

<sup>210</sup> The “percentile standing” is Dr. Cohen’s  $U_3$  measure. See *Cohen* at 21-23 and Table 2.2.1 (where  $d=1$ ,  $U_3=84.1\%$ ).

The Federal Circuit’s second citation of *Grissom*<sup>211</sup> must also be taken in its complete context:

Glass *et al.* (1981) suggested the use of Equations 3.1 and 3.2 because treatment can affect variances and, therefore, cause heteroscedasticity. However, *if the two populations that are being compared are assumed to have equal variances*, then a better estimate of the denominator of a standardized difference between population means can be made if one pools the data from both samples to estimate the common  $\sigma$  {*i.e.*, the standard deviation of a population} instead of using  $s_b$  {*i.e.*, the standard deviation of sample data  $b$ } that is based on the data of only one sample.<sup>212</sup>

Equations 3.1 and 3.2 define the denominator of the effect size as the standard deviation of the control (*i.e.*, comparison) group, whereas Dr. Grissom is stating that, in the situation involving sampling where the variances are equal, the denominator can be an average of the two variances.<sup>213</sup> This does not indicate that the use of the calculated standard deviations distorts the calculation or estimation of the effect size, but only suggests an alternative approach to calculate the denominator of the “*d*” coefficient in Dr. Grissom’s equations.<sup>214</sup>

As cited by the Federal Circuit, Professor Coe states that “the interpretation of the ‘standardized mean difference’ measure of effect size {(e.g., Cohen’s *d*)} is very sensitive to violations of the assumptions of normality,”<sup>215</sup> including where “interpretation of effect sizes in terms of percentiles is very sensitive to violations of this assumption {of a normal distribution}.”<sup>216</sup> This involves the same issue raised with respect to sampled data discussed in *Cohen* and *Grissom* above, that the interpretation of the effect size, based on non-overlap or

---

<sup>211</sup> See *Stupp*, 5 F.4th at 1358 (quoting *Grissom* at 68).

<sup>212</sup> See *Grissom* at 68 (emphasis as quoted in *Stupp*).

<sup>213</sup> Although if the variances are equal between the test and comparison groups, then presumably the average of these two values would be the same as the value of the standard deviation for either group.

<sup>214</sup> See *Grissom* at 63 (the “*d*” coefficient is equal to the ratio of the difference in the means of the sampled data of the experimental and control groups divided by the standard deviation of the sampled data of the control group).

<sup>215</sup> See *Stupp*, 5 F.4th at 1358 (quoting *Coe* at 14).

<sup>216</sup> See *Coe* at 5.

standing percentile, must necessarily be based on a normal distribution to permit the calculation of the percentages in those analyses.

Further, Professor Coe discusses the issue of a non-normal distribution *within the context of sampled data* and its potential impact on the estimation of effect size when the effect size is identical.<sup>217</sup> In Professor Coe's example, as with the hypothetical sample data in *Grissom*, 84 percent of the data in the comparison group with a normal distribution is less than the mean of the test group, but with the non-normal distribution, 97 percent of the data in the comparison group is less than the mean of the test group. Because these two comparisons both have an effect size of one, the effect size of the data with a non-normal distribution is underestimated because the difference in the means, as seen in Figure 3(b), is greater than the data with a normal distribution in Figure 3(a). Thus, the effect size of the non-normal distribution, equal to one, underestimates the actual difference in the means. This suggests that a non-normal distribution has the opposite effect from the allegation that the estimated effect size is positively *biased*, and resolves the concerns expressed by the Federal Circuit about finding "false positives." If anything, this aspect of the Cohen's *d* coefficient makes it less likely that Commerce's methodology will result in finding prices that differ significantly among purchasers, regions, or time periods. Moreover, when using the entire population as opposed to a sample, the issue concerning an inherent bias in an estimated effect size is no longer relevant.

The Federal Circuit also referenced Dr. Lane's online text concerning the interpretation of effect size.<sup>218</sup> Dr. Lane's statement is simply a recognition, as discussed above, that the

---

<sup>217</sup> *Id.* at 12-13 ("The interpretations of effect-sizes given in Table I {*i.e.*, standing percentiles} depend on the assumption that both control and experimental groups have a 'Normal' distribution").

<sup>218</sup> *See Stupp*, 5 F.4th at 1358.

measure of effect size uses the variability of the underlying data to determine the yardstick by which the difference in the means is measured:

When the effect size is measured in standard deviation units as it is for Hedges's  $g$  and Cohen's  $d$  {*i.e.*, both different measures of effect size}, it is important to recognize that the variability in the subjects has a large influence on the effect size measure. Therefore, if two experiments both compared the same treatment to a control but the subjects were much more homogeneous in Experiment 1 than in Experiment 2, then a standardized effect size measure would be much larger in the former experiment than in the latter.<sup>219</sup>

In other words, the variability in the data (*i.e.*, variance) is the yardstick by which the difference in the means is measured. For a given difference in the means, the effect size is smaller when the variability in the underlying data is larger; conversely, the effect size is larger when the variability in the underlying data is smaller.

The Federal Circuit also identified a concern regarding a conclusion by Dr. Algina and his co-authors<sup>220</sup> that:

After simulating Cohen's  $d$  on various data that followed a mixed-normal distribution, *e.g.*, a heavy-tailed distribution, they concluded that Cohen's  $d$  was not robust to mixed-normal distributions, and that applying Cohen's  $d$  to such data caused serious flaws in interpreting the resulting parameter.<sup>221</sup>

The purpose of the *Algina* paper is to propose for specific circumstances an alternative formula to calculate effect size based on the difference of the means,<sup>222</sup> analogous to those proposed by Glass and Hedges as different approaches to quantify the variations in the data. As a result of their analysis, the authors ask:

---

<sup>219</sup> See Lane, David, *et al.*, *Introduction to Statistics*, Online Edition, Chapter XIX, Part 3: "Difference Between Two Means" (included in Academic Texts from the Federal Circuit Memorandum at Attachment 4).

<sup>220</sup> See Algina, James, H.J. Keselman, and Randall D. Penfield, "An Alternative to Cohen's Standardized Mean Difference Effect Size: A Robust Parameter and Confidence Interval in the Two Independent Groups Case," *Psychological Methods*, Volume 10, Number 3, pp. 317-28 (2005) (*Algina*) (included in Academic Texts from the Federal Circuit Memorandum at Attachment 6).

<sup>221</sup> See *Stupp*, 5 F.4th at 1358.

<sup>222</sup> See *Algina* at 317 ("The authors argue that a robust version of Cohen's effect size constructed by replacing population means with 20% trimmed means and the population standard deviation with the square root of a 20% Winsorized variance is a better measure of population separation than is Cohen's effect size.").

Why then is  $\delta$  so much smaller for the mixed normal distributions? The answer is that because the mixed normal distribution is a heavy-tailed distribution and there are more scores in the tails than one would find in a normal distribution, the standard deviation, which is very sensitive to the tails of a distribution, is quite large. This, in turn, reduces  $\delta$ .<sup>223</sup>

The situation addressed here is the same as that discussed in *Coe* concerning a heavy-tailed distribution. As noted in *Coe* and *Algina*, this results in an estimated effect size that *understates* the magnitude of the difference in the means, which contradicts the claim that violations of its alleged statistical criteria result in false positives. Further, this does not impact Dr. Cohen's definition of his thresholds as discussed above.

Dr. Johnson Ching-Hong Li further analyzed the robustness of six proposed alternative approaches to Dr. Cohen's *d* coefficient.<sup>224</sup> The Federal Circuit noted the conclusion of Dr. Li's analysis that:

Li concluded that Cohen's *d* "was found to be inaccurate when the normality and homogeneity-of-variances assumptions were violated in this study, thereby severely affecting the accuracy of *d* in evaluating the true {effect size} in the research literature."<sup>225</sup>

Again, the inaccuracies identified by Dr. Li, as well as others, involve "the accuracy of *d* in evaluating the true {effect size}" where "*d*" is the estimated Cohen's *d* coefficient of the sampled data in comparison with the actual value of the Cohen's *d* coefficient for the population. In Commerce's Cohen's *d* test, Commerce *does not estimate* the Cohen's *d* coefficient in the Cohen's *d* test, but *calculates the actual* Cohen's *d* coefficient based on the entire population of sale prices, not on a limited sample of the sale price data. Thus, the concerns raised by Dr. Li and others are not germane to the results of Commerce's Cohen's *d* test.

---

<sup>223</sup> *Id.* at 319.

<sup>224</sup> See Li, Johnson Ching-Hong, "Effect Size Measures in a Two-Independent Samples Case with Nonnormal and Nonhomogeneous Data," *Behavior Research* 48, pp. 1560-74, Springer (2016) (*Li*) (included in Academic Texts from the Federal Circuit Memorandum at Attachment 7).

<sup>225</sup> See *Stupp*, 5 F.4th at 1358 (quoting *Li* at 1571).

Lastly, the Federal Circuit returned to *Grissom* with the concern that:

The use of Cohen's  $d$  with test groups consisting of very few observations may be particularly problematic. {} The literature concludes that using Cohen's  $d$  in such a situation may produce an upward bias in the calculated effect size. "Both Cohen's  $d$  and Glass's  $d_G$  have some positive bias (*i.e.*, tending to overestimate their respective parameters), the more so the smaller the sample sizes and the larger the effect size in the population." An upward bias might produce more "passing" results under the Cohen's  $d$  test, which would tend to exaggerate dumping margins.<sup>226</sup>

Bias is the systematic tendency to overestimate, or underestimate, the parameter's actual value because of some characteristic of the sampled data. Here, the Federal Circuit points to *Grissom* which observes that a positive bias may be introduced to an estimated effect size based on smaller sample sizes. However, as discussed above, Commerce's Cohen's  $d$  test *does not estimate* the Cohen's  $d$  coefficient, let alone overestimate it, but rather *calculates the actual* Cohen's  $d$  coefficient based on the entire populations of sale prices. Accordingly, *there is no bias*, positive or negative, in the results of Commerce's Cohen's  $d$  test. Additionally, as discussed below, the results of the Cohen's  $d$  test determine whether the requisite U.S. prices differ significantly among purchasers, regions, or time periods, and do not "exaggerate dumping margins," which require the comparison of U.S. prices to normal values (*i.e.*, prices and/or costs outside the United States).

e. The Federal Circuit's Hypothetical Example

The Federal Circuit also raised a concern about a situation when the prices in a test group "hover around the same value."<sup>227</sup> The Federal Circuit proposed a hypothetical example:

Consider, for example, ten purchasers of a product, each of which purchases five units. Assume that the per-unit sales prices for a particular purchaser are not normally distributed and are all the same, or nearly the same (*e.g.*, \$100.01, \$100.01, \$100.01, \$100.01, and \$99.99). Assume further that the per-unit sales

---

<sup>226</sup> *Id.*, 5 F.4th at 1358-59 (quoting *Grissom* at 70).

<sup>227</sup> *Id.*

prices across the entire set of purchasers are also very similar, falling within a relatively small range (such as between \$99.92 and \$101.01).

As the variance within each test group approaches zero, the denominator in the Cohen's *d* equation is greatly reduced and ... the resulting effect-size parameter is increased, tending to artificially inflate the dumping margins for a set of export sales prices that has minimal variance.<sup>228</sup> An objective examiner inspecting those export sales prices would be unlikely to conclude that they embody a "pattern" of prices that "differ significantly."

Although the problem in that situation is a function of Commerce's use of the simple average pooled standard deviation, our concern is also related to the number of observations being compared and the distribution of those observations—requiring larger test groups tends to decrease the likelihood that a test group would have sales prices with near-zero variance, and requiring normality also tends to decrease that likelihood as the number of observations increases.<sup>229</sup>

Underpinning this concern of the Federal Circuit appears to be the continued supposition that not adhering to the statistical criteria will have "some positive bias ... tending to overestimate {the} respective parameters."<sup>230</sup>

First, we offer the following observation regarding the Federal Circuit's statement that, "{a}s the denominator is reduced, the resulting effect size parameter is increased, tending to artificially inflate the dumping margins for a set of export sales prices that has minimal variance."<sup>231</sup> The term "dumping margin" means the amount by which the normal value exceeds the export price or constructed export price (*i.e.*, the U.S. price) of the subject merchandise. The Cohen's *d* test only examines the relationship of prices of the subject merchandise within the

---

<sup>228</sup> The Federal Circuit also includes the specific assumption that, as the variance of the test group approaches zero, the value of the denominator approaches one half of the standard deviation of the comparison group. As discussed above, the specifics on the formula for the denominator is the subject of *Mid Continent I* and *Mid Continent II*; however, the general proposition is true that as the variance of either or both the test and comparison groups is made smaller and smaller, the denominator will be reduced and the calculated effect size will increase. This is simply an arithmetic tautology, and not evidence that the test is flawed. Also, as discussed above, this is the overall premise of effect size based on the difference of the means, that the significance of the difference between the means of the two groups is based on the variation of the underlying data.

<sup>229</sup> See *Stupp*, 5 F.4th at 1359.

<sup>230</sup> *Id.* (quoting *Grissom* at 70).

<sup>231</sup> *Id.*, 5 F.4th at 1359.

U.S. market and does not examine whether the U.S. price is at less than normal value. The magnitude of the Cohen's *d* coefficient, or whether it is small, medium, or large, does not involve the comparison of U.S. price with normal value, and, therefore, it is unrelated to and cannot create dumping margins. Whether U.S. prices are dumped, *i.e.*, sold at less than normal value, is not part of the Cohen's *d* test. Rather, the Cohen's *d* test is part of Commerce's analysis to determine whether there is a pattern of prices that differ significantly in the U.S. market. Whether prices differ significantly between purchasers, regions, or time periods in the U.S. market does not change whether dumping exists due to the respondent's overall U.S. pricing behavior when the U.S. price is compared with normal value. It is also important to recognize that when U.S. prices differ significantly, it does not mean that the U.S. prices passing the Cohen's *d* test are dumped. In fact, U.S. sale prices that pass the Cohen's *d* test may not be dumped at all when those prices are greater than the normal value. Thus, there is no basis to conclude that the approach used to consider whether U.S. price differences are significant can artificially inflate the dumping margins.

Further, as noted above, Commerce has had several approaches, including the "Pasta Test,"<sup>232</sup> the "P/2 Test,"<sup>233</sup> the "Nails Test,"<sup>234</sup> and now the differential pricing analysis,<sup>235</sup> to

---

<sup>232</sup> See *Borden, Inc. v. United States*, 23 CIT 372, Slip Op. 99-50 (CIT June 4, 1999).

<sup>233</sup> See *Notice of Final Determination of Sales at Less Than Fair Value: Coated Free Sheet Paper from the Republic of Korea*, 72 FR 60630 (October 25, 2007), and accompanying IDM.

<sup>234</sup> See *Certain Steel Nails from the People's Republic of China: Final Determination of Sales at Less Than Fair Value and Partial Affirmative Determination of Critical Circumstances*, 73 FR 33977 (June 16, 2008); *Certain Steel Nails from the United Arab Emirates: Notice of Final Determination of Sales at Not Less Than Fair Value*, 73 FR 33985 (June 16, 2008), as modified in *Multilayered Wood Flooring from the People's Republic of China: Final Determination of Sales at Less Than Fair Value*, 76 FR 64318 (October 18, 2011); *Mid Continent Nail Corp. v. United States*, Slip. Op. 2010-47 (CIT 2010); and *Mid Continent Nail Corp. v. United States*, Slip. Op. 2010-48 (CIT 2010).

<sup>235</sup> See *Xanthan Gum from Austria: Final Determination of Sales at Less Than Fair Value*, 78 FR 33354 (June 4, 2013), and accompanying IDM; *Xanthan Gum from the People's Republic of China: Final Determination of Sales at Less Than Fair Value*, 78 FR 33351 (June 4, 2013), and accompanying IDM; *Polyester Staple Fiber from Taiwan: Preliminary Results of Antidumping Duty Administrative Review; 2011-2012*, 78 FR 17637 (March 22, 2013), and accompanying PDM; and *Polyester Staple Fiber from Taiwan: Final Results of Antidumping Duty Administrative Review; 2011-2012*, 78 FR 38938 (June 28, 2013).

ascertain whether there is a pattern of prices that differ significantly. One thing which Commerce has determined to be consistent across the application of these various tests is that if U.S. prices differ, that fact alone does not always mean that masked dumping or dumping exists. In general, the purpose of the differential pricing analysis is to examine whether masked, or “targeted,” dumping exists when one of the standard comparison methods cannot account for a respondent’s U.S. pricing behavior. Even without a differential pricing analysis, dumping, masked or not, exists as a result of the respondent’s pricing behavior in the U.S. market. It is only a question of what extent of the dumping that is part of a respondent’s U.S. pricing behavior is accounted for by Commerce’s margin calculations. Thus, there is no logical basis to conclude that the approach used to consider whether U.S. price differences are significant, *i.e.*, the Cohen’s *d* test, could “artificially inflate the dumping margins.”

For example, this means that Resolute FP’s estimated weighted-average dumping margin from the *Final Determination*, 3.20 percent<sup>236</sup> in the investigation before the Panel, was not created because of the Cohen’s *d* test. It is mistaken to infer that an analysis of any differences in Resolute FP’s U.S. prices, *i.e.*, the Cohen’s *d* test, results in “excessive” dumping margins. Resolute FP’s U.S. pricing behavior, along with any dumping of subject merchandise in the U.S. market, was determined by Resolute FP’s pricing decisions in both the U.S. and Canadian markets during the POI, not by Commerce’s present use of the Cohen’s *d* test. As noted above, dumping is measured by comparing U.S. price with normal value for each U.S. sale, and the results of this comparison are not measured by the Cohen’s *d* test. The Cohen’s *d* test only indicates that U.S. prices, whether dumped or not, differ significantly amongst themselves in the

---

<sup>236</sup> We note that this rate may be revised as a result of this redetermination, and could be any respondent in an investigation or administrative review where the A-to-A method cannot account for the differences in the respondent’s U.S. pricing behavior.

U.S. market. In this instance, the analysis and results of the Cohen's *d* test have no bearing on the amount of dumping found to exist for Resolute.

Second, as discussed above, the parameters (*e.g.*, mean, standard deviation, effect size) calculated in the Cohen's *d* test are not estimates of the parameters based on sampled data from a larger population. In contrast, the values of parameters based on sampled data will change with each sample of data drawn from a population; thus, such calculated values estimate the values of the actual parameters of the population data. Such estimates may be biased, upward or downward, based on the characteristics of the sampled data. For example, the fat-tailed, non-normal distribution discussed in *Coe* underestimates the significance of the difference in the means in Figure 3(b) vis-à-vis Figure 3(a).<sup>237</sup> However, Commerce's application of the Cohen's *d* test does not rely on sampled data or on estimated parameters, but rather calculates the actual parameters, including the Cohen's *d* coefficient, of the U.S. prices which reflects the actual measure of the significance of the difference in prices between the test and comparison groups.

To address the Federal Circuit's concern that prices that "are all the same or nearly the same" may invalidate the interpretation of the results of the Cohen's *d* test, Commerce analyzed the extreme situation from the hypothetical scenario where all prices to each purchaser are identical and where there are two purchasers, A and B (there could be more purchasers, but limiting the example to two purchasers will simplify this discussion). All sales to purchaser A are priced at \$100, and all sales to purchaser B are priced at \$101, within the range of the Federal Circuit's example. In this situation, the standard deviation of the prices to each purchaser is zero (*i.e.*, the "denominator is reduced" to the lowest value possible), the number of sales to each purchaser is not material (there could be two sales to one purchaser and 200 sales to the other

---

<sup>237</sup> See *Coe* at 12-13.

purchaser), the prices to each customer are not normally distributed (as assumed in the example), and the resulting Cohen's  $d$  coefficient is infinite (*i.e.*,  $d = \$1 \div 0$ ).

If Commerce were to apply its differential pricing analysis to this hypothetical example, then Commerce's analysis would not have resulted in the application of an alternative calculation methodology. The pattern requirement requires that Commerce first define "significant" and then identify prices that differ significantly. Although there are many possible definitions of significance ranging from qualitative to quantitative measures, based on the concept of effect size, Commerce has defined significance based on the result of the Cohen's  $d$  test. In the above example, when the prices are uniform to each purchaser and when these two prices differ, the difference is significant because the Cohen's  $d$  coefficient is infinite, *i.e.*, "large." This is consistent with the analyses by Dr. Cohen and others who visualize the effect size based on the non-overlap of the two sets of data. In this example, there is no overlap of the two groups of prices at all. Indeed, Professor Coe concludes that "if there were no overlap at all ... then this would seem like a very substantial difference."<sup>238</sup> Therefore, by this definition, the prices differ significantly. However, the meaningful difference requirement, discussed in the following paragraph, imposes a contextual interpretation on the results of the Cohen's  $d$  test that there also be meaningful amount of masked dumping that the A-to-A method cannot account for it.

In general, there are five scenarios to describe the possible relationship between a difference in U.S. prices and normal value.<sup>239</sup> Of these five scenarios, there is only one that will result in a "meaningful difference" according to the differential pricing analysis: where the normal value falls within the range of U.S. prices and where the difference between normal value

---

<sup>238</sup> *Id.* at 2.

<sup>239</sup> See, e.g., *Certain Hot-Rolled Steel Flat Products from Japan: Final Determination of Sales at Less Than Fair Value and Final Affirmative Determination of Critical Circumstances*, 81 FR 53409 (August 12, 2016), and accompanying IDM at 31-34.

and U.S. prices are large enough that there is “a non-*de minimis* amount of dumping, but there is also a meaningful amount of offsets to impact the identified amount of dumping under the A-to-A method with offsets.”<sup>240</sup> In the above example, if the normal value is less than \$100, then no sales are dumped; if the normal value is greater than \$101, then all sales are dumped. In either scenario, there is no meaningful difference when the overall dumping margin is calculated using the A-to-A method or the A-to-T method. When the normal value is within the range of U.S. prices, the maximum amount of an individual dumping margin will be where the normal value is \$101, and the individual dumping margin for a sale to purchaser A will be one percent, which is below the *de minimis* threshold in an LTFV investigation. For all sales to purchasers A and B, if the quantity sold to each is identical, then the maximum weighted-average dumping margin will be one half of one percent. Thus, there can be no meaningful difference where the largest possible weighted-average dumping margin is less than the *de minimis* threshold, which in an LTFV investigation is two percent. Consequently, one could conclude in the context of the differential pricing analysis that, although there are significant price differences based on the Cohen’s *d* test, those price differences in U.S. prices are not meaningful and the results of the differential pricing analysis, even in this extreme hypothetical example, would not permit the application of an alternative comparison methodology.

f. Commerce Application of an Alternative Comparison Methodology

In the Binational Panel Order, the Panel notes that “Resolute FP also argued that Commerce has turned its differential pricing analysis into the general rule, rather than the exception.”<sup>241</sup> Resolute FP states that “Commerce found that at least one company had 33% or

---

<sup>240</sup> *Id.* at 33.

<sup>241</sup> *See* Binational Panel Order at 21.

more of its sales pass the Cohen's *d* test in 145 of the 165 antidumping investigations and reviews undertaken during the POI, amounting to 87% of the determinations."<sup>242</sup>

Resolute FP's logic distorts the language of the Act. First, the "exception" of section 777A(d)(1)(B) of the Act relates to the application of the A-to-T method, and not to finding that the pattern requirement is present. Further, "exception" does not necessarily connote something that is rare or unusual, but just that the A-to-T method is an "exception" or "alternative" to one of the two standard comparison methods of section 777A(d)(1)(A) of the Act. Contrary to Resolute FP's suggestion, Commerce's actual application of the Cohen's *d* test in the context of the differential pricing analysis resulted in the application of an alternative comparison methodology to a relatively small number of respondents. The significance of the price differences which exist within a company's pricing behavior in the U.S. market will limit the application of an alternative comparison methodology to situations only where masked dumping meaningfully impedes the A-to-A method from calculating an accurate weighted-average dumping margin. By way of example, for calendar year 2015, the year in which the final determination underlying *Stupp* was published, for all published final determinations in LTFV investigations,<sup>243</sup> Commerce calculated final rates for 18 companies.<sup>244</sup> Of those 18 respondents, Commerce applied an alternative comparison methodology to four companies, including SeAH Steel Corp., the plaintiff in *Stupp*. This means that only 22 percent of respondents with calculated rates had their weighted-average dumping margin calculated using an alternative comparison methodology. Further, of the 14 companies whose rates were based on the A-to-A method, two of the calculated rates were zero. Similar overall results were repeated in calendar

---

<sup>242</sup> *Id.* (internal citation omitted).

<sup>243</sup> See Draft Redetermination at Attachment I.

<sup>244</sup> This excludes rates calculated based on section 776 of the Act (*i.e.*, total adverse facts available).

year 2021, the year of the *Stupp* opinion, where Commerce applied an alternative comparison methodology for 15 companies (21 percent of the total) and applied the A-to-A method for 58 companies, eight of which had a zero rate.<sup>245</sup>

Therefore, the Federal Circuit’s concern that groups with small variations in prices or a small sample size result in “false positives” or “artificially {inflated} dumping margins” is not borne out by the data regarding the real-world application of the Cohen’s *d* test in the context of the differential pricing analysis. Using Dr. Cohen’s thresholds is a reasonable approach to interpret whether the difference in the prices is significant and the further interpretation of the difference in the prices in the context of the calculation of dumping margins ensures the reasonable and limited application of the alternative comparison methodology.

## 2. Calculation of the Denominator of the Cohen’s *d* Coefficient

The Cohen’s *d* test is based on a measure of effect size, the Cohen’s *d* coefficient, which in general is the ratio of the difference in the means, divided by the “standard deviation,” *i.e.*, the dispersion of values in the underlying data. It is this denominator of the Cohen’s *d* coefficient, the “standard deviation,” that is the subject of the Federal Circuit’s opinions in *Mid Continent I* and *Mid Continent II*.<sup>246</sup>

In the *Final Determination*, as in the final determination in the underlying investigation at issue in *Mid Continent II*, Commerce calculated the denominator of the Cohen’s *d* coefficient as the “simple average” of the standard deviations<sup>247</sup> of the test group and the comparison group.

The Federal Circuit remanded Commerce’s use of the simple average, requiring that “Commerce

---

<sup>245</sup> See Draft Redetermination at Attachment II.

<sup>246</sup> See *Mid Continent II*, 31 F.4th at 1377 (“Commerce recognized that the function of the denominator in the Cohen's *d* coefficient is to be a "yardstick to gauge the significance of the difference of the means" of the sales prices of the test and comparison groups.”).

<sup>247</sup> The “simple average” is really the square root of the simple average of the variances of the test group and the comparison group, as discussed below. See *Cohen* at 44.

must either provide an adequate explanation for its choice of simple averaging or make a different choice, such as use of weighted averaging or use of the standard deviation for the entire population.”<sup>248</sup>

First, we address each of the three options for calculating the denominator of the Cohen’s *d* coefficient as contemplated by the Federal Circuit. Next, we address the Federal Circuit’s understanding of the role of sample size in Dr. Cohen’s text (or perhaps “in the academic literature”). Lastly, we discuss the circumstances of Commerce’s Cohen’s *d* test and the use of a simple average to calculate the denominator of the Cohen’s *d* coefficient.

a. A Simple Average of the Standard Deviations for the Two Groups

As detailed above, the Federal Circuit’s decision in *Mid Continent II* is premised on a finding that Commerce departed from academic literature in relying on a simple average in the Cohen’s *d* denominator. Accordingly, we have examined the academic literature on the record of this remand segment and the circumstances in which the Cohen’s *d* test is performed in Commerce’s differential pricing analysis in light of the Federal Circuit’s decision. Based on this evaluation, as described below, we find that the literature does support Commerce’s use of a simple average when sampled data are not used, the standard deviations of the full populations are known, and the standard deviations of both populations are not equal.

As discussed above, Dr. Cohen presented effect size as part of his concept of power analysis,<sup>249</sup> where effect size is one element of Dr. Cohen’s power analysis and represents “the degree to which the phenomenon is present *in the population*.”<sup>250</sup> In Dr. Cohen’s general formulation of “the effect size (ES) we wish to detect,” he defines the “*d*” coefficient as the

---

<sup>248</sup> *Id.*, 31 F.4th at 1381.

<sup>249</sup> See *Cohen* at 1 (“The purpose of this book is to provide a self-contained comprehensive treatment of statistical power analysis from an ‘applied’ viewpoint.”).

<sup>250</sup> *Id.* at 9 (emphasis in original, and added).

“standardizing of the raw effect size as expressed in the measurement unit of the dependent variable {i.e., the difference in the means} by dividing it by the (common) standard deviation of the measures *in their respective populations*, the latter also in the original measurement unit.”<sup>251</sup>

Mathematically, for the difference in the means, Dr. Cohen expressed the effect size as,

$$d = \frac{m_A - m_B}{\sigma}$$

for a one-tailed case, or as

$$d = \frac{|m_A - m_B|}{\sigma}$$

for a two-tailed case,<sup>252</sup> where  $m_A$  and  $m_B$  are the “*population means*” and  $\sigma$  is “the standard deviation of *either population (since they are assumed equal)*.”<sup>253</sup> Dr. Cohen repeated this definition of effect size for a population in his discussion of the “power tables,” where “ $\sigma$  is the common within-population standard deviation (i.e.,  $\sigma_A = \sigma_B = \sigma$ ).”<sup>254</sup> Thus, the common within-population standard deviation is defined by Dr. Cohen as equal to the standard deviation of population A or the standard deviation of population B, which are assumed to be equal.

In Dr. Cohen’s general formulation of effect size, the denominator of the ratio, *i.e.*, the “standard deviation,” is the standard deviation of population A or the standard deviation of population B, which are assumed to be identical. Thus, when the standard deviations of population A and population B are equal, either of the standard deviations of the two populations is used as the denominator. However, when the standard deviations of population A and population B are not equal,<sup>255</sup>

the definition of  $d$  will be slightly modified. Since there is no longer a common within-population  $\sigma$ ,  $d$  is defined as above (formulas (2.2.1) and (2.2.2)), but instead

---

<sup>251</sup> *Id.* at 20 (emphasis added).

<sup>252</sup> *Id.* (referencing *Cohen* at 20 (equations 2.2.1 and 2.2.2, respectively)).

<sup>253</sup> *Id.* (emphasis added).

<sup>254</sup> *Id.* at 27.

<sup>255</sup> *Id.* at 43-44 and equation 2.3.2.

of  $\sigma$  in the denominator, the formula requires the root mean square of  $\sigma_A$  and  $\sigma_B$ , that is, the square root of the mean of the two variances:

$$\sigma' = \sqrt{\frac{\sigma_A^2 + \sigma_B^2}{2}}$$

In other words, when the standard deviations of the two populations are not equal, then the denominator of the effect size should be the simple average of the two, unequal standard deviations of population A and population B. In this scenario, there is no common within-population standard deviation. Moreover, unlike a common within-population standard deviation where one of the population standard deviations is used as the denominator, the denominator in this scenario is defined as the root mean square, *i.e.*, the simple average, of the standard deviations of population A and population B.<sup>256</sup> Throughout *Cohen*, when the standard deviations of the two populations are known, the denominator of the effect size is either the common population standard deviation when the standard deviations of the two populations are equal,<sup>257</sup> or the root square mean of the two standard deviations when the standard deviations of the two populations are unequal.<sup>258</sup>

Consistent with Dr. Cohen’s general formulation of effect size based on the means and standard deviations of two populations, Dr. Ellis recognized:

{t}he best way to measure an effect is to conduct a census of an entire population but this is seldom feasible in practice. Census-based research may not even be desirable if researchers can identify samples that are representative of broader

---

<sup>256</sup> *Id.* at 44-45 (“Note that this value is not the standard deviation of either the population of men workers or that of women workers, but the root mean square of their respective population standard deviations,  $\sigma'$  (formula (2.3.2)).”).

<sup>257</sup> *Id.* at 20 and 27.

<sup>258</sup> *Id.* at 44, 60 (“The inequality of population  $\sigma$  values results only in a standardization of the difference in population means by the root mean square of the population variances (formula (2.3.2)) instead of the common population standard deviation.”), 61 (“Since she is assuming that  $\sigma_S^2 \neq \sigma_C^2$ , the standardizing unit cannot be the common within-population standard deviation, but is instead the square root of the mean of the two variances, *i.e.*,  $\sqrt{(\sigma_S^2 + \sigma_C^2)/2}$  (formula (2.3.2)).”), 63 (“Note that  $d_4'$  is simply the  $m_P - m_C$  difference, standardized by the common within-population standard deviation (or, if  $\sigma_P^2 \neq \sigma_C^2$ , their root mean square,  $\sigma'$ , formula (2.3.2)).”), and 65 (“where  $\sigma$  is either the common population standard deviation or  $\sigma'$  from formula (2.3.2)”).

populations and then use inferential statistics to determine whether sample-based observations reflect population-level parameters.<sup>259</sup>

However, given Dr. Cohen's general formulation of effect size and the  $d$  coefficient where the denominator of the ratio was defined generally as the "standard deviation," Dr. Ellis observed:

{t}he only tricky part in this calculation is figuring out the population standard deviation. *If this number is unknown, some approximate value must be used instead.* When he originally developed this index, Cohen (1962) was not clear on how to solve this problem, but there are now at least three solutions. These solutions are referred to as Cohen's  $d$ , Glass's delta or  $\Delta$ , and Hedges'  $g$ . As we can see from the following equations, the only difference between these metrics is the method used for calculating the standard deviation:

$$\text{Cohen's } d = \frac{M_1 - M_2}{SD_{pooled}}$$

$$\text{Glass's } \Delta = \frac{M_1 - M_2}{SD_{control}}$$

$$\text{Hedges' } g = \frac{M_1 - M_2}{SD^*_{pooled}}$$

Choosing among these three equations requires an examination of the standard deviations of each group.<sup>260</sup>

Thus, when the standard deviations of the two populations are *unknown*, Dr. Ellis and other academic authors provide alternatives with which to estimate the denominator of the effect size.

b. A Weighted Average of the Standard Deviations of the Two Groups

As noted in the equations above, Dr. Ellis provides different formulations for the "pooled standard deviation" when the actual population standard deviations are unknown. Additionally, as noted above, when calculations are based on sampled data, the calculated values estimate the

---

<sup>259</sup> See *Ellis* at 5.

<sup>260</sup> *Id.* at 10 (emphasis added).

actual values of the parameters for the full population. As an estimate for the denominator of the effect size, for two of the three options, Dr. Ellis presented the following equations:<sup>261</sup>

For Cohen's  $d$ :<sup>262</sup>

$$SD_{pooled} = \sqrt{\frac{\sum(X_A - \bar{X}_A)^2 + \sum(X_B - \bar{X}_B)^2}{n_A + n_B - 2}}$$

For Hedges'  $g$ :<sup>263</sup>

$$SD_{*pooled} = \sqrt{\frac{(n_A - 1)SD_A^2 + (n_B - 1)SD_B^2}{n_A + n_B - 2}}$$

In each of these equations, the variable  $n$  represents the sample size of each group of data.

When based on sampled data, Dr. Cohen states that “generally, we can define the effect size *in the sample* (ES<sub>s</sub>) using sample statistics in the same way as we define it for the population, and a statistically significant ES<sub>s</sub> is one which exceeds an appropriate criterion value.”<sup>264</sup> Dr. Cohen also provides an estimation of effect size when the analysis is based on sampled data:<sup>265</sup>

Accordingly, we redefine our ES index,  $d$ , so that its elements are sample results, *rather than population parameters*, and call it  $d_s$ . For all tests of the difference between means of independent samples,

$$d_s = \frac{\bar{X}_A - \bar{X}_B}{s}$$

---

<sup>261</sup> The denominator for Dr. Ellis' third option, Glass'  $\Delta$ , is simply the standard deviation of the control (*i.e.*, comparison) group and does not include the standard deviation of the test group to derive the “pooled standard deviation.”

<sup>262</sup> *Id.* at 26.

<sup>263</sup> *Id.* at 27.

<sup>264</sup> See Cohen at 17 (emphasis in the original).

<sup>265</sup> *Id.* at 66-67 and equations 2.5.1 and 2.5.2 (emphasis added).

where  $\bar{X}_A$  and  $\bar{X}_B$  = the two sample means, and  
 $s$  = the usual *pooled within sample estimate of the population standard deviations*,  
that is,

$$s = \sqrt{\frac{\sum(X_A - \bar{X}_A)^2 + \sum(X_B - \bar{X}_B)^2}{n_A + n_B - 2}}$$

The equation to estimate the denominator of the effect size based on sampled data, the “pooled” standard deviation, is identical to that included by Dr. Ellis for the Cohen’s  $d$  coefficient, *i.e.*, the “pooled standard deviation.”

Professor Coe’s discussion of effect size is consistent with that of Dr. Cohen and Dr.

Ellis:

{t}he ‘standard deviation’ is a measure of the spread of a set of values. Here it refers to the standard deviation of the population from which the different treatment groups were taken. In practice, however, this is almost never known, so it must be estimated either from the standard deviation of the control group, or from a ‘pooled’ value from both groups.<sup>266</sup>

In his discussion of “Which ‘standard deviation’?,” Professor Coe presents different arguments for and against using different approaches to provide the “best *estimate* of standard deviation.”<sup>267</sup>

One option is the standard deviation of a “control group,” *i.e.*, Glass  $\Delta$  as presented by Dr. Ellis.

A second option is a “‘pooled’ estimate of standard deviation,” which is “essentially an average of the standard deviations of the experimental and control groups (Equation 4).”<sup>268</sup> Each of

Professor Coe’s approaches is an estimate of the actual standard deviation,  $\sigma$ , of Dr. Cohen’s general formulation of effect size, and rely on sampled data rather than on the actual standard deviations of the populations for which the difference in the means is tested.

---

<sup>266</sup> See Coe at 2.

<sup>267</sup> *Id.* at 6 (emphasis added)

<sup>268</sup> *Id.* at 6-7. Equation 4 is identical to the  $SD^*_{pooled}$  for Hedges’  $g$  in Ellis at 27.

Thus, the academic literature provides for a weighted average of the standard deviations of the two groups in the context of sampled data where the actual value for the standard deviations of the two groups is unknown.

c. A Single Standard Deviation of the Data in Both Groups

In *Mid Continent II*, the Federal Circuit held that,

{t}he cited literature makes clear that one way to form the more general data-pool dispersion figure for the denominator—seemingly the preferred way if the full set of population data is available—is to use the standard deviation for the entire population.<sup>269</sup>

The Federal Circuit further held that “Commerce did not use the standard deviation of all the data for its denominator. It made that choice even while recognizing that it had the full set of data for U.S. sales for the period Commerce was reviewing.”<sup>270</sup> It then concluded that “{i}ndeed, when the entire population is known, the cited literature points toward using the standard deviation of the entire population as the denominator in Cohen’s  $d$ —which Commerce has not done.”<sup>271</sup> Consequently, the Federal Circuit indicated that Commerce may choose on remand to “use ... the standard deviation for the entire population” in the denominator of the Cohen’s  $d$  coefficient in lieu of a simple average.<sup>272</sup>

Based on Commerce’s examination of the academic literature, we find that the option that the Federal Circuit identified of using as the denominator the single standard deviation of all sale prices of the comparable merchandise, *i.e.*, of all sale prices in the test and comparison groups, is not appropriate. Under this formulation, Commerce would calculate the standard deviation of all prices included in populations A and B as a single group of commingled data. However, the

---

<sup>269</sup> See *Mid Continent II*, 31 F.4th at 1377.

<sup>270</sup> *Id.*, 31 F.4th at 1378.

<sup>271</sup> *Id.*, 31 F.4th at 1380.

<sup>272</sup> *Id.*, 31 F.4th at 1381.

academic literature demonstrates that Dr. Cohen delineates between the two distinct populations that are the source of the means whose difference is being assessed.<sup>273</sup> In Dr. Cohen's general formulation (*i.e.*, equations 2.2.1 and 2.2.2), the denominator of the effect size is "the (common) standard deviation of the measures in their respective populations," or, in other words, "the standard deviation of either population (since they are assumed equal)."<sup>274</sup> When "there is no longer a common with-in population  $\sigma$ ," then Dr. Cohen provides that " $d$  is defined as above (formulas (2.2.1) and (2.2.2)), but instead of  $\sigma$  in the denominator, the formula requires the root mean square of  $\sigma_A$  and  $\sigma_B$ ."<sup>275</sup> In Dr. Cohen's words, there may be a *common value of the standard deviations* of two populations, but Dr. Cohen does not provide for a *single standard deviation* in which the two populations are combined into one set of data. For Dr. Cohen, there are two separate populations of data, each of whose standard deviation is part of the calculation of effect size, but there is not a single standard deviation based on commingled data.

As with Dr. Cohen, Professor Coe describes the effect size as the difference in the means divided by the "standard deviation":

{t}he "standard deviation" is a measure of the spread of a set of values. Here it refers to the standard deviation of the population from which the different treatment groups were taken. In practice, however, this is almost never known, so it must be estimated either from the standard deviation of the control group, or from a "pooled" value from both groups (see question 7, below, for more discussion of this).<sup>276</sup>

Under question 7, "Which 'standard deviation'?" Professor Coe first proposes using the standard deviation of the control group, as with Glass'  $\Delta$ .<sup>277</sup> Alternatively, given difficulties in selecting a

---

<sup>273</sup> See *Cohen* at 20, 27, and 44. Even when the analysis is based on sampled data, Dr. Cohen, as well as Dr. Ellis and Professor Coe, maintain the distinction between the two groups of data in their formulas to estimate the "standard deviation," *i.e.*, the denominator, of the effect size. *Id.* at 67; *Ellis* at 26-27; and *Coe* at 6.

<sup>274</sup> See *Cohen* at 20.

<sup>275</sup> *Id.* at 44.

<sup>276</sup> See *Coe* at 2.

<sup>277</sup> *Id.* at 6; see also *Ellis* at 10.

control group, Professor Coe states that, with sampled data, “it is often better to use a ‘pooled’ estimate of standard deviation. The pooled estimate is essentially an average of the standard deviations of the experimental and control groups (Equation 4).”<sup>278</sup>

Thus, in identifying options for calculating the standard deviation in instances where full population data are not available, Professor Coe recognizes that there are two populations each with its own standard deviation. Indeed, in describing the calculation of a pooled estimate of standard deviation, Professor Coe distinguishes a pooled average of the standard deviations of the experimental and control groups from a single “pooled” standard deviation:

{n}ote that this is not the same as the standard deviation of all the values in both groups ‘pooled’ together. If, for example each group had a low standard deviation but the two means were substantially different, the true pooled estimate (as calculated by Equation 4) would be much lower than the value obtained by pooling all the values together and calculating the standard deviation.<sup>279</sup>

Professor Coe’s “Equation 4” is discussed above, and is identical to Dr. Ellis’ equation for the pooled standard deviation for Hedges  $g$ .<sup>280</sup> The cause for this overestimation is that the standard deviation within each group is calculated based on the mean within each group, whereas the standard deviation for both groups together would be the mean of all observations in both groups. This includes the variations in the data with each group as well as the variation in the data between the two groups. As noted by Professor Coe, as the difference in the means increases between the two groups, the standard deviation of all observations in both groups will also increase rather than remain constant when based on the standard deviation of the observations within each group.

---

<sup>278</sup> See Coe at 6.

<sup>279</sup> *Id.*

<sup>280</sup> Compare Coe at 6 with Ellis at 27.

To illustrate the differences in the calculations, when the standard deviations in *Coe* Equation 4 is expanded, the equation can be restated as,

$$SD_{pooled} = \sqrt{\frac{(N_E - 1) \frac{\sum(X - \bar{X}_E)^2}{N_E - 1} + (N_C - 1) \frac{\sum(X - \bar{X}_C)^2}{N_C - 1}}{N_E + N_C - 2}}$$

which simplifies to,

$$SD_{pooled} = \sqrt{\frac{\sum(X - \bar{X}_E)^2 + \sum(X - \bar{X}_C)^2}{N_E + N_C - 2}}$$

Note that this is the same equation as *Cohen* equation 2.5.2, as well as the equations for the denominator for Cohen's *d* and Hedges' *g* in *Ellis*.<sup>281</sup> Even with sampled data, each of the formulas used to estimate the denominator of the effect size maintain the separate group of data and do not commingle all of the observations to calculate a single standard deviation for all of the data combined. The standard deviation for each group is based on the square of the difference between each observation within the group and that group's mean. The standard deviation of each group, whether sampled or based on the population, is centered on the mean of each group.

The equation for a "single standard deviation" of all observations combined together differs substantially:

$$SD_{single} = \sqrt{\frac{\sum(X - \bar{X}_{E\&C})^2}{N_E + N_C - 1}}$$

In the equation for the proposed  $SD_{single}$ , the standard deviation is based on the square of the difference of each observation from the single mean of the commingled observations in both

---

<sup>281</sup> See *Ellis* at 26-27.

groups. The pooled standard deviation reflects only the variation in the data within each group because the differences are measured relative to the mean price within each group (*i.e.*,  $\bar{X}_E$  and  $\bar{X}_C$ ). However, the “single standard deviation” is measured relative to the mean of all data (*i.e.*,  $\bar{X}_{E\&C}$ ), which also includes the differences between the overall mean and the means of the two individual groups. Thus, the “single standard deviation” not only reflects the variation of the data within each group, but also the difference in the means between the two groups. Further, as the difference in the means of the two groups increases, the “single standard deviation” will also increase even though there is no change in the variances, *i.e.*, the dispersion, in the data within each of the two groups. Accordingly, the value of  $SD_{pooled}$  will remain constant because it is based on the relationship of the data within each group, however, the value of  $SD_{single}$  will increase as the difference in the means between the two groups increases.

Therefore, the option to use a single standard deviation of all data when the data are explicitly separated into two separate populations is not contemplated in the academic literature and we do not consider it to be a reasonable approach for Commerce’s Cohen’s  $d$  test because the results of the calculation reflect not just the dispersion of the data within each group, but also the dispersion of the data between the two groups, the precise aspect, *i.e.*, the difference in prices, that the effect size is meant to quantify. The academic literature, whether the effect size is based on population parameters or estimates based on sampled data, bases the calculations on the standard deviations of the data within each group of data, and not on an overall, single, commingled group of data.

#### d. Sample Sizes Do Not Limit the Use of a Simple Average

In Commerce’s use of the Cohen’s  $d$  test, the standard deviations of population A (test group) and population B (comparison group) are known, but the standard deviations of

population A and population B are not equal. Consequently, we formulate the denominator of the effect size using equation 2.3.2, where the standard deviations of population A and population B are known and not equal.

The Federal Circuit described Dr. Cohen's alternative formula set forth in equation 2.3.2 as "designed to be applied when the two groups, though of the same size, have different standard deviations."<sup>282</sup> However, after examination of the academic literature, we find that the sample size limitation, *i.e.*, that the sample sizes must be equal, does not impact or preclude the use of equation 2.3.2. Dr. Cohen does not apply the limitation of equal sample sizes, *i.e.*,  $n_A = n_B$ , in his description of equation 2.3.2 to calculate the denominator of the effect size. Rather, the sample size,  $n$ , is an important factor in the determination of the reliability of the results of an analysis result based on sampled data:

{t}he reliability (or precision) of a sample value is the closeness with which it can be expected to approximate the relevant population value. It is necessarily an estimated value in practice, since the population value is generally unknown. Depending upon the statistic in question, and the specific statistical model on which the test is based, reliability may or may not be directly dependent upon the unit of measurement, the population value, and the shape of the population distribution. However, *it {reliability} is always dependent upon the size of the sample.*<sup>283</sup>

Thus, sample sizes are an input for the  $t$ -test and the determination of whether the results of the analysis are statistically significant.<sup>284</sup> When the effect size is based on sampled data, the sample size is also an input into the calculation of the estimated effect size as can be seen in the equations presented above from *Cohen, Ellis, and Coe* for calculating a pooled standard deviation. Indeed, Dr. Cohen, with reference to the pooled standard deviation used to estimate the denominator of the effect size,<sup>285</sup> notes that "we have defined  $s$  quite generally so that it will

---

<sup>282</sup> See *Mid Continent II*, 31 F.4th at 1372 (referencing *Cohen* at 44 (equation 2.3.2)).

<sup>283</sup> See *Cohen* at 6 (emphasis added).

<sup>284</sup> *Id.* at 19-20 and 43.

<sup>285</sup> *Id.* at 67 (equation 2.5.2).

hold for all cases involving two independent samples, whether or not sample sizes are equal.”<sup>286</sup>

If the effect size were based on sampled data, then it would be determined by equations 2.5.1 and 2.5.2 where the sample sizes are an input into the estimation of the effect size of the full populations of data.<sup>287</sup>

Further, as described by Dr. Cohen, the sample size is also an input into Dr. Cohen’s power tables:

{n}ote that if  $\sigma_A \neq \sigma_B$  and it is also the case that  $n_A \neq n_B$ , the nominal values for  $t$  and power at a given significance criterion,  $\alpha$ , may differ greatly from the true values. Under these conditions ( $\sigma_A \neq \sigma_B$  and  $n_A \neq n_B$ , simultaneously), the values in Tables 2.3 {i.e., “Power of t test”) may be greatly in error.<sup>288</sup>

The condition that the sample sizes be equal for this aspect of Dr. Cohen’s power analysis is also unrelated to the measure of effect size. Effect size of the population, along with sample size,  $n$ , significance criteria,  $\alpha$ , and Dr. Cohen’s power analysis are the four parameters of statistical inference.<sup>289</sup>

There is a distinction between the nomenclature used to distinguish between a parameter (i.e., variable) for a population and a parameter based on sampled data. Dr. Algina *et al.* highlight that population parameters are designated by Greek letters and sample parameters by Latin letters:

For two populations, Cohen’s ES is,

$$\delta = \frac{\mu_2 - \mu_1}{\sigma}$$

---

<sup>286</sup> *Id.* at 67. Note that “ $s$ ” is Dr. Cohen’s calculation of the denominator of the effect size based on sampled data, whereas “ $\sigma$ ” is the denominator of the effect size based on the full populations. The formula for “ $s$ ,” *Cohen* equation 2.5.2, is included above.

<sup>287</sup> *Id.* at 66-67.

<sup>288</sup> *Id.* at 44.

<sup>289</sup> *Id.* at 14.

where  $\mu_j$  is the population mean for the  $j$ th ( $j = 1, 2$ ) level of the grouping factor and  $\sigma$  is the population standard deviation, assumed to be equal for both levels. The parameter  $\delta$  is often estimated by,

$$d = \frac{\bar{Y}_2 - \bar{Y}_1}{S}$$

where  $\bar{Y}_j$  is the mean for the  $j$ th level of the grouping factor and  $S$  is the square root of the pooled variance, which we refer to as the pooled standard deviation.<sup>290</sup>

with Dr. Algina *et al.* noting that,

Cohen used the Latin letter  $d$  to refer to the population ES. Following more typical practice we use  $d$  to refer to the sample ES and the Greek letter  $\delta$  to refer to the population ES.<sup>291</sup>

Accordingly, in Dr. Cohen's equation 2.3.2, " $\sigma$ " references the standard deviations of populations and not estimated standard deviations of sampled data.

Therefore, even though equal sample sizes are a requirement for Dr. Cohen's "Case 2" for the  $t$ -test and power analysis, the sample size is not relevant to the measure of the effect size that continues to be presented as representing the effect size of the full populations of data. Even if Dr. Cohen may have been inconsistent in following the "typical practice" of using a Greek letter to designate a population parameter and a Latin letter to designate a sample parameter, Dr. Cohen did consistently use " $\sigma$ " and " $s$ " to distinguish between the standard deviation of a population and sampled data, respectively. If the effect size were based on the sampled data, then Dr. Cohen's equation 2.5.2 would be the basis to calculate the denominator of the effect size and not Dr. Cohen's equation 2.3.2.

e. Commerce's Cohen's  $d$  Test and Use of a Simple Average

In sum, the academic literature allows for the use of a simple average when the standard deviations of the two groups of data encompass the full populations of data and when those

---

<sup>290</sup> See Algina at 318.

<sup>291</sup> *Id.*, footnote 1.

standard deviations are unequal. The academic literature also provides for the use of a weighted average of the (estimated) standard deviations of the two groups of sampled data. The academic literature does not support the use of a single standard deviation of all the data in both groups as if they constitute a single group of data.

Therefore, Commerce's use of the simple average, *i.e.*, *Cohen* equation 2.3.2, as the denominator of the Cohen's *d* coefficient, given that the actual standard deviation of each population, is known and they are unequal.<sup>292</sup> Commerce's calculation of the effect size in the Cohen's *d* test is based on the full population of sale prices of comparable merchandise to a given purchaser, region, or time period and the full population of all other sale prices of comparable merchandise (*i.e.*, the test and comparison groups, respectively).<sup>293</sup> Accordingly, Commerce's calculation of the Cohen's *d* coefficient is based on the actual means and standard deviations of the test and comparison groups. Commerce's calculation of the Cohen's *d* coefficient is not based on sampled data, and there is no estimation of the actual mean and standard deviation of the test group and of the comparison group. The academic literature provides for the use of a weighted average as a possible approach when estimating the denominator of the effect size when the actual standard deviations are not known, which is not the situation with Commerce's application of the Cohen's *d* test. Therefore, the academic literature allows for the use of the simple average to calculate the denominator of the effect size, and it does not necessarily support the use of a weighted average.

---

<sup>292</sup> We note that in the exceptional situation where the standard deviations of the test and comparison group would be equal, *i.e.*,  $\sigma_A = \sigma_B$ , the results of the simple average, *i.e.*, *Cohen* equation 2.3.2, would simply be  $\sigma = \sigma_A = \sigma_B$ . (*i.e.*,  $\sigma$  = the standard deviation of either population (since they are assumed equal); *see Cohen* at 20).

<sup>293</sup> *See Mid Continent II*, 31 F.4th at 1378 ("Indeed, in each test-group/comparison-group pair, the test and comparison groups together make up 'the entire universe, *i.e.*, population, of the available data,' because for each test group, the comparison group is all other sales data." (Internal citation omitted)).

f. Equal Reliability

Notwithstanding the explanation above, even if it is assumed, *arguendo*, that sample sizes must be equal to use Dr. Cohen’s equation 2.3.2,<sup>294</sup> which, as discussed above, involves the comparison of two full populations and not two groups of sampled data, then use of a simple average is nonetheless reasonable. The CIT, in a subsequent opinion in the *Mid Continent III* litigation, concluded that “Commerce’s assertion that sampling is not implicated in equation (2.3.2) is unsupported, as Cohen seems to use this equation in calculating statistical power.”<sup>295</sup> Thus, “{i}f Commerce continues to rely on the academic literature to support its methodology,” the CIT held that Commerce “must further explain why its choice of the simple average is reasonable in light of this inconsistency” and the issue was remanded to Commerce for further explanation or reconsideration.<sup>296</sup> Consistent with Commerce’s analysis on remand in *Mid Continent III*, Commerce now provides further reasonable justification to support the continued use of a simple average, even with the understanding (which Commerce disagrees with) that use of Dr. Cohen’s equation 2.3.2 may be limited to where the sample sizes of the two groups must be equal.

In his presentation of the parameters of the statistical power analysis, Dr. Cohen describes the “reliability of sample results and sample size”:

The reliability (or precision) of a sample value is the closeness with which it can be expected to approximate the relevant population value. It is necessarily an estimated value in practice, since the population value is generally unknown. Depending upon the statistic in question, and the specific statistical model on which the test is based, reliability may or may not be directly dependent upon the unit of measurement, the population value, and the shape of the population distribution. However, it is *always* dependent upon the size of the sample.<sup>297</sup>

---

<sup>294</sup> See *Mid Continent Steel & Wire, Inc. v. United States*, 628 F. Supp. 3d 1316 (CIT 2023) (*Mid Continent III*).

<sup>295</sup> *Id.*, 628 F. Supp. 3d at 1325.

<sup>296</sup> *Id.*, 628 F. Supp. 3d at 1326.

<sup>297</sup> See *Cohen* at 6 (emphasis in original).

Dr. Cohen further notes that:

The nature of the dependence of reliability upon  $n$  {i.e., sample size} is obvious from the illustrative formulas, and, indeed, intuitively. The larger the sample size, other things being equal, the smaller the error and the greater the reliability or precision of the results.<sup>298</sup>

Indeed, when Dr. Cohen defines the four parameters of statistical inference, “sample size (n)” represents the reliability of the sample results.<sup>299</sup> Accordingly, the sample size is a gauge of the reliability of sample results as part of Dr. Cohen’s power analysis. The larger the sample size vis-à-vis the population, the more reliable the sample results.

As stated above, it is assumed that the use of the simple average of differing standard deviations, i.e., Dr. Cohen’s equation 2.3.2, applies to an analysis involving sampled data because it is part of a power analysis which involves sampled data.<sup>300</sup> Further, it is assumed that Dr. Cohen’s equation 2.3.2 applies only when the sample sizes are equal, i.e.,  $n_A = n_B$ .<sup>301</sup> Because the sample sizes are equal in size and reliability, the estimated standard deviation for each of the sampled groups also has the same “reliability (or precision) of a sample value {which} is the closeness with which it can be expected to approximate the relevant population value.”<sup>302</sup> Consequently, a simple average of the standard deviations of the two groups is appropriate because the reliability of each value of the standard deviation is equal. In other words, when the sample sizes of the two groups are equal, then the reliability of the estimated standard deviations are the same, and it is appropriate to give equal weights, i.e., a simple

---

<sup>298</sup> *Id.* at 7.

<sup>299</sup> *Id.* at 14 (“Four parameters of statistical inference have been described: power, significance criterion (a), sample size (n), and effect size (ES).”).

<sup>300</sup> *Id.* at 43-44.

<sup>301</sup> See *Mid Continent II*, 31 F.4th at 1378 (“{Equation 2.3.2, the simple average,} comes from a section of *Cohen* that addresses a situation in which the two groups at issue are of the same size. (“CASE 2:  $\sigma_A \neq \sigma_B$ ,  $n_A = n_B$ ”).” (internal citations omitted, emphasis added)).

<sup>302</sup> See *Cohen* at 6.

average, when averaging the two estimated standard deviations to calculate the denominator of the Cohen's *d* coefficient.

In contrast, “when the sampled groups have unequal sizes {*i.e.*,  $n_A \neq n_B$ }, the cited literature uniformly teaches use of a pooled standard deviation estimate that involves weighted averaging.”<sup>303</sup> With the weighted average, the standard deviation of the group with the larger sample size is given more weight than the group with the smaller sample size.<sup>304</sup> If the sample size of group A is larger than the sample size of group B, then the reliability of the standard deviation of group A will be greater than the reliability of group B. In such a situation, the standard deviation of group A has more reliability and is given more weight than the standard deviation of group B when calculating the denominator of the Cohen's *d* coefficient. Because the group with the larger sample size has greater reliability, the weights reflect the relative reliability of the standard deviations from the two groups.

As discussed above, in Commerce's application of the Cohen's *d* test, Commerce uses the full populations of data, *i.e.*, all prices of comparable merchandise to a given purchaser, region, or time period (*i.e.*, the test group) and all prices of comparable merchandise to all other purchasers, regions, or time periods (*i.e.*, the comparison group). As a result, the standard deviations calculated for the test and comparison groups each have a reliability of 100 percent, *i.e.*, “the closeness with which {the calculated value} can be expected to approximate the relevant population value.”<sup>305</sup> In other words, the reliability of the calculated standard deviations based on the full population of sale prices to each group is identical. Because the reliability of the standard deviations based on full populations is equal, to calculate the denominator of the

---

<sup>303</sup> See *Mid Continent II*, 31 F.4th at 1378 (referencing *Cohen* at 67; *Ellis* at 26-27; and *Coe* at 6).

<sup>304</sup> See, *e.g.*, *Coe* at 6 (equation 4).

<sup>305</sup> See *Cohen* at 6.

Cohen's *d* coefficient, it is reasonable to weight these standard deviations equally, *i.e.*, a simple average, as presented in Dr. Cohen's equation 2.3.2, just as when the reliability is equal for standard deviations based on sampled data with equal sample sizes.

Commerce's use of the simple average to calculate the denominator of the Cohen's *d* coefficient is reasonable. The parameters calculated in the Cohen's *d* test, which are used to calculate the Cohen's *d* coefficient (*i.e.*, the standard deviation and mean of each group), reflect a 100 percent reliability that they represent the parameters of the population because the Cohen's *d* test includes all sale prices in the test and comparison groups. With sampled data, Dr. Cohen presumptively provides the use of a simple average, *i.e.*, equation 2.3.2, when sample sizes are equal and standard deviations differ.<sup>306</sup> The use of the simple average when the sample sizes are equal reflects that the calculated parameters used to calculate the Cohen's *d* coefficient are equally reliable. Therefore, and because the reliability of these values is also equal when the calculated parameters are based on the full population of U.S. sale prices, it is reasonable to combine the standard deviations using a simple average to calculate the denominator of the Cohen's *d* coefficient.

As discussed above, on this basis, the CIT sustained Commerce's use of a simple average to calculate the denominator of the Cohen's *d* coefficient.<sup>307</sup> Although Commerce continues to believe that Dr. Cohen's equation 2.3.2 is in the context of full populations and not sampled data,

---

<sup>306</sup> See *Mid Continent III*, 628 F. Supp. 3d at 1324 (“A test for full populations in the context of power analysis would be redundant on its face, as there would be no question of statistical significance to analyze. Thus, Commerce does not explain, and it is not discernable why Commerce believes that equations (2.2.1) and (2.2.2)—still less equation (2.3.2), which expressly implicates sample size—are intended for testing full populations {noting in footnote 12 that, in} *Mid Continent II*, the Court of Appeals discusses the use of equation (2.3.2) with sample groups, rather than full populations, implicitly recognizing that the equation does not apply only to full populations.” (internal citations omitted)).

<sup>307</sup> See *Mid Continent IV*, 680 F. Supp. 3d at 1357 (“Commerce has provided a reasonable explanation for its use of a simple average as instructed by the {Federal Circuit and the CIT} and its determination is sustained.”), appeal docketed Court No. 2024-1556 (March 11, 2024).

and its use need not be limited to where the sample sizes of the two groups must be equal, in light of that CIT's holding in *Mid Continent III*, we have provided our affirmed analysis here as well to show that in either situation, Commerce's use of a simple average to calculate the denominator of the Cohen's *d* coefficient is reasonable and lawful.

## Interested Parties' Comments

### *COALITION's Comments*

The COALITION argues that Commerce's Draft Redetermination,

provide{s} a reasonable and lawful explanation for its continued use of differential pricing and the Cohen's *d* test. Specifically, Commerce has adequately addressed questions raised by the {Federal Circuit} in {*Stupp*, *Mid Continent I*, and *Mid Continent II*} with respect to three statistical criteria (*i.e.*, normality of distribution, number of observations, and homogeneity of the variances), and with respect to {Commerce's} use of a simple average when calculating the denominator of the Cohen's *d* coefficient. The {Draft Redetermination} reflect{s} the same analyses presented in numerous other proceedings that have been affirmed by the {CIT}. {Commerce} should therefore maintain its use of and explanation for the differential pricing methodology in its final remand redetermination.<sup>308</sup>

Specific to the issue which questions whether Commerce must observe the three statistical criteria, the Panel restated that the Federal Circuit ask Commerce "to clarify {the agency's} argument that having the entire universe of data rather than a sample makes it permissible to disregard the otherwise-applicable limitations on the use of the Cohen's *d* test."<sup>309</sup> Commerce's redetermination found that the three statistical criteria are not relevant because the Cohen's *d* test "examines the entire population."<sup>310</sup> In 2023, the CIT affirmed Commerce's explanation,<sup>311</sup> just as the CIT has in other opinions.<sup>312</sup>

---

<sup>308</sup> See COALITION Comments on the Draft Redetermination at 2 (citing *Stupp*; *Mid Continent I*; and *Mid Continent II*).

<sup>309</sup> *Id.* at 5 (quoting *Stupp*, 5 F.4th at 1360 (citing Binational Panel Order at 30)).

<sup>310</sup> *Id.* (internal citation omitted).

<sup>311</sup> *Id.* at 5-6 (citing *Stupp*, 619 F. Supp. 3d at 1321, 1327).

<sup>312</sup> *Id.* at 6 (citing Draft Redetermination at 18).

The Panel also recognized that the Federal Circuit had asked Commerce to address its use of a simple average to calculate the denominator of the Cohen's *d* coefficient in light of the academic literature.<sup>313</sup> Commerce's analysis in the Draft Redetermination is consistent with its explanations in response to the Federal Circuit's remand orders, and this has now been sustained by the CIT.<sup>314</sup> Specifically, Commerce "explained that economic literature on the record supports using a simple average to calculate the denominator of the Cohen's *d* coefficient when sampling is not used, the standard deviations of the full populations are known, and the standard deviations of both populations are not equal," and that simple averaging "need not be limited to where the sample sizes of the two groups must be equal."<sup>315</sup> Thus, Commerce should continue to use the simple average in the final results of this redetermination.

#### *Canadian Parties' Comments*

The Canadian Parties contend that Commerce "bifurcates" its explanation when addressing its use of the Cohen's *d* test for the issues raised in *Stupp* and *Mid Continent I* and *Mid Continent II*, yet "the two issues are closely intertwined."<sup>316</sup> "The Cohen's *d* coefficient provides a consistent, meaningful metric" only when the three statistical criteria have been satisfied.<sup>317</sup> Without satisfying the statistical criteria, the value of the Cohen's *d* coefficient is arbitrary and the comparison with Dr. Cohen's thresholds is not meaningful.<sup>318</sup> Further, the "denominator problems identified in the *Mid Continent* litigation do not arise when the assumptions are met," thus compounding the degree to which Commerce's Cohen's *d* coefficient is not comparable with Dr. Cohen's thresholds.<sup>319</sup>

---

<sup>313</sup> *Id.*

<sup>314</sup> *Id.* (citing *Mid Continent IV*).

<sup>315</sup> *Id.* at 7 (internal citations omitted).

<sup>316</sup> See Canadian Parties Comments on the Draft Redetermination at 8.

<sup>317</sup> *Id.*

<sup>318</sup> *Id.*

<sup>319</sup> *Id.* at 8-9.

## 1. Dr. Cohen and Effect Size

According to the Canadian Parties, they present “an objective account of the derivation” of Dr. Cohen’s *d* coefficient as a measure of effect size.<sup>320</sup> Dr. Cohen and other statisticians “establish that the assumptions {*i.e.*, the statistical criteria} and the U measures are essential to the use of {the} Cohen’s *d* {coefficient}—not optional.”<sup>321</sup> “When any of the assumptions are not satisfied, the *d* coefficient does not correspond to the measures of nonoverlap, which are the foundation of Professor Cohen’s interpretive thresholds (or as {Commerce} calls them, ‘operative definitions’).”<sup>322</sup> Thus, if Commerce violates any one of the three statistical criteria, and the calculated Cohen’s *d* coefficient does not match the percent non-overlap, then the Cohen’s *d* coefficient “will not reasonably indicate the degree of difference between {the} two groups of prices, and cannot “reasonably be used as part of {Commerce’s methodology to detect so-called ‘targeted dumping.’”<sup>323</sup>

Dr. Cohen’ provided equations 2.2.1 and 2.2.2 to calculate the *d* coefficient.<sup>324</sup> “Importantly, in both equations, the standard deviations of the groups being compared are assumed to be equal, and thus, the assumption of equal variances is satisfied.”<sup>325</sup> Dr. Cohen “then defines the *d* coefficient in terms of all three assumptions and the U measures”:<sup>326</sup>

**d** as Percent Nonoverlap: The U Measure. If we maintain the *assumption* that the *populations* being compared are *normal* and with *equal variability*, and conceive them further as *equally numerous*, it is possible to *define measures of nonoverlap* (U) associated with *d* which are intuitively compelling and meaningful.<sup>327</sup>

---

<sup>320</sup> *Id.* at 9.

<sup>321</sup> *Id.*

<sup>322</sup> *Id.* (“The U measures (U<sub>1</sub>, U<sub>2</sub>, and U<sub>3</sub>) reflect the degree of overlap or nonoverlap that is measured by the Cohen’s *d* coefficient. The U measures require that all three assumptions are satisfied.” (Citing *Cohen* at 21)).

<sup>323</sup> *Id.* at 9-10.

<sup>324</sup> *Id.* at 10 (citing *Cohen* at 20).

<sup>325</sup> *Id.*

<sup>326</sup> *Id.*

<sup>327</sup> *Id.* at 10-11 (quoting *Cohen* at 21 (emphasis added by the Canadian Parties)).

Therefore, “regardless of whether one uses populations or samples, the assumptions and the U measures cannot be divorced from the Cohen’s *d* test.”<sup>328</sup>

Dr. Cohen also provides Table 2.2.1 “defining different values for *d* in terms of the U measures.”<sup>329</sup> Dr. Cohen “uses this table to create the small, medium, and large thresholds (0.2, 0.5, and 0.8) for interpreting *d*.”<sup>330</sup> For example, a large, 0.8, threshold “must correlate to a U<sub>1</sub> of 47.4%, a U<sub>2</sub> of 65.5%, and a U<sub>3</sub> of 78.8%.”<sup>331</sup> Drs. Grissom and Kim also “conclude that when the assumption of normality is not satisfied, ‘the usual interpretation of’ Cohen’s *d* ‘would be invalid.’”<sup>332</sup> “Professor Coe also discusses Cohen’s *d* and measures of effect size in terms of overlap and plots distributions on graphs when measuring Cohen’s *d* using normal distributions.”<sup>333</sup> Thus, Dr. Cohen requires that his the statistical criteria must be satisfied “such that the U measures correspond to the measures of nonoverlap provided in Table 2.2.1.”<sup>334</sup>

Dr. Cohen also “used certain real-world observations to illustrate and calibrate the differences indicated by the interpretive thresholds.”<sup>335</sup> However, the data underlying Dr. Cohen’s illustrative examples satisfy “the assumptions of normality, equal variances and equal size.”<sup>336</sup> Thus, Dr. Cohen’s use of these examples “reinforces the conclusion” that the data used in Commerce’s Cohen’s *d* test must also satisfy the same statistical criteria.<sup>337</sup>

Further, Dr. Cohen “acknowledges that when the assumption of equal variances is violated to a minor extent . . . , the denominator of the *d* coefficient may still be used,” but “the

---

<sup>328</sup> *Id.* at 11.

<sup>329</sup> *Id.* at 11-12 (citing *Cohen* at 22).

<sup>330</sup> *Id.* at 13.

<sup>331</sup> *Id.* (internal citation omitted).

<sup>332</sup> *Id.* at footnote 23 (quoting *Grissom* at 66).

<sup>333</sup> *Id.* at footnote 25 (citing *Coe* at 2).

<sup>334</sup> *Id.*

<sup>335</sup> *Id.* at 14 (citing *Cohen* at 24-27).

<sup>336</sup> *Id.*

<sup>337</sup> *Id.*

assumption of equal size must be observed.”<sup>338</sup> “Even then, Professor Cohen cautions that the U measures will no longer ‘generally’ obtain from (correlate to) the *d* coefficient.”<sup>339</sup> In other words, if the denominator of the Cohen’s *d* coefficient does not measure what Dr. Cohen intended, then the *d* coefficient does not mean what it is supposed to mean and cannot be the basis of a determination that the difference in prices is large.

Dr. Algina *et al.* also identify flaws in Dr. Cohen’s approach, including when the statistical criteria are not satisfied, and recommend alternative approaches as appropriate.<sup>340</sup> Drs. Grissom and Kim conclude that when the statistical criteria are not met, “‘the usual interpretation of’ Cohen’s *d* ‘would be invalid.’”<sup>341</sup> Dr. Ellis cautions about “the risk of error that result from violating the assumptions.”<sup>342</sup> Accordingly, Dr. Ellis presents alternatives to calculating the denominator of the Cohen’s *d* coefficient depending on the underlying data, noting, however, that use of the Cohen’s *d* coefficient is only appropriate when the standard deviations of the two groups are similar.<sup>343</sup> Therefore, when the assumption of equal variances is violated, the Cohen’s *d* coefficient should not be used.

## 2. Commerce’s Explanation Fails to Address the Need for the Statistical Criteria

The academic literature clearly explains that the statistical criteria must be satisfied, and when Commerce’s Cohen’s *d* test violates these assumptions, then its analysis fails to identify prices that differ significantly. Commerce’s reasoning is based on five explanations, as addressed below.<sup>344</sup>

---

<sup>338</sup> *Id.* at 15 (citing *Cohen* at 43-44).

<sup>339</sup> *Id.* at 15-16 (citing *Cohen* at 44).

<sup>340</sup> *Id.* at 16 (citing *Algina* at 318).

<sup>341</sup> *Id.* (quoting *Grissom* at 66).

<sup>342</sup> *Id.* at 17 (citing *Ellis* at 54).

<sup>343</sup> *Id.* (citing *Ellis* at 9-11).

<sup>344</sup> *Id.* at 17-18.

First, Commerce asserts that it does not need to observe the three statistical criteria because the Cohen's  $d$  test is based on the full population of prices in each of the test and comparison groups; that the statistical criteria are only relevant when sampled data is used to estimate the actual values of the population parameters.<sup>345</sup> "This is a red herring."<sup>346</sup> There is no difference between an analysis based on full populations or sampled data, and Commerce's claim to the contrary ignores the academic literature and does not permit Commerce to use Dr. Cohen's thresholds when the statistical criteria have not been met. Indeed, Commerce's explanation contradicts Dr. Cohen's description of the U measures, and Commerce has failed to reconcile its failure to observe the statistical criteria in light of Dr. Cohen's presentation of his  $d$  coefficient.<sup>347</sup> In *Stupp*, the Federal Circuit stated that "violating the assumptions 'can subvert the usefulness of the interpretive cutoffs.'"<sup>348</sup> A Cohen's  $d$  coefficient of 0.8 provides meaningful information only when the statistical criteria are satisfied, and if not satisfied, then it does not have the same meaning, whether the underlying data are full populations or sampled data. Without satisfying the statistical criteria, Commerce is not measuring the same thing as Dr. Cohen did with the U measures.<sup>349</sup> Therefore, Commerce reliance on its "'population' rationale does not mitigate {Commerce's} unreasonable application of the Cohen's  $d$  test."<sup>350</sup>

Second, Commerce claims that Dr. Cohen's thresholds are not dependent on the statistical criteria or Dr. Cohen's U measures.<sup>351</sup> In Commerce's claims, Dr. Cohen "conceived of the thresholds as based solely on 'real-world observations' and independent of the U measures

---

<sup>345</sup> *Id.* at 19-20.

<sup>346</sup> *Id.* at 20.

<sup>347</sup> *Id.* (citing *Cohen* at 21).

<sup>348</sup> *Id.* (quoting *Stupp*, 5 F.4th at 1360).

<sup>349</sup> *Id.* at 21.

<sup>350</sup> *Id.*

<sup>351</sup> *Id.* (internal citation omitted).

and the assumptions.”<sup>352</sup> Dr. Cohen’s text simply does not support Commerce’s position as Dr. Cohen “defines {each threshold} in terms of the U measures such that the interpretive power of *d* cannot be understood as anything other than a measure of nonoverlap.”<sup>353</sup> Dr. Cohen did provide “real-world examples” to illustrate each threshold, each threshold itself is defined by the U measures, *e.g.*, the percent nonoverlap.<sup>354</sup> Further, the “real-world examples” used by Dr. Cohen themselves satisfy the statistical criteria.<sup>355</sup> Thus, Commerce has “cherry-picked certain statements” and failed to consider the record as a whole.<sup>356</sup> In *Stupp*, the Federal Circuit observed “that ‘{*i*}n developing {the interpretive} cutoffs, including the 0.8 cutoff, Professor Cohen noted that “we maintain the assumption that the populations being compared are normal and with equal variability and conceive them further as equally numerous.””<sup>357</sup> Thus, Commerce’s explanation is also contrary to the understanding of the Federal Circuit. Commerce itself has even recognized the relationship between the percent nonoverlap and the Cohen’s *d* coefficient.<sup>358</sup> Therefore, Commerce’s reasoning that Dr. Cohen’s thresholds are independent of the statistical criteria and the U measures is unreasonable, and Commerce must reconsider its application of the Cohen’s *d* test in the final results of redetermination.<sup>359</sup>

Third, Commerce contends that the Federal Circuit misunderstood the academic literature discussed in *Stupp*,<sup>360</sup> and that none of the academic literature requires that Commerce observe

---

<sup>352</sup> *Id.* at 22 (internal citation omitted).

<sup>353</sup> *Id.* (citing *Cohen* at 22-27).

<sup>354</sup> *Id.*

<sup>355</sup> *Id.*

<sup>356</sup> *Id.*

<sup>357</sup> *Id.* at 22-23 (quoting *Stupp*, 5 F.4th at 1357 (quoting *Cohen* at 21)) (emphasis added by Canadian Parties).

<sup>358</sup> *Id.* at 23 (citing *Certain Frozen Warmwater Shrimp from the Socialist Republic of Vietnam: Final Results of Antidumping Duty Administrative Review, 2014–2015*, 81 FR 62717 (September 12, 2016) (*Shrimp from Vietnam*), and accompanying IDM at 9; and *Certain Frozen Warmwater Shrimp from India: Final Results of Antidumping Duty Administrative Review, 2012–2013*, 79 FR 51309 (August 28, 2014) (*Shrimp from India*), and accompanying IDM at 24).

<sup>359</sup> *Id.* at 23-24.

<sup>360</sup> *Id.* at 24.

the statistical criteria. However, the academic literature does undermine Commerce's assertion that it need not observe the statistical criteria. The Federal Circuit pointed to Drs. Grissom and Kim that Dr. Cohen did derive his thresholds based on the statistical criteria:

When the distribution of scores of a comparison population is not normal, the usual interpretation of a  $d_G$  or {Cohen's}  $d$  in terms of estimating the percentile standing of the average-scoring members of another group with respect to the supposed normal distribution of the comparison group's scores would be invalid. Also, because standard deviations can be very sensitive to a distribution's shape, ... nonnormality can greatly influence the value of a standardized-mean-difference effect size and its estimate.<sup>361</sup>

Commerce insists that the statement from Grissom does not relate to Dr. Cohen's thresholds but only impacts the calculation of certain U measures:

As discussed above, Professor Cohen defined the thresholds in terms of the U measures and with the understanding that the assumptions would be satisfied. That conclusion is reinforced by the fact that the 'real-world examples,' which illustrate Professor Cohen's thresholds, involve data that conform to the assumptions. Professor Cohen did not, as {Commerce} avers, create the thresholds with the understanding that they would have the same meaning regardless of whether the assumptions were satisfied.<sup>362</sup>

Commerce dismisses the second passage from Drs. Grissom and Kim as "'an alternative approach' to calculate the denominator of the  $d$  coefficient when dealing with sampling."<sup>363</sup>

However, Grissom does not "refer to estimates and samples (as well as to populations)," which are not relevant in any case. In fact, Drs. Grissom and Kim recommend "the use of a single population standard deviation (or an estimate thereof) as the denominator ... because of the sensitivity of the coefficient to the violations of the assumption of equal variances."<sup>364</sup> This sensitivity is relevant whether a population or a sample is used.<sup>365</sup>

---

<sup>361</sup> *Id.* (quoting *Grissom* at 66).

<sup>362</sup> *Id.* at 25 (internal citation omitted).

<sup>363</sup> *Id.* at 25-26 (citing *Grissom* at 68; and Draft Redetermination at 31).

<sup>364</sup> *Id.* at 26 (citing *Grissom* at 66).

<sup>365</sup> *Id.* (citing *Grissom* at 68).

Professor Coe also explains that a measure of effect size is sensitive to the assumption that the data exhibit a normal distribution.<sup>366</sup> Commerce states that this concern only applies to sampled data, however, “{Professor Coe’s} explanation applies mathematically ... regardless of whether those groups are populations or samples.”<sup>367</sup>

Dr. Li explained “that violating the assumptions of normality and roughly equal variances ‘severely affect{s} the accuracy of  $d$  in evaluating the true {effect size}.”<sup>368</sup> Commerce dismisses Dr. Li’s statement because the Cohen’s  $d$  test uses the full populations of data and Dr. Li’s “concern applies only to estimating the  $d$  coefficient using samples of data.”<sup>369</sup> However, “estimates have nothing to do with the issue” because a “given value of { $a$ } Cohen’s  $d$  {coefficient} that is interpreted when the assumptions are satisfied cannot be interpreted to mean the same thing when the assumptions are violated.”<sup>370</sup>

For other situations, Commerce argues that when the statistical criteria are not met, the Cohen’s  $d$  coefficient is smaller than if the statistical criteria had been met, thus making it less likely that the difference in prices would be found to be significant.<sup>371</sup> Here, Commerce is wrong that the academic literature “demonstrates a systematic tendency for {the} Cohen’s  $d$  {coefficient} to understate the effect size when the assumptions are violated.”<sup>372</sup> For a given group of data, where the difference in the means (*i.e.*, the numerator) is the same but the distributions differ, then where the standard deviation of the non-normal distribution is larger than the standard deviation of the normal distribution, the Cohen’s  $d$  coefficient will be smaller, and *vice versa*; this is simply a function of the fact that the standard deviation is the denominator

---

<sup>366</sup> *Id.* (citing *Coe* at 14).

<sup>367</sup> *Id.* at 27.

<sup>368</sup> *Id.* (quoting *Li* at 1560).

<sup>369</sup> *Id.* (internal citation omitted).

<sup>370</sup> *Id.*

<sup>371</sup> *Id.* at 27-28.

<sup>372</sup> *Id.* at 28.

of the Cohen's  $d$  coefficient.<sup>373</sup> Although the examples from *Coe*, *Algina*, and *Li* represent "violations" that result in a smaller  $d$  coefficient,<sup>374</sup> the examples from Grissom and Kim show a larger  $d$  coefficient with smaller sample size.<sup>375</sup>

The Canadian Parties argue that Commerce's fundamental error is that "the {academic} literature discussed in *Stupp* demonstrates that when a {Cohen's}  $d$  coefficient is calculated using data that violate the assumptions of normality, equal variance, and equal size, the {Cohen's}  $d$  coefficient will not describe the degree of nonoverlap in the same way that {Dr. Cohen} describes the nonoverlap measures underlying the thresholds for small, medium, and large effect sizes."<sup>376</sup> They argue that "as the Federal Circuit recognized in *Stupp*, 'violating these assumptions can subvert the usefulness of the interpretive cutoffs, transforming what might be a conservative cutoff into a meaningless comparator.'"<sup>377</sup> Thus, they claim that "for these reasons, {Commerce} must reconsider its application of {its} Cohen's  $d$  {test} in this {investigation}."<sup>378</sup>

Fourth, in response to the Federal Circuit's hypothetical example, Commerce states that its use of the Cohen's  $d$  test would not result in an upward bias because it uses full populations and not sampled data.<sup>379</sup> Further, Commerce states that even if there are results that are "false positive" (*i.e.*, Commerce finds that the difference in prices is significant when it is not), subsequent steps in the differential pricing analysis will "launder such flaws" such that the differential pricing analysis, as a whole, is reasonable.<sup>380</sup> However, the Canadian Parties argue

---

<sup>373</sup> *Id.*

<sup>374</sup> *Id.* at 28-29 (citing Draft Redetermination at 32-34).

<sup>375</sup> *Id.* at 29 (citing *Grissom* at 68-70).

<sup>376</sup> *Id.*

<sup>377</sup> *Id.* (quoting *Stupp*, 5 F.4th at 1360).

<sup>378</sup> *Id.*

<sup>379</sup> *Id.* (citing Draft Redetermination at 39-40).

<sup>380</sup> *Id.* at 29-30 (citing Draft Redetermination at 40-43).

that Commerce’s explanation fails to address the Federal Circuit’s concerns that the results of the Cohen’s *d* test in its hypothetical example is “problematic.”<sup>381</sup> The Federal Circuit posited that, as the variances in the prices approach zero, the Cohen’s *d* coefficient increases and creates “an ‘artificially inflate {d} ... dumping margin.’”<sup>382</sup> “As a result, {Commerce’s} methodology would indicate a large effect size when clearly the prices between the two groups differ very little.”<sup>383</sup> The Canadian Parties argue that “the Federal Circuit also noted that ‘requiring larger test groups tends to decrease the likelihood that a test group would have sales prices with near-zero variance, and requiring normality also tends to decrease that likelihood as the number of observations increases.’”<sup>384</sup> Commerce’s response, that its use of populations eliminates any potential bias, does not address the Federal Circuit’s concern, *i.e.*, “whether violating the assumptions produces a *d* coefficient of 0.8 or greater when there is a small difference between the groups.”<sup>385</sup> They claim that Commerce’s explanation “*assumes* that violating the assumptions can produce a *d* that accurately indicates the effect size” as long as the full populations are used.<sup>386</sup> The Federal Circuit was aware of Commerce’s full-population explanation, and that the data in its hypothetical example are full populations; “{n}evertheless, the Federal Circuit was not persuaded that using full populations addressed its concerns about violating the assumptions.”<sup>387</sup> Commerce even constructs a “more ‘extreme’” example where there is no variance in prices within each group which would produce a significant difference, however, “the difference between the two prices is clearly not significant.”<sup>388</sup>

---

<sup>381</sup> *Id.* at 30 (citing *Stupp*, 5 F.4th at 1359).

<sup>382</sup> *Id.* (quoting *Stupp*, 5 F.4th at 1359).

<sup>383</sup> *Id.*

<sup>384</sup> *Id.* (quoting *Stupp*, 5 F.4th at 1359).

<sup>385</sup> *Id.* at 31.

<sup>386</sup> *Id.* (emphasis in original).

<sup>387</sup> *Id.*

<sup>388</sup> *Id.* at 32 (internal citations omitted).

Further, Commerce claims that the meaningful difference test will resolve the concern raised by the Federal Circuit. Commerce explains that there are five scenarios that describe the relationship between U.S. price and normal value, and only the fifth scenario will result in a meaningful difference.<sup>389</sup> Commerce asserts that its “more extreme” example would not result in a meaningful difference.<sup>390</sup> However, this is “yet another red herring.”<sup>391</sup> Commerce concedes that “its application of the Cohen’s *d* test in its hypothetical would indicate a large effect size when the two groups of prices do not differ significantly.”<sup>392</sup> Therefore, Commerce’s Cohen’s *d* test does not measure the significance of the difference in prices, as the statute requires, such that Commerce must reconsider its use of the Cohen’s *d* test.<sup>393</sup>

Commerce’s differential pricing analysis includes three distinct steps: (1) the Cohen’s *d* test; (2) the ratio test; and (3) the meaningful difference test. Each test “serves a distinct function” and they “do not operate as an amalgamated whole.”<sup>394</sup> If the Cohen’s *d* test fails to determine whether the prices differ significantly when the statistical criteria are not satisfied, then the subsequent tests cannot perform their intended purposes.<sup>395</sup>

The Canadian Parties argue that, accordingly, “{Commerce’s} defense that subsequent steps of the {differential pricing analysis} launder flaws in the Cohen’s *d* test is unreasonable, and {Commerce} must reconsider its application of the Cohen’s *d* test for the {final results of redetermination}.”<sup>396</sup>

---

<sup>389</sup> *Id.* at 32-33 (citing *Certain Hot-Rolled Steel Flat Products from Japan: Final Determination of Sales at Less Than Fair Value and Final Affirmative Determination of Critical Circumstances*, 81 FR 53409 (August 12, 2016), and accompanying IDM at 32).

<sup>390</sup> *Id.* at 33.

<sup>391</sup> *Id.*

<sup>392</sup> *Id.* at 33-34.

<sup>393</sup> *Id.* at 34.

<sup>394</sup> *Id.* at 35.

<sup>395</sup> *Id.*

<sup>396</sup> *Id.* at 36.

Fifth, Resolute FP states that Commerce has used the differential pricing analysis “and found targeted dumping,” and applied an alternative comparison method in 145 of 165 investigations and reviews during the POI of the *Final Determination*.<sup>397</sup> The “alternative {comparison} methods are supposed to be the ‘exception’ to the preferred {A-to-A} method according to the statute.”<sup>398</sup> The frequency with which Commerce resorts to an alternative comparison method “demonstrates that the ‘exception’ has impermissibly swallowed the rule.”<sup>399</sup> Commerce claims that “exception” does not need to be rare or unusual, rather, that the alternative comparison methods are but one of two comparison methodologies, depending upon the circumstances, used to calculate a weighted-average dumping margin. “Additionally, {Commerce} asserts that the figures provided by {Resolute FP} fail to take into account the number of times that {Commerce} does not actually calculate margins for respondent,” such as when Commerce relies on total adverse facts available.<sup>400</sup> However, Resolute argues that Commerce ignores the plain meaning of the word “exception” in that the rule does not generally apply, and given the text of the statute, Congress’ intent is clear, that Commerce should generally use the A-to-A method, and use an alternative comparison method “sparingly.”<sup>401</sup> Commerce does not dispute Resolute FP’s analysis that Commerce used an alternative comparison method in 87 percent of the investigations and reviews during the POI. Commerce finds “purported targeted dumping” with such frequency because its Cohen’s *d* test “does not meaningfully identify significant price differences ... when the assumptions are violated.”<sup>402</sup>

---

<sup>397</sup> *Id.* (citing Rule 57(3) Reply Brief of Resolute FP Canada Inc. (October 2, 2018) (Resolute FP Rule 57(3) Reply Brief), at 22–23 and Appendix B).

<sup>398</sup> *Id.* at 36–37 (citing section 777A(d)(1)(A) of the Act; and 19 CFR 351.414).

<sup>399</sup> *Id.* at 37.

<sup>400</sup> *Id.*

<sup>401</sup> *Id.*

<sup>402</sup> *Id.* at 38.

### 3. Commerce's Use of a Simple Average to Calculate the Denominator of the Cohen's *d* Coefficient is Unreasonable

In responding to the Binational Panel Order to address the Federal Circuit's opinion in *Mid Continent II*, the Canadian Parties state that Commerce explained that using a simple average is supported by the academic literature itself when using full populations when the standard deviations of the test and comparison groups differ, or, alternatively, by the fact that Commerce uses the full populations such that the standard deviations of the two populations are 100 percent reliable, and, therefore each should be equally weighted.<sup>403</sup> Commerce's "explanations are unreasonable and must be reconsidered."<sup>404</sup>

First, Commerce incorrectly concludes that the academic literature supports using Dr. Cohen's equation 2.3.2 when using full populations,<sup>405</sup> "even if the assumptions are violated."<sup>406</sup> Commerce mischaracterizes the requirement of equal sample sizes "as a proxy for determining the reliability of each standard deviation,"<sup>407</sup> which is then the basis for "determining how to weigh the standard deviations."<sup>408</sup> Commerce supports this conclusion with Dr. Algina *et al.*'s description of the standard practice concerning the use of Greek and Latin variables. "None of this reasoning makes any sense."<sup>409</sup>

According to the Canadian Parties, Commerce claims that the academic literature supports the use of Dr. Cohen's equation 2.3.2 when full populations are known, and that Commerce "acknowledges that Professor Cohen assumes that the sample sizes would be equal in equation 2.3.2, but insists 'Dr. Cohen does not apply the limitation of equal sample sizes, *i.e.*,

---

<sup>403</sup> *Id.* at 39.

<sup>404</sup> *Id.*

<sup>405</sup> *Id.* at 38, footnote 125 (citing *Cohen* at 43-44).

<sup>406</sup> *Id.* at 39.

<sup>407</sup> *Id.* at 39-40.

<sup>408</sup> *Id.* at 40.

<sup>409</sup> *Id.*

$n_A = n_B$ , in his description of equation 2.3.2.”<sup>410</sup> Instead, Commerce reduces sample size to a determinant of reliability of the sampled data, as confirmed by Dr. Cohen’s equations 2.5.1 and 2.5.2, which Commerce concludes are the equations used by Dr. Cohen to calculate the denominator of the  $d$  coefficient for sampled data.<sup>411</sup> However, “the Federal Circuit has already rejected {Commerce’s} assertion that the literature suggests simple averaging when sample sizes are unequal.”<sup>412</sup> Further, Commerce’s logic that equal sample sizes results in equal reliability which results in equal weighting is incorrect. The reason for a simple average when sample sizes are equal is simply that it is a weighted average with equal weights.<sup>413</sup> “*This is just arithmetic.*”<sup>414</sup>

The Canadian Parties note that Dr. Cohen does discuss both reliability and equal group sizes, but “Professor Cohen’s discussion of the relationship between reliability and sample size has nothing to do with the assumption of equal group size when calculating {the} Cohen’s  $d$  {coefficient}.”<sup>415</sup> Reliability describes the closeness of a sample value to approximate a population value, and, as discussed below, reliability “has nothing to do with {Commerce’s} use of {the} Cohen’s  $d$  {test}.”<sup>416</sup> “Simply put, {Commerce’s} use of simple averaging when the assumptions of equal size and equal variances are violated creates a denominator that does not do what Professor Cohen said it should do.”<sup>417</sup>

Further, the CIT noted that Dr. Cohen’s equation 2.3.2 is part of a power analysis, in which a test involving full populations “would be redundant.”<sup>418</sup> The Canadian Parties argue

---

<sup>410</sup> *Id.* at 40-41 (quoting Draft Redetermination at 55).

<sup>411</sup> *Id.* at 41.

<sup>412</sup> *Id.* (citing *Mid Continent II*, 5 F.4th at 1380).

<sup>413</sup> *Id.* at 41-42.

<sup>414</sup> *Id.* at 42 (emphasis in original).

<sup>415</sup> *Id.* (internal citations omitted).

<sup>416</sup> *Id.* (citing *Cohen* at 6).

<sup>417</sup> *Id.*

<sup>418</sup> *Id.* (quoting *Mid Continent III*, 628 F. Supp. 3d at 1324).

that Commerce has failed to explain how its “‘populations’ rationale” is reasonable in this situation.<sup>419</sup> The CIT also rejected Commerce’s “‘populations’ rationale” based on Dr. Cohen’s “warning ‘that power values “may be greatly in error” if both sample sizes and standard deviations are unequal.’”<sup>420</sup> Commerce dismisses the CIT’s opinion, stating that Dr. Cohen’s concern relates to a power analysis and not to calculating the  $d$  coefficient itself; however, this does not change the fact that Dr. Cohen assumes equal sample sizes when using the simple average of equation 2.3.2.<sup>421</sup>

The Canadian Parties claim that Commerce’s claim that Dr. Cohen’s equations 2.5.1 and 2.5.2 are used for sampled data misunderstands the academic literature.<sup>422</sup> These equations are used to calculate  $d_s$ , “which is calculated using experimental data to detect ‘the palpable characteristics of the sample and their bearing on the null hypothesis.’”<sup>423</sup> On the other hand, the Cohen’s  $d$  coefficient is one of the four variables that are part of a power analysis, and Dr. Cohen does not “substitute equation 2.3.2 with equation 2.5.2 when using sample data because the two equations perform distinct functions.”<sup>424</sup> “{W}hile Professor Coe and Dr. Ellis ‘prescribe equation (2.5.2) for situations where size is being calculated from experimental data,’ neither ‘discusses using a simple unweighted average.’”<sup>425</sup>

Additionally, Commerce asserts that Dr. Cohen’s equation 2.3.2 is specific to the use of full populations based on the notion, from *Algina*, that Greek letters denote a population variable.<sup>426</sup> Commerce’s assertion that Dr. Cohen used such a convention in equation 2.3.2 is

---

<sup>419</sup> *Id.* at 43.

<sup>420</sup> *Id.* at 43 (quoting *Mid Continent III*, 628 F. Supp. 3d at 1325 (quoting *Cohen* at 44); and citing Draft Redetermination at 56 (quoting the same)).

<sup>421</sup> *Id.*

<sup>422</sup> *Id.* (citing *Cohen* at 66-67).

<sup>423</sup> *Id.* at 43-44 (quoting *Cohen* at 66).

<sup>424</sup> *Id.* at 44.

<sup>425</sup> *Id.* (quoting *Mid Continent III*, 628 F. Supp. 3d at 1325-26 (citing *Coe* at 10; and *Ellis* at 10, footnote 8)).

<sup>426</sup> *Id.*

without merit. In fact, Dr. Algina *et al.*'s statement demonstrates that Dr. Cohen did not follow this convention and used Latin letters to denote population values.

Irrespective of the nomenclature used by Dr. Cohen, the Canadian Parties argue that for Commerce to use the thresholds described by Dr. Cohen, whether it is using “populations, samples, or anything else,” Commerce is calculating a denominator of the effect size different from what is provided in the academic literature. “The literature does not support {Commerce’s} reliance on equation 2.3.2 to calculate the denominator when the assumptions of equal variances and equal size are violated.”<sup>427</sup>

Additionally, Commerce “fundamentally misunderstands the concept of reliability,” which “characterizes the degree to which a value in one group of data approximates the same value in another group of data.”<sup>428</sup> Commerce misuses reliability “to evaluate how closely a value in a data set resembles itself in that same dataset,” and it “cannot compensate for failing to observe the assumptions in calculating the denominator {of the Cohen’s *d* coefficient}.”<sup>429</sup> Commerce asserts that when the sample sizes are equal, the standard deviations of the sampled data are equally reliable, and, therefore, equal weights, *i.e.*, a simple average, is appropriate.<sup>430</sup> Commerce “is confusing different concepts that have nothing to do with its Cohen’s *d* test and relying on a CIT decision that did not address that disconnect.”<sup>431</sup> In *Mid Continent IV*, the CIT found that Commerce “‘provided an explanation that logically connects the relevance of full populations to the use of simple averaging’ in its reliability analysis.”<sup>432</sup> However, the CIT

---

<sup>427</sup> *Id.* at 45.

<sup>428</sup> *Id.* at 40.

<sup>429</sup> *Id.*

<sup>430</sup> *Id.* at 45-46 (citing Draft Redetermination at 60-61).

<sup>431</sup> *Id.* at 46 (citing *Mid Continent IV*).

<sup>432</sup> *Id.* (quoting *Mid Continent IV*, 680 F. Supp. 3d at 1353).

holding “that it is appropriate to use a simple average for equal sample sizes because the two samples have equal reliability”<sup>433</sup> is “both false and irrelevant.”<sup>434</sup>

The academic literature does not support that simple averaging is the result of “different sample sizes that are equally reliable.”<sup>435</sup> As discussed above, simple averaging is simply the result of a weighted average with equal sample sizes. Further, Commerce’s reliance on reliability is misplaced, where reliability “describes the degree to which a value in a subset of data resembles that same value in the larger set of data (*i.e.*, the data set from which the subset was selected).”<sup>436</sup> Commerce’s Cohen’s *d* test “has nothing to do with drawing inferences about an unknown population of data from a known sample.”<sup>437</sup> Dr. Cohen did not present his measures of effect size in the context of reliability of the standard deviations, and the fact that Commerce asserts that the reliability of the standard deviations in its Cohen’s *d* test are 100 percent reliable is meaningless.<sup>438</sup> In addition, “{Commerce’s} reliability rationale provides no information that would entitle the standard deviations of the test and comparison groups to equal weight.”<sup>439</sup> The Canadian Parties argue that “{t}he weight assigned to the standard deviations in equation 2.3.2 must be based on two considerations: whether the standard deviations are equal and whether the group sizes are equal.”<sup>440</sup> “If the standard deviations are equal, then there is no need to perform a separate denominator calculation—the denominator is just the common standard deviation.”<sup>441</sup> “If the groups are of equal size, then there is no need to calculate a weighted average standard deviation—the two standard deviations already carry equal

---

<sup>433</sup> *Id.* (quoting *Mid Continent IV*, 680 F. Supp. 3d at 1354-55).

<sup>434</sup> *Id.*

<sup>435</sup> *Id.*

<sup>436</sup> *Id.* at 47 (citing *Cohen* at 6).

<sup>437</sup> *Id.*

<sup>438</sup> *Id.*

<sup>439</sup> *Id.*

<sup>440</sup> *Id.* at 48 (citing *Cohen* at 43-44).

<sup>441</sup> *Id.* (citing *Cohen* at 20).

weight.”<sup>442</sup> If neither condition is satisfied, then Commerce’s approach will not be consistent with the basis “on which the thresholds are based, or in accordance with the literature.”<sup>443</sup>

Therefore, Commerce must reconsider its use of a simple average.

#### 4. Commerce May Properly Account for Seasonality Using Shorter Averaging Periods

The Cohen’s *d* test identifies significant price differences between time periods where “goods that exhibit price volatility due to seasonal market conditions or simply price variability over time.”<sup>444</sup> The results of Commerce’s Cohen’s *d* test “produced the overwhelming majority of ‘passing’ values across periods of time.”<sup>445</sup> Commerce’s meaningful difference test fails to “consider whether ‘such differences’ can be taken into account using the {A-to-A} method adjusted for time periods.”<sup>446</sup> Commerce’s own regulations provide the remedy to account for seasonality in resorting to “time periods that account for price differences over time.”<sup>447</sup> With the A-to-A method, Commerce is not required to use period-wide averages, but may use shorter averaging periods such that the A-to-A method would account for the price differences by time period. Given the flexibility afforded Commerce, “it is difficult to see how {Commerce} would ever be unable to account for differences among periods of time.”<sup>448</sup>

#### *Resolute FP’s Comments*

##### 1. The Cohen’s *d* Test and the Statistical Criteria

Resolute FP argues that Commerce, contrary to the academic literature, has failed to observe the statistical criteria, *i.e.*, normal distribution, sufficient size and roughly equal

---

<sup>442</sup> *Id.* (citing *Cohen* at 43-44).

<sup>443</sup> *Id.*

<sup>444</sup> *Id.* at 35.

<sup>445</sup> *Id.* (citing Resolute FP Rule 57(3) Reply Brief at 22–23, Appendix C).

<sup>446</sup> *Id.*

<sup>447</sup> *Id.* at 36 (citing 19 CFR 351.404(d)(3)).

<sup>448</sup> *Id.*

variances, which govern the Cohen's *d* test.<sup>449</sup> Without Commerce's use of the Cohen's *d* test, Resolute FP's estimated weighted-average dumping margin would be *de minimis*, and Resolute FP would not be subject to the *Order*.<sup>450</sup> The law has evolved significantly since Commerce's *Final Determination*, and there continues to be litigation before the CIT, the Federal Circuit, and binational panels.<sup>451</sup> This Panel appropriately understood these circumstances, and remanded this issue "for an explanation of whether the limits on the use of the Cohen's *d* test were met in this case or whether those limits need not be followed by Commerce ... {and,} “{i}n this regard, Commerce is invited to clarify its argument concerning availability of the full universe of data.”<sup>452</sup> The Panel's instructions were simple, yet Commerce has again “complicated the issue.”<sup>453</sup>

Contrary to Commerce's explanation, Resolute FP argues that the academic literature “demonstrates that there is no viable excuse for rejecting the underlying assumptions of the Cohen's *d* test.”<sup>454</sup> The Cohen's *d* test uses a measure of effect size to determine whether prices differ significantly. Resolute FP argues that “Dr. Cohen designed his test using two datasets of equal size having normal (*i.e.*, bell-shaped) distributions and equal variances.”<sup>455</sup> It notes that “Dr. Cohen stated, ‘{i}f we maintain the assumption that the populations being compared are normal and with equal variability, and conceive them further as equally numerous,’ then his equation can be used to derive a meaningful measurement of effects size.”<sup>456</sup> Thus, a researcher can see the percent nonoverlap between the two groups of data which meet the statistical criteria,

---

<sup>449</sup> See Resolute FP Comments on the Draft Redetermination at 19 (internal citation omitted).

<sup>450</sup> *Id.*

<sup>451</sup> *Id.* (internal citations omitted).

<sup>452</sup> *Id.* at 20 (quoting Binational Panel Report at 30).

<sup>453</sup> *Id.*

<sup>454</sup> *Id.* at 21.

<sup>455</sup> *Id.* (citing *Cohen* at 27).

<sup>456</sup> *Id.* (quoting *Cohen* at 21).

and can gauge the significance of the difference between the two groups.<sup>457</sup> Further, Resolute FP argues that “Dr. Cohen, based on the results of his modeling, derived interpretive cutoffs of effect size.”<sup>458</sup> “Despite any arbitrariness, the thresholds adopted by Dr. Cohen can measure the nonoverlap of datasets when the appropriate conditions for the analysis are met.”<sup>459</sup>

Resolute FP states that Commerce cites no authority for the proposition that the statistical criteria are not relevant simply because the Cohen’s *d* test is based on the full populations of U.S. prices. Dr. Cohen discusses populations and does not limit the three statistical criteria to only sampled data.<sup>460</sup> It claims that Commerce’s analysis is meaningless when the statistical criteria are disregarded. Further, even though the Federal Circuit found that Dr. Cohen’s thresholds are “widely accepted,” the Federal Circuit still remanded this issue to Commerce to further explain whether the statistical criteria need be observed.<sup>461</sup> Commerce notes that Dr. Cohen’s thresholds are “illustrated by real-world observations and experience,” but Resolute FP states that it continues to defy “the logic that underpins {Dr. Cohen’s work}.”<sup>462</sup> It argues that Dr. Cohen’s discussion “is anything but divorced from the underlying {statistical criteria}.” Dr. Cohen’s presentation of the “real-world, easy-to-understand examples” comes “after his discussion of the *d* coefficient and the illustrations of {non}overlap ({i.e.}, ‘U measures’).”<sup>463</sup> There is no reason for Commerce to think that Dr. Cohen ignored the preceding analysis, including the statistical criteria to derive his small, medium, and large thresholds. Commerce’s explanation is simply not supported by the academic literature.

---

<sup>457</sup> *Id.*

<sup>458</sup> *Id.* at 22 (citing *Cohen* at 24-27).

<sup>459</sup> *Id.*

<sup>460</sup> *Id.*

<sup>461</sup> *Id.* at 23 (internal citation omitted).

<sup>462</sup> *Id.* (internal citation omitted).

<sup>463</sup> *Id.* (citing *Cohen* at 21).

According to Resolute FP, Commerce addresses the academic literature cited by the Federal Circuit in *Stupp* only in the context of its flawed analysis.<sup>464</sup> Its argument focuses on the claim that the academic literature relates to sampled data and its own analysis is based on full populations. Further, Commerce claims again that Dr. Cohen’s thresholds are not based on the statistical criteria and are only based on “real-world experience.”<sup>465</sup> Resolute FP states that Commerce’s claim to distinguish between the academic literature and its use of the Cohen’s *d* test is “a distinction without a difference” and is not supported by the academic literature.<sup>466</sup>

2. The Average-to-Transaction Method Is an Exception to the Average-to-Average Method and Should Be Applied Rarely

Resolute FP argues that Commerce incorrectly argues that the A-to-A method is not limited to “only rare or unusual cases.”<sup>467</sup> It claims that the statute provides two standard comparison methodologies – the A-to-A method and the T-to-T method – which are preferred over the alternative A-to-T method<sup>468</sup> and “the SAA emphasizes that the {A-to-T} method must be used with restraint.”<sup>469</sup> The SAA states that “Commerce ‘must establish and provide an explanation why it cannot account for such differences through the use of {an} {A-to-A} or {T-to-T} comparison{ }’ and that ‘Commerce will proceed on a case-by-case basis, because small differences may be significant for one industry or one type of product, but not for another.’”<sup>470</sup>

Resolute claims that the statute is unambiguous—the A-to-A method is the default comparison method and the A-to-T method may be used to only address “targeted dumping as an uncommon practice when the statutory conditions are met.”<sup>471</sup> In the Draft Redetermination,

---

<sup>464</sup> *Id.* at 24.

<sup>465</sup> *Id.*

<sup>466</sup> *Id.*

<sup>467</sup> *Id.* at 25.

<sup>468</sup> *Id.*

<sup>469</sup> *Id.*

<sup>470</sup> *Id.* (quoting the SAA at 843).

<sup>471</sup> *Id.*

Commerce provided “two examples” where it tries to demonstrate that the A-to-T method is “rarely used.”<sup>472</sup> However, Commerce’s information is flawed because it includes results based on total adverse facts available (AFA) and “did not attempt to distinguish and eliminate cases where it found insufficient sales to conduct a targeted dumping analysis.”<sup>473</sup> Commerce “also ignores administrative reviews, making the sample size too small to be statistically significant.”<sup>474</sup> However, in its Rule 57(3) Brief, Resolute FP provided information to the Panel which demonstrates that Commerce finds “targeted dumping in 145 of 165 investigations and reviews (87%) conducted during the period of investigation of this {investigation} (October 2015–September 2016), making {the A-to-T method} the norm rather than an exception.”<sup>475</sup> Commerce “has not disputed the accuracy of those figures.”<sup>476</sup>

Resolute FP does not claim that Commerce cannot test for “targeted dumping” in all investigations and reviews, but if Commerce’s test is “accurate and properly used,” then it must strictly follow the statutory requirements before using the “alternative and less-favored” A-to-T method. However, the differential pricing analysis is flawed and fails to meet this standard. Further, the differential pricing analysis “ignores the SAA’s caution that ‘small differences may be significant for one industry or one type of product, but not for another’ which, in this {investigation}, is the seasonality of lumber.”<sup>477</sup>

### 3. Use of the Simple Average Is Indefensible

Resolute FP argues that the academic literature provides no support for Commerce’s use of a simple average of the standard deviations of the test and comparison groups as the

---

<sup>472</sup> *Id.*

<sup>473</sup> *Id.* at footnote 118.

<sup>474</sup> *Id.*

<sup>475</sup> *Id.* at 26.

<sup>476</sup> *Id.*

<sup>477</sup> *Id.* (quoting the SAA at 843).

denominator of the Cohen's  $d$  coefficient. That denominator "is generally defined as the 'common within-population standard deviation' ( $\sigma$ ), assuming equal variances, equal size, and normal distribution."<sup>478</sup> It claims that when "normally distributed populations of equal size {have} unequal variances," then Dr. Cohen provides for a simple average.<sup>479</sup> Further, it argues that "Dr. Cohen notes that where variances and sample sizes are different, his table of estimates (Table 2.3) of power values 'may be greatly in error.'"<sup>480</sup> It also claims that "when the two groups being compared are of unequal size, scholarly literature generally requires the use of a weighted average."<sup>481</sup>

The Federal Circuit also stated in *Mid Continent II* that Commerce had not used the single standard deviation "even while recognizing that it had the full set of data for U.S. sales for the period Commerce was reviewing."<sup>482</sup> Therefore, Resolute FP states that the single standard deviation "fulfills Dr. Cohen's criterion for use of a single standard deviation:  $\sigma = \sigma_A = \sigma_B$ ."<sup>483</sup>

Instead, Resolute FP argues that Commerce continues to use the simple average which is unsupported by the academic literature: "Dr. Cohen was very specific about the conditions for using a simple average: 'normal populations of unequal variance' where each group being compared is the same size (*i.e.*,  $n_A = n_B$ )."<sup>484</sup> It claims that these conditions do not exist in Commerce's Cohen's  $d$  test, which Commerce continues to erroneously rely on a simple average.<sup>485</sup>

---

<sup>478</sup> *Id.* at 27 (citing *Cohen* at 27).

<sup>479</sup> *Id.* (citing *Cohen* at 43-44).

<sup>480</sup> *Id.* at 28 (quoting *Cohen* at 44).

<sup>481</sup> *Id.*

<sup>482</sup> *Id.* (quoting *Mid Continent II*, 31 F.4th at 1378).

<sup>483</sup> *Id.* at 28-29 (citing *Cohen* at 27).

<sup>484</sup> *Id.* at 29 (citing *Cohen* at 43-44).

<sup>485</sup> *Id.*

Resolute FP argues that Commerce’s alternative argument, based on “reliability,” is not supported by the academic literature.<sup>486</sup> It states that “the choice between simple average or weighted average is based on the value of the {standard deviations} and the sample size.”<sup>487</sup> In this investigation, the test and control groups generally are not the same size, have standard deviations which differ, and are not normally distributed. It claims, therefore that the academic literature requires a different approach to calculate the denominator of the Cohen’s *d* coefficient and argues that Commerce must not use a simple average if it continues to use the Cohen’s *d* test.<sup>488</sup>

**Commerce’s Position:**

We disagree with the arguments submitted by the Canadian Parties and Resolute FP that the Cohen’s *d* test is unreasonable or contrary to law. The courts have repeatedly sustained the Cohen’s *d* test.<sup>489</sup> Therefore, as in the *Final Determination*, Commerce continues to use the Cohen’s *d* test, as part of the differential pricing analysis, to examine whether U.S. prices differ significantly among purchasers, regions, or time periods pursuant to section 777A(d)(1)(B)(i) of the Act.

As explained in the Draft Redetermination, based on the academic literature, Dr. Cohen’s thresholds were not derived from any particular statistical criterion. Rather, Dr. Cohen proposed his thresholds as conventions. While he recognized that these conventions would be seen as “arbitrary” and perhaps susceptible to misunderstanding, Dr. Cohen believed that “the proposed

---

<sup>486</sup> *Id.*

<sup>487</sup> *Id.*

<sup>488</sup> *Id.* at 30.

<sup>489</sup> See, e.g., *Apex Frozen Foods Pvt. Ltd. v. United States*, 144 F. Supp. 3d 1308, 1327-28 (CIT 2016) (*Apex I*), *aff’d Apex Frozen Foods Pvt. Ltd. v. United States*, 862 F.3d 1337 (Fed. Cir. 2017) (*Apex II*); *Stupp Corp.*, 619 F. Supp. 3d 1314, appeal docketed Court No. 2023-1663 (March 27, 2023); *Marmen Inc.*, 627 F. Supp. 3d 1312, appeal docketed Court No. 2023-1877 (May 11, 2023); *NEXTEEL*, 676 F. Supp. 3d 1345; *SeAH*, 619 F. Supp. 3d at 1313.

conventions will be found to be reasonable by reasonable people.”<sup>490</sup> Moreover, the statistical criteria are not relevant to Commerce’s Cohen’s *d* test to consider whether the calculated parameters are representative of the actual parameters because the calculated parameters are not estimates based on sampled data, but rather are the actual parameter values based on the full populations of prices in each of the test and comparison groups. Lastly, the Federal Circuit’s concern about the results of the Cohen’s *d* test when there is a test group with a small number of observations and a very small variance in prices is not borne out by the data regarding the application of the Cohen’s *d* test in the context of Commerce’s differential pricing analysis.<sup>491</sup>

With respect to the calculation of the denominator of the Cohen’s *d* coefficient, Commerce explained in the Draft Redetermination how Dr. Cohen’s equation 2.3.2, *i.e.*, the simple average, applies to the calculation of the Cohen’s *d* coefficient for full populations where the sample size is not relevant. This is distinct from the situation where a weighted average is used when the calculation of the Cohen’s *d* coefficient is based on sampled data where the weighting is determined by the sample size in each group. Alternatively, the use of a simple average is also justified based on consideration of the equal reliability of parameters calculated using sampled data with equal sample sizes and parameters calculated based on full populations.

As an “exception” to one of the standard statutory comparison methodologies, Commerce’s reliance on the alternative A-to-T method is reasonable. Commerce is permitted to resort to an alternative comparison methodology when the statutory requirements of section 777A(d)(1)(B) of the Act have been examined and found to have been satisfied. Nothing in the

---

<sup>490</sup> See *Cohen* at 12-13.

<sup>491</sup> See, e.g., West Fraser’s “Response to Sections B, C and D of Initial Antidumping Duty Questionnaire,” dated March 21, 2017, at Exhibit C-11; Resolute’s “Response to Sections B-D of Initial Antidumping Questionnaire,” dated March 21, 2017, at Exhibit C-2; Canfor’s “Sections B-D Questionnaire Response,” dated March 23, 2017, at Exhibit C-1 (Canfor’s Questionnaire Response); Tolko’s “Sections B-D Questionnaire Response,” dated March 22, 2017, at Exhibit C-1.

statute, regulations, or legislative history beyond the statutory requirements dictates when or how Commerce may use the A-to-T method.

Further, nothing in the statute, regulations, or legislative history requires Commerce to use shorter averaging periods when comparing U.S. price with normal value. Commerce has used shorter price averaging periods, or limited the window for contemporaneous normal values, under certain circumstances, such as when using shorter cost-averaging periods. Finally, seasonality is a condition which may be accounted for when making comparisons of U.S. price with normal value, but seasonality is not part of an examination of whether prices differ significantly in the U.S. market.

1. The Statistical Criteria Need Not Be Observed as Part of the Cohen's *d* Test

In *Stupp*, the Federal Circuit held that “the evidence and arguments before us call into question whether Commerce’s application of the Cohen’s *d* test to the data in this case violated the assumptions of normality, sufficient observation size, and roughly equal variances associated with that test.”<sup>492</sup> Because this question could have a material impact on the results, the Federal Circuit remanded:

to give Commerce an opportunity to explain whether the limits on the use of the Cohen’s *d* test prescribed by Professor Cohen and other authorities were satisfied in this case or whether those limits need not be observed when Commerce uses the Cohen’s *d* test in less-than-fair-value adjudications. In that regard, we invite Commerce to clarify its argument that having the entire universe of data rather than a sample makes it permissible to disregard the otherwise-applicable limitations on the use of the Cohen’s *d* test.<sup>493</sup>

In the Draft Redetermination, Commerce explained further that the three statistical criteria cited by the Federal Circuit are not relevant to the Cohen’s *d* test. The Canadian Parties and Resolute FP contend that Commerce’s explanations are unsupported and unreasonable. For the reasons

---

<sup>492</sup> See *Stupp*, 5 F.4th at 1360.

<sup>493</sup> *Id.*

discussed below, we disagree, and continue to find that Commerce need not observe the statistical criteria identified by the Federal Circuit as part of the Cohen's *d* test.

As explained above, there are two purposes which the statistical criteria might be necessary for the calculation of effect size, specifically with respect to the Cohen's *d* coefficient: (1) to establish the reliability of estimated parameters based on sampled data to represent the actual parameters in a full population; and (2) to be able to calculate the U measures for given values of the Cohen's *d* coefficient, or to calculate the Cohen's *d* coefficient values for given values of the U measures. Resolute FP and the Canadian Parties continue to conflate these distinct purposes and improperly attempt to extend statistical limitations beyond their intended purposes which confuses the issue. For example, Resolute FP asserts that the denominator "is generally defined as the 'common within-population standard deviation' ( $\sigma$ ), assuming equal variances, equal size, and normal distribution."<sup>494</sup> However, Dr. Cohen's definition of " $\sigma$ " as the "common within-population standard deviation" is only limited to equal variances.<sup>495</sup> When Dr. Cohen discusses the use of his power tables,<sup>496</sup> this same assumption is valid, as Dr. Cohen includes the *d* coefficient here as presented originally at page 20. However, Dr. Cohen's power tables, for Case 0, also include the assumptions of normality and equal sample sizes, which apply to the use of the power tables,<sup>497</sup> yet Resolute FP projects these additional assumptions to also apply to the Cohen's *d* coefficient as originally formulated, which clearly is not supported by Dr. Cohen's presentation. This is but one example of where concepts and assumptions in the

---

<sup>494</sup> See, e.g., Resolute FP Comments on the Draft Redetermination at 27 (citing *Cohen* at 27).

<sup>495</sup> See *Cohen* at 20 (" $\sigma$  = the standard deviation of either population (since they are assumed equal).").

<sup>496</sup> *Id.* at 27-52.

<sup>497</sup> *Id.* at 27 ("CASE 0:  $\sigma_A = \sigma_B$ ,  $n_A = n_B$ . The power tables are designed to yield power values for the *t* test for the difference between the means of two independent samples of equal size drawn from normal populations having equal variances (Case 0).").

academic literature have been misrepresented to confuse an issue, here, the claim that the statistical criteria are required.

First, “when applying the Cohen’s *d* test, the three statistical criteria are relevant when using sampled data to ensure that the sample results reliably represent the full population of data, but they are not relevant here because the results of the Cohen’s *d* test are based on the full universe of prices in each test and comparison group and are not based on sampled data.”<sup>498</sup>

Certainly, when determining whether an estimated value based on a sample reliably represents the actual value of the population, characteristics such as the normality of the sample data and the sample size are important.<sup>499</sup> Violations of those assumptions (*i.e.*, normality, equal variances and sufficient sample size) could lead to distortions, including systemic bias, in the estimated value, and reduce the estimated value’s usefulness to represent the actual value of the parameter in the full population.<sup>500</sup> However, if the calculated value is based on the full population of data, then the calculated value is not an estimate and concerns related to the reliability of the calculated value do not exist. Thus, when the full population of data is known, the statistical criteria are not relevant to the reliability or accuracy of the calculated results and there is no sampling error in Commerce’s analysis of whether prices differ significantly.

As explained above, Commerce’s application of the Cohen’s *d* test is based on the full populations of price in each of the test and comparison groups.<sup>501</sup> The purpose of the differential pricing analysis is to determine whether one of the standard comparison methods provided for in section 777A(d)(1)(A) of the Act, is appropriate or whether an alternative comparison method is

---

<sup>498</sup> See “A Summary of The Statistical Criteria” section, *supra*.

<sup>499</sup> See *Cohen* at 6.

<sup>500</sup> See *Stupp*, 5 F.4th at 1358 (“There is extensive literature describing the problems associated with applying the Cohen's *d* test to data that are not normally distributed or that are lacking equal variances”)

<sup>501</sup> See “Full Population or Sampled Data” section, *supra*.

warranted pursuant to section 777A(d)(1)(B) of the Act once both the pattern and meaningful difference requirements have been met. Commerce's dumping analysis includes all U.S. sales during the POI or POR, and, thus, all the same U.S. sale prices are examined as part of the pattern requirement. For the pattern requirement, the statute directs Commerce to look at pricing patterns for comparable merchandise between purchasers, regions, or time periods. Thus, Commerce creates test groups which include all U.S. sale prices of comparable merchandise to a given purchaser, region, or time period, and comparison groups, which include all other U.S. sale prices of comparable merchandise. The mean and standard deviation calculated for each test and comparison group, as well as the effect size comparing the difference in prices between the two groups, are the actual values and not estimated parameters. Therefore, for Commerce's Cohen's *d* test, which determines if the prices differ significantly, the reliability of the calculated results to estimate the actual parameter values, including the Cohen's *d* coefficient, is not relevant, and, thus, the statistical criteria are not relevant to this aspect of the Federal Circuit's opinion in *Stupp*.

This distinction between estimated values based on sampled data and actual values based on full populations of data is supported in the academic literature. Dr. Ellis states that “{t}he best way to measure an effect is to conduct a census of an entire population but this is seldom feasible in practice {and}... may not even be desirable if researchers can identify samples that are representative of broader populations and then use inferential statistics to determine whether sample-based observations reflect population-level parameters.”<sup>502</sup> Dr. Ellis compares and contrasts two distinct methods of measuring the effect size. The first method is more accurate, as it results in the actual value and not an estimated value of the effect size, but it is less feasible in

---

<sup>502</sup> See *Ellis* at 5.

practice—it requires examination of the entire population and does not involve the use of inferential statistics with the underlying statistical assumptions. The second method uses samples, requiring reliance on inferential statistics to estimate the parameters of the entire population with assumptions concerning the underlying statistical criteria and, thus, is less accurate, but it is more feasible in practice. Drs. Grissom and Kim describe the difference between the Cohen’s *d* coefficient of the population and the estimate of the Cohen’s *d* coefficient based on sampled data.<sup>503</sup> Dr. Algina *et al.* also recognize the difference between the actual value of effect size based on full populations and the estimated value of effect size based on sampled data.<sup>504</sup> Commerce’s calculation of the Cohen’s *d* coefficient is not an estimate of the population’s actual measure of effect size, but in fact, the Cohen’s *d* coefficient is *the actual value* of the measure of effect size.

The Canadian Parties’ rebuttal to Commerce’s explanation is that “{e}ven if {Commerce} is using the full populations of data, that does not give {Commerce} license to claim that the *d* coefficient can be reasonably interpreted using Professor Cohen’s thresholds when the assumptions are violated.”<sup>505</sup> Again, the Canadian Parties confuse the issue of reliability of estimated values with the need for the statistical criteria to have a mathematical relationship between a given *d* coefficient value and the U measures.<sup>506</sup> Neither Resolute FP nor the Canadian Parties argue that the statistical criteria need to be observed here to establish the reliability of the calculated values to represent the population parameters.

---

<sup>503</sup> See *Grissom* at 63 (equations 3.1 and 3.2), 65 (equations 3.3 and 3.4), and 68 (equations 3.5 and 3.6), for a similar presentation with respect to Glass’  $\Delta$ ; see also *Ellis* at 10.

<sup>504</sup> See *Algina* at 318.

<sup>505</sup> See Canadian Parties Comments on the Draft Redetermination at 20.

<sup>506</sup> *Id.* (“As the Federal Circuit explained in *Stupp*, violating the assumptions ‘can subvert the usefulness of the interpretive cutoffs.’” (quoting *Stupp*, 5 F.4th at 1360)).

The Canadian Parties also argue that Dr. Cohen used real-world observations to illustrate differences that are representative of the interpretive thresholds, but that the data in Dr. Cohen's examples satisfy the assumptions, which supports the idea that the data used in Commerce's Cohen's *d* test must also satisfy the statistical criteria.<sup>507</sup> Even if one were to assume *arguendo* that certain examples given by Dr. Cohen may have satisfied the statistical criteria, the Canadian Parties fail to establish that it is a statutory requirement for Commerce's use of the Cohen's *d* test to also do the same when Commerce uses full populations of data and not samples. As already discussed, the assumptions in the statistical criteria are only necessary when dealing with sampled data, whereas the application of the Cohen's *d* test in Commerce's differential pricing analysis does not involve samples or estimates, but rather involves the entire population of prices. Further, Dr. Cohen did not develop his thresholds based on the statistical criteria, but rather he proposed various numerical values to represent small, medium, and large effect sizes which Dr. Cohen believed would be found to be reasonable in the academic community.<sup>508</sup> Dr. Cohen places no assumptions or conditions on the examples which he presents to illustrate each of the three threshold levels beyond the numerical value of the *d* coefficient.

Next, in response to the Canadian Parties' and Resolute FP's focus on the U measures and whether Dr. Cohen's thresholds are dependent upon the statistical criteria, the Canadian Parties' and Resolute FP's arguments that statistical criteria must be observed for the use of Dr. Cohen's *d* coefficient, or else "the resulting coefficient does not measure what {Commerce} claims it measures,"<sup>509</sup> are unpersuasive. The Canadian Parties claim that "{t}his is true regardless of whether {Commerce} uses full populations of data or sampled data in conducting

---

<sup>507</sup> See Canadian Parties Comments on the Draft Redetermination at 14.

<sup>508</sup> See *Cohen* at 24-27.

<sup>509</sup> See Canadian Parties Comments on the Draft Redetermination at 4.

its test.”<sup>510</sup> Central to the Canadian Parties’ and Resolute FP’s argument is that Dr. Cohen’s thresholds are defined by the U measures (*e.g.*,  $U_1$  as the percent nonoverlap). In the introduction to the U measures, Dr. Cohen states:

If we maintain the assumption that the populations being compared are normal and with equal variability, and conceive them further as equally numerous, it is possible to define measures of nonoverlap (U) associated with  $d$  which are intuitively compelling and meaningful.<sup>511</sup>

The Canadian Parties assert that “{w}hen any of the assumptions are not satisfied, the  $d$  coefficient does not correspond to the measures of nonoverlap, which are the foundation of Professor Cohen’s interpretive thresholds.”<sup>512</sup> Therefore, if Commerce does not observe the statistical criteria, then the results of Commerce’s Cohen’s  $d$  test “will not reasonably indicate the degree of difference between two groups of prices.”<sup>513</sup>

We disagree with the Canadian Parties that Dr. Cohen used the U measures “to create the small, medium and large thresholds ... for interpreting  $d$ .”<sup>514</sup> Once again, the Canadian Parties’ reasoning is incorrect. The Canadian Parties assert that Dr. Cohen used his table of U measures<sup>515</sup> to derive his small, medium, and large thresholds, thus requiring the statistical criteria to be satisfied in order to translate each U measure, *e.g.*,  $U_1$ , the percent nonoverlap, into a value of the  $d$  coefficient. However, the Canadian Parties fail to cite any statements by Dr. Cohen that demonstrate that this is how Dr. Cohen derived his thresholds. If thresholds were indeed derived from specific tables, it is reasonable to expect that Dr. Cohen would have expressly stated so and provided a specific reference to Table 2.2.1 as the source for small,

---

<sup>510</sup> *Id.*

<sup>511</sup> *See Cohen* at 21. Also quoted by the Federal Circuit in *Stupp*, 5 F.4th at 1357.

<sup>512</sup> *See* Canadian Parties Comments on the Draft Redetermination at 9.

<sup>513</sup> *Id.* at 10.

<sup>514</sup> *Id.* at 13.

<sup>515</sup> *Id.* at 11-13 (quoting *Cohen* at 22, Table 2.2.1).

medium, and large thresholds, given that he repeatedly referenced Table 2.2.1 whenever Table 2.2.1 was used to support findings, calculations, or analysis. However, Dr. Cohen did not define the thresholds using the U measures for incremental values of the  $d$  coefficient, and instead illustrated his thresholds by examples from real-world observations or specific values of nonoverlap.<sup>516</sup>

As explained above, Dr. Cohen proposed, as a convention, “operational definitions” of small, medium, and large thresholds when other means of interpreting the value of an effect size, *i.e.*, a Cohen’s  $d$  coefficient, are not available.<sup>517</sup> Dr. Cohen recognized that these conventions were “arbitrary” and perhaps open to misunderstanding, but Dr. Cohen thought that “the proposed conventions will be found to be reasonable by reasonable people.”<sup>518</sup> With the analysis of the difference in the means,<sup>519</sup> Dr. Cohen presents the general formulation of the Cohen’s  $d$  coefficient.<sup>520</sup> Dr. Cohen also takes that  $d$  coefficient and introduces the U measures, which are calculated given the assumptions of normality, equal variances, and equally numerous.<sup>521</sup> Specifically, Dr. Cohen presents a table of the calculated U measures for incremental values of the Cohen’s  $d$  coefficient.<sup>522</sup> Separately, Dr. Cohen introduces “‘Small,’ ‘Medium,’ and ‘Large’  $d$  Values”<sup>523</sup> for his proposed thresholds:

The terms “small,” “medium,” and “large” are relative, not only to each other, but to the area of behavioral science or even more particularly to the specific content and research method being employed in any given investigation (see Sections 1.4 and 11.1). In the face of this relativity, there is a certain risk inherent in offering conventional *operational definitions* for these terms for use in power analysis in as diverse a field of inquiry as behavioral science. This risk is nevertheless accepted in the belief that more is to be gained than lost by supplying a common conventional

---

<sup>516</sup> See Cohen at 24-27.

<sup>517</sup> *Id.* at 12-13.

<sup>518</sup> *Id.*

<sup>519</sup> *Id.* at 19-20.

<sup>520</sup> *Id.* at 20-21.

<sup>521</sup> *Id.* at 21-23.

<sup>522</sup> *Id.* at 22, Table 2.2.1; see also Canadian Parties Comments on the Draft Redetermination at 11-13.

<sup>523</sup> See Cohen at 24-27.

frame of reference which is recommended for use only when no better basis for estimating the ES index is available.<sup>524</sup>

Dr. Cohen also describes each threshold: “Small Effect Size:  $d=.2$ ,”<sup>525</sup> “Medium Effect Size:  $d=.5$ ,”<sup>526</sup> and “Large Effect Size:  $d=.8$ .”<sup>527</sup> Dr. Cohen states the “arbitrary” value for each threshold level, and then illustrates each threshold using the values for various U measures, as well as with real-world observations which represent each threshold level, and a qualitative description for each threshold. For the large threshold, Dr. Cohen states that “almost half ( $U_1=47.4\%$ ) of their areas are not overlapped,”<sup>528</sup> seemingly deemphasizing the exact percentage of nonoverlap as if that detail, contrary to the Canadian Parties’ logic, is incidental to the fact that the  $d$  coefficient is 0.8. Dr. Cohen also lists the values for the remaining two U measures as well as real-world observations where the difference between the two groups “seem like grossly perceptible and therefore large differences.”<sup>529</sup> This characterizes Dr. Cohen’s introduction of each threshold level, including the U measures and the real-world observations, which will, in Dr. Cohen’s opinion, “be found to be reasonable by reasonable people,”<sup>530</sup> and which in actual use have been found to be reasonable and widely adopted.<sup>531</sup>

The Canadian Parties arguments against Commerce’s explanation that Dr. Cohen’s thresholds are based on operational definitions and not the U measures are unsupported.

---

<sup>524</sup> *Id.* at 25 (emphasis added).

<sup>525</sup> *Id.*

<sup>526</sup> *Id.* at 26.

<sup>527</sup> *Id.*

<sup>528</sup> *Id.*

<sup>529</sup> *Id.* at 27.

<sup>530</sup> *Id.* at 13.

<sup>531</sup> *See Mid Continent I*, 940 F.3d at 673 (“Commerce reasoned {before the CIT} that even a small absolute difference in the means of the two groups can be significant (for the present statutory purpose) if there is a small enough dispersion of prices within the overall pool as measured by a proper pooled variance or standard deviation; the 0.8 standard is “widely adopted” as part of a “commonly used measure” of the difference relative to such overall price dispersion; and it is reasonable to adopt that measure where there is no better, objective measure of effect size. We agree with the {CIT} that this rationale adequately supports Commerce’s exercise of the wide discretion left to it under {section 777A(d)(1)(B) of the Act}.” (internal citations omitted)).

Nowhere in Dr. Cohen’s discussion of the U measures does he define any of the threshold levels of his proposed conventions.<sup>532</sup> Dr. Cohen’s U measures, quantified in Table 2.2.1, includes various incremental values for the *d* coefficient over a wide range, and contrary to the claims by the Canadian Parties, Dr. Cohen does not describe any of the values of the U measures as representing small, medium, or large effect sizes. In fact, once Dr. Cohen does present his proposed values for each threshold level, he then simply refers to the values of the U measures which correspond to that threshold value.<sup>533</sup> Dr. Cohen states, for example, that “*the implication of  $d = .2$  as the operational definition of a small difference between means can be seen in Table 2.2.1.*”<sup>534</sup> His choice of terminology suggests that he did not view the operational threshold as being derived from Table 2.2.1, as the Canadian Parties contend. Thus, contrary to the Canadian Parties’ claim, the progression of Dr. Cohen’s presentation of effect size, and specifically for the difference of the means, demonstrates that the U measures do not define Dr. Cohen’s proposed thresholds, but rather the U measures serve as one interpretative approach to understand any given value of the Cohen’s *d* coefficient, and not only the three specific numerical values which Dr. Cohen proposed to represent a small, medium, or large effect size.

The Canadian Parties raise that Commerce “has previously acknowledged the intrinsic relations between the degree of overlap (or nonoverlap) between two groups and {the value of the} Cohen’s *d* {coefficient}.”<sup>535</sup> The Canadian Parties assert that Commerce has not explained its changing position on the role of the percent nonoverlap, *i.e.*, the U Measures, where now Commerce asserts that the “interpretive thresholds {are} independent of the assumptions and the

---

<sup>532</sup> See *Cohen* at 21-23.

<sup>533</sup> *Id.* at 24-27.

<sup>534</sup> *Id.* at 25 (emphasis added).

<sup>535</sup> See Canadian Parties Comments on the Draft Redetermination at 23 (citing *Shrimp from India*; and *Shrimp from Vietnam*).

U measures.”<sup>536</sup> We disagree with the Canadian Parties’ assertion. Commerce has not changed its position. The U measures provide an interpretation of a given value of the Cohen’s *d* coefficient where the calculation of the values of the U measures are dependent upon the assumptions provided by Dr. Cohen. There is an intrinsic relationship between the values of the Cohen’s *d* coefficient and the values of the U measures when the assumptions are satisfied. The assumptions are required to formulate the mathematical relationship between the Cohen’s *d* coefficient and the U measures. However, as discussed above, Dr. Cohen did not define the thresholds based on the values of the U measures. Rather, Dr. Cohen used the values of the various U measures as one of several different ways to interpret, to understand, to illustrate, and to visualize a given value of the Cohen’s *d* coefficient when the assumptions, which represent an idealized situation, are satisfied and permit the calculation of the U measure values.<sup>537</sup> Similarly, Dr. Cohen provides real-world observations also to interpret, to understand, to illustrate, and in some circumstances to visualize a given threshold value of the Cohen’s *d* coefficient.

The Canadian Parties disagree with Commerce’s discussion of the Federal Circuit’s various quotations from the academic literature and argue that Commerce’s explanation “does not amount to a reasonable justification for its erroneous application of the Cohen’s *d* {test}.”<sup>538</sup> The Canadian Parties’ rejection of Commerce’s explanation is based on their faulty understanding that: (1) Dr. Cohen defined his thresholds based on the U measures; (2) Dr. Cohen’s proposed thresholds are, therefore, dependent on the statistical criteria; and (3) the need

---

<sup>536</sup> *Id.* at 24.

<sup>537</sup> *Id.* at 14 (Figure A, citing *Grissom* at 62). Figure A illustrates two overlapping bell curves, and because of the assumptions of normality and equal variances, Figure A demonstrates the calculation of the percentile standing.

<sup>538</sup> *Id.* at 24.

for the statistical criteria cannot be dismissed simply because Commerce's analysis is based on the full populations of sale prices.<sup>539</sup> We disagree with the Canadian Parties' conclusions.

As discussed above, there are two distinct issues for which consideration of the statistical criteria is not relevant in Commerce's Cohen's *d* test. First, U measures, which are the focus of the Canadian Parties and Resolute FP and which are dependent on the statistical criteria, were not used by Dr. Cohen to define his proposed thresholds. Rather these thresholds are "arbitrary" numbers selected by Dr. Cohen,<sup>540</sup> which he illustrated by the U measures and real-world observations, and which he believed would be found to be reasonable by reasonable people. Moreover, Commerce does not calculate U measures as part of its differential pricing analysis and, thus, any preconditions or limitations for calculating U measures, are not pertinent to the use of the Cohen's *d* test in Commerce's margin calculations. Second, the results of Commerce's Cohen's *d* test are based on the full populations of data in the test and comparison groups, such that the calculated parameters are not estimates, and the assessment of the reliability of those estimates, based on the statistical criteria, is not relevant.

Citing Drs. Grissom and Kim, the Canadian Parties "urge the use of a single population standard deviation ... because of the sensitivity ... to violations of the assumption of equal variances," and argue further that this sensitivity "is mathematically obvious, and does not depend on whether the groups being compared are populations or samples."<sup>541</sup> However, the

---

<sup>539</sup> *Id.* at 24-27.

<sup>540</sup> *See Cohen* at 12 ("{T}he author proposes, as a convention, {effect size} values to serve as operational definitions of the qualitative adjectives 'small,' 'medium,' and 'large.' This is an operation fraught with many dangers: The definitions are arbitrary ...") and 24-27 (for an analysis based on the difference in the means, Dr. Cohen defines the arbitrary numerical values for the proposed "small," "medium," and "large" thresholds); *see also Algina* at 318 (where Dr. Algina *et al.* note that Dr. Cohen had proposed different values for the three thresholds in an earlier text from 1969).

<sup>541</sup> *See Canadian Parties Comments on the Draft Redetermination* at 26 (citing *Grissom* at 66: "When the distribution of scores of a comparison population is not normal the usual interpretation of a  $d_G$  or  $d$  in terms of estimating the percentile standing of the average scoring members of another group with respect to the supposed normal distribution of the comparison group's scores would be invalid."); *see also Stupp*, 5 F.4th at 1358.

citation of Drs. Grissom and Kim does not support the proposition that the statistical criteria must be satisfied. This is a simple restatement of the Canadian Parties' misunderstanding that the U measures define the thresholds, except that here they substitute the percentile standing (*i.e.*,  $U_3$ ) or the percent nonoverlap (*i.e.*,  $U_1$ ). As stated above, the calculation of the U measures for incremental values of the Cohen's  $d$  coefficient are dependent upon the statistical criteria to enable the mathematical calculation.<sup>542</sup> The fact that the statistical criteria are required, mathematically, to be able to calculate the U measures for incremental values of the Cohen's  $d$  coefficient does not impact Commerce's Cohen's  $d$  test or the use of Dr. Cohen's proposed thresholds, the values for which were not defined using the statistical criteria. Finally, the percentile standing (*i.e.*,  $U_3$ ) and the percent nonoverlap (*i.e.*,  $U_1$ ) are not relevant to Commerce's calculations or analysis, because Commerce does not calculate or consider percentile standing or percentile nonoverlap. Therefore, the Canadian Parties' argument is without merit.

Next, the Canadian Parties simply reject Commerce's response to the Federal Circuit's citation of Dr. Li,<sup>543</sup> where the Federal Circuit highlighted that "{Dr.} Li concluded that Cohen's  $d$  'was found to be inaccurate when the normality and homogeneity-of-variances assumptions were violated in this study, thereby severely affecting the accuracy of  $d$  in evaluating the true [ ] in the research literature.'" <sup>544</sup> For Dr. Li, " $d$ " is the estimated value of the Cohen's  $d$  coefficient for full populations:

$$d = (\bar{Y}_1 - \bar{Y}_2) / s_p$$

Therefore, Dr. Li's statement concerning the statistical criteria involves that accuracy of the estimated value to reliably represent the actual value of the Cohen's  $d$  coefficient. However, the

---

<sup>542</sup> See "The Statistical Criteria Need Not Be Observed as Part of the Cohen's  $d$  Test" section, *supra*.

<sup>543</sup> See Canadian Parties Comments on the Draft Redetermination at 27 (citing Draft Redetermination at 27).

<sup>544</sup> See *Stupp*, 5 F.4th at 1358 (quoting *Li* at 1571).

accuracy of estimated values is not pertinent to Commerce's calculations or analysis.

Conversely, Commerce uses full populations and its calculated values are not estimates, but rather, they are the actual parameter values. Consequently, the Canadian Parties' assertion that "estimates have nothing to do with the issue" is erroneous.<sup>545</sup>

The Canadian Parties also assert that Commerce is incorrect when it concludes that "the referenced literature demonstrates a systematic tendency for Cohen's  $d$  to understate effect size when the assumptions are violated."<sup>546</sup> They point to Drs. Grissom and Kim who state that the estimate of the Cohen's  $d$  coefficient will be larger with smaller sample sizes.<sup>547</sup> We acknowledge that, when samples are used, it is possible to both overestimate and underestimate the effect size depending on specific facts. However, these statements in the academic literature discuss underestimation or overestimation of the Cohen's  $d$  coefficient which is based on *sampled* data. Commerce does not use sampled data to estimate the value of the Cohen's  $d$  coefficient but calculates the actual parameter value for the full population. Accordingly, the argument of the Canadian Parties is misplaced.

The Canadian Parties cite the texts of Drs. Grissom and Kim, Dr. Algina *et al.*, and Dr. Ellis to argue that the statistical literature "confirm that the assumptions and measures of overlap are necessary to any application of Cohen's  $d$ ."<sup>548</sup> This interpretation by the Canadian Parties misunderstands the literature and Commerce's use of full populations of data. As discussed in detail above, the U.S. price data used in the Cohen's  $d$  test include the full population of U.S. sales prices in each test and comparison group. Thus, Commerce's calculation of the Cohen's  $d$  coefficient is not an estimate of the population's actual measure of effect size, but in fact, the

---

<sup>545</sup> See Canadian Parties Comments on the Draft Redetermination at 27.

<sup>546</sup> *Id.* at 28.

<sup>547</sup> *Id.* at 29 (citing *Grissom* at 68-70).

<sup>548</sup> *Id.* at 16-17 (citing *Grissom* at 66).

Cohen's *d* coefficient is the actual value of that parameter. It is important to distinguish the three statistical criteria, which ensure that a sample statistic is a representative estimate of the parameter of an entire population, from situations where the entire population is considered and, thus, there is no need to make an estimate or ensure that a statistic based on sampled data is representative of the parameter based on the entire population data.

Lastly, concerning the Federal Circuit's hypothetical example involving small price variances, the Canadian Parties reject the findings that a calculated Cohen's *d* coefficient will not be artificially inflated, and that the subsequent steps in the differential pricing analysis are relevant to Commerce's conclusion whether to apply the alternative comparison methodology.<sup>549</sup> The Canadian Parties disagree that the use of full populations eliminates the concern that the Cohen's *d* coefficient will find significant differences when "there is a small difference between the groups."<sup>550</sup> According to the Canadian Parties, Commerce claims that the meaningful difference test will "resolve issues raised by the Federal Circuit's hypothetical {example}."<sup>551</sup> However, according to the Canadian Parties, the Federal Circuit's hypothetical example represents another situation where the statistical criteria are not met, and, thus, the results of the Cohen's *d* test "{do} not actually measure 'significant difference.'"<sup>552</sup>

We continue to disagree that test or comparison groups with very small variances in prices are problematic simply because it is assumed that such price patterns would result in an artificially large Cohen's *d* coefficient where the difference is small and would not satisfy the statistical criteria. A measure of effect size, including the Cohen's *d* coefficient, is designed to

---

<sup>549</sup> *Id.* at 29-30.

<sup>550</sup> *Id.* at 31 and footnote 91 ("{Commerce's} use of the Cohen's *d* test when values hover around the same price point can lead to 'passes' where the prices do not significantly differ.")

<sup>551</sup> *Id.* at 32.

<sup>552</sup> *Id.* at 34.

gauge the significance of the difference in the means of two groups relative to the variance within the two groups. Thus, by definition, when the variances within the two groups become smaller, it is an arithmetic tautology that the Cohen's *d* coefficient will become larger. The reverse is also mathematically true. When the variances within the two groups become larger, the Cohen's *d* coefficient will become smaller. This simply reflects the designated measure of the significance of the difference in the means of the two groups. That difference is not the absolute difference between the two groups (*i.e.*, \$1), or the absolute difference relative to the absolute price level (*i.e.*, \$1 / \$100 or 1%); the difference is measured by the variance of prices within the two groups. This reflects the SAA statement that "the Administration intends that in determining whether a pattern of significant price differences exist, Commerce will proceed on a case-by-case basis, because small differences may be significant for one industry or one type of product, but not for another."<sup>553</sup>

Nonetheless, the Canadian Parties argue that "the difference between the two prices is clearly not significant."<sup>554</sup> The Canadian Parties view the \$1 difference between the two groups either as the absolute difference in the prices (*i.e.*, one dollar) or that difference relative to the absolute price level (*i.e.*, one percent). However, to implement the statute's provision to find significant price differences, Commerce has reasonably used the concept of effect size to gauge the difference in prices relative to the variance of prices within the test and comparison groups.<sup>555</sup> Accordingly, we continue to use the Cohen's *d* test with its measure of effect size to determine whether prices differ significantly among purchasers, regions, or time periods.

---

<sup>553</sup> See SAA at 843.

<sup>554</sup> See Canadian Parties Comments on the Draft Redetermination at 32.

<sup>555</sup> See *Coe* at 5 ("Effect size quantifies the size of the difference between two groups, and may therefore be said to be a true measure of the significance of the difference.")

In the Federal Circuit’s hypothetical example, given this definition of “significant difference” based on effect size and the Cohen’s *d* coefficient, it may reasonably be determined that the difference in the means of price groups with very small variances is significant even if one may judge that difference to be small by some other measure. As Professor Coe stated, “{e}ffect size quantifies the size of the difference between two groups, and may therefore be said to be a true measure of the significance of the difference.”<sup>556</sup> Commerce continues to find that such a definition of significance based on a measure of effect size gives meaning to the pattern requirement and is fully consistent with the intent of the statute. Moreover, when Commerce’s differential pricing analysis is applied to the facts in the Federal Circuit’s hypothetical, the outcome of the differential pricing analysis, including Cohen’s *d* test, is the application of the standard A-to-A comparison method because the meaningful difference requirement is not satisfied in that hypothetical scenario involving small price variances. Moreover, the relevant question is not whether it is possible to construct a hypothetical scenario in which an application of methodology to unusual hypothetical facts could result in an unusual outcome, but rather the question is whether the methodology is reasonable.

Finally, in a recent holding, the CIT accepted Commerce’s explanation “that even if {the} Cohen’s {*d*} test can produce positive results under unusual circumstances, this possibility does not mean its use of {the} Cohen’s *d* {test} is unreasonable when combined with the ratio test and meaningful difference test.”<sup>557</sup> Nonetheless, the Canadian Parties argue that the Cohen’s *d* test must function ideally in order for it to be found to reasonably implement the statutory requirement.<sup>558</sup> They argue that the tests in the differential pricing analysis “do not operate as an

---

<sup>556</sup> See *Coe* at 5.

<sup>557</sup> See *Stupp*, 619 F. Supp. 3d at 1326.

<sup>558</sup> See Canadian Parties Comments on the Draft Redetermination at 34-35.

amalgamated whole” and that if the Cohen’s *d* test fails, subsequent tests cannot perform their intended purposes.<sup>559</sup> We disagree. Even if there are distortions or inaccuracies when calculating the Cohen’s *d* values of small test groups, this does not by itself render Commerce’s use of the Cohen’s *d* test unreasonable because the ratio test and meaningful difference test compensate for inaccuracies. Contrary to the Canadian Parties’ argument, as the CIT held, “Commerce reasonably explains that Cohen’s *d* test does not operate in a vacuum, but as part of the differential pricing analysis as a whole.”<sup>560</sup>

## 2. A Simple Average Is Reasonable to Calculate the Denominator of the Cohen’s *d* Coefficient

Commerce provided a reasoned explanation above for its determination to use the simple average of the standard deviations<sup>561</sup> to calculate the denominator of the Cohen’s *d* coefficient.

As a threshold matter, we clarify that Commerce provided two alternative bases for why its use of the simple average is reasonable. First, Commerce explained that it reexamined the academic literature on the record of the remand segment and the circumstances in which the Cohen’s *d* test is performed in Commerce’s differential pricing analysis in light of the Federal Circuit’s decision in *Mid Continent II*, which was premised on a finding that Commerce departed from the academic literature.<sup>562</sup> Accordingly, Commerce evaluated the academic literature on the record of this remand segment and concluded that its use of the simple average was consistent with the academic literature. Commerce explained that Dr. Cohen’s presentation of effect size, and specifically the *d* coefficient, differentiates between the effect size of the difference in the means of two populations in contrast with the effect size of the difference in the

---

<sup>559</sup> *Id.* at 35.

<sup>560</sup> *See Stupp*, 619 F. Supp. 3d at 1324.

<sup>561</sup> The “simple average” is “the square root of the {simple} mean of the two variances {of the two populations}” where the variance of each population is the square of the standard deviation of each population. *See Cohen* at 44.

<sup>562</sup> *See* “Calculation of the Denominator of the Cohen’s *d* Coefficient” section (subsections A-E), *supra*.

means of two groups of sampled data.<sup>563</sup> Notwithstanding the fact that Dr. Cohen’s text encompasses “power analysis” which by definition involves analysis of sampled data, the effect size is one element of Dr. Cohen’s power analysis and represents “the degree to which the phenomenon is present in the population.”<sup>564</sup> Dr. Cohen presents the general formulation for the *d* coefficient based on the comparison of two populations, equations 2.2.1 and 2.2.2, where the standard deviation of each population is assumed to be equal.<sup>565</sup> Further, when the two within-population standard deviations differ, Dr. Cohen provides equation 2.3.2 which defines the denominator of the *d* coefficient at the simple average of the two differing standard deviations.<sup>566</sup> Dr. Cohen and other authors discuss the calculation of effect size, and specifically Dr. Cohen’s *d* coefficient, based on the difference in the means of sampled data. For example, Dr. Ellis, while recognizing that effect size can be measured based on populations, presents equations for the calculation of effect size based on sampled data.<sup>567</sup> Thus, based on an evaluation of the academic literature, Commerce found that the literature does have support for Commerce’s reliance on a simple average when sampling is not used, the standard deviations of the full populations are known, and the standard deviations of both populations are not equal.

However, as Commerce explained above, in *Mid Continent III* the CIT remanded a similar explanation because according to the CIT, the Federal Circuit in *Mid Continent II* held “that Commerce needed to justify its departure from the established statistical practice,” and that even if the Federal Circuit had left Commerce with the option “to offer an explanation {} of its view of the literature,” Commerce’s explanation “fail{s} to support its position.”<sup>568</sup> Thus, as we

---

<sup>563</sup> *Id.* at subsection A.

<sup>564</sup> See “Effect Size as a Measure of Significance” section, *supra*.

<sup>565</sup> See *Cohen* at 20; see also *Algina* at 318, *Grissom* at 68.

<sup>566</sup> See *Cohen* at 43-44.

<sup>567</sup> See *Ellis* at 5 and 10.

<sup>568</sup> See *Mid Continent III*, 628 F. Supp. 3d at 1322-23.

have explained, consistent with our analysis on remand following the CIT's decision in *Mid Continent III*, even if it is assumed, *arguendo*, that Commerce has departed from the academic literature, and the sample sizes must be equal to use Dr. Cohen's equation 2.3.2, then use of the simple average is nonetheless reasonable.<sup>569</sup> Assuming that the use of the simple average of differing standard deviation, *i.e.*, Dr. Cohen's equation 2.3.2, applies to an analysis involving sampled data, and that it applies only when sample sizes are equal, because the sample sizes are equal in size and reliability, the estimated standard deviation for each of the sampled groups also has the same reliability or precision of a sample value.<sup>570</sup> Thus, a simple average of the standard deviations of the two groups is appropriate because the reliability of each value of the standard deviation is equal.<sup>571</sup> In contrast, when the sampled groups have unequal sizes, the literature teaches use of a pooled standard deviation estimate that involves weight averaging.<sup>572</sup>

Commerce took from these principles in the academic literature and explained that because it uses full populations in its application of the Cohen's *d* test, the standard deviations for the test and comparison groups each have a reliability of 100 percent. Accordingly, it is reasonable to weigh these standard deviations equally, *i.e.*, a simple average, as presented in Dr. Cohen's equation 2.3.2, in the same way as when the reliability is equal for standard deviations based on sampled data with equal sample sizes (*i.e.*,  $n_A = n_B$ ). Commerce provided this explanation as further reasonable justification to support the continued use of a simple average, even though it disagrees with the understanding that the use of Dr. Cohen's equation 2.3.2 may

---

<sup>569</sup> See "Calculation of the Denominator of the Cohen's *d* Coefficient" section (subsection F), *supra*.

<sup>570</sup> *Id.*

<sup>571</sup> *Id.*

<sup>572</sup> See *Cohen* at 66-67 and equation 2.5.2 (the formula for a weighted average of the standard deviations of groups A and B, where "*s* {is} the usual pooled within sample estimate of the population standard deviation," noting "that we have defined *s* quite generally so that it will hold for all cases involving two independent samples, whether or not sample sizes are equal."); see also *Ellis* at 26-27, footnotes 8 and 9.

be limited to where the sample sizes of the two groups must be equal. Commerce’s secondary explanation has been sustained by the CIT as reasonable.<sup>573</sup> The CIT found:

Commerce has provided an explanation that logically connects the relevance of full populations to the use of the simple averaging. Commerce is not relying solely upon the academic literature to support its choice, but rather argues that the principle it derives from the academic literature leads to a logical conclusion that simple averaging in this case is a reasonable choice.<sup>574</sup>

The CIT held that “{a}lthough there may be other reasonable alternatives, the Court cannot find fault with Commerce’s logic here. Commerce’s reliability analysis is reasonable.”<sup>575</sup>

In response to Commerce’s original explanation, the Canadian Parties argue that Commerce failed to explain how “its ‘populations’ rationale is reasonable in this {investigation}” in light of the CIT’s statement that ““ {a} test for full populations in the context of power analysis would be redundant on its face, as there would be no question of statistical significance to analyze.”<sup>576</sup> However, this understanding ignores what Dr. Cohen and other authors state in their presentations of effect size and statistical analysis. Dr. Cohen introduces his text as, “{t}he purpose of this book is to provide a self-contained comprehensive treatment of statistical power analysis from an ‘applied’ viewpoint.”<sup>577</sup> Further, “{t}he power of a statistical test depends upon three parameters: the significance criterion, the reliability of the sample results, and the ‘effect size,’ that is, the *degree* to which the phenomenon exists”<sup>578</sup> Although effect size is one element that is part of a power analysis, a statistical analysis that is based on sampled data, Dr. Cohen emphasizes that “the phrase ‘effect size’ {means} ‘the *degree* to which

---

<sup>573</sup> See *Mid Continent IV*.

<sup>574</sup> *Id.*, 680 F. Supp. 3d at 1353.

<sup>575</sup> *Id.*

<sup>576</sup> See Canadian Parties Comments on the Draft Redetermination at 42-43 (quoting *Mid Continent III*, 628 F. Supp. 3d at 1324)).

<sup>577</sup> See *Cohen* at 1 and 4 (“*The power of a statistical test of a null hypothesis is the probability that it will lead to the rejection of the null hypothesis, i.e., the probability that it will result in the conclusion that the phenomenon exists.*” (emphasis in original)).

<sup>578</sup> *Id.* at 4 (emphasis in original); see also Draft Redetermination at 21.

the phenomenon is present in the population,’ or ‘the degree to which the null hypothesis is false.’”<sup>579</sup> Thus, Dr. Cohen continues that, “it can now readily be made clear that when the null hypothesis is false, it is false to some specific degree, *i.e.*, *the effect size (ES) is some specific nonzero value in the population.* The larger this value, the greater the *degree* to which the phenomenon under study is manifested.”<sup>580</sup> What is inescapable from Dr. Cohen’s description (and the Canadian Parties provide no counter evidence or argument except to point to the statement by the CIT) is that effect size exists as a parameter of a population. The effect size may serve as a component of a statistical analysis, where the effect size may be estimated (like other parameters such as the mean) based on sampled data; however, the effect size does not exist solely as “applied” as part of a statistical analysis but also is “the *degree* to which the phenomenon is present in the population.”

The authors of the other academic literature on the record are consistent in this perspective. For example:

- Dr. Ellis presents the application of effective size in the context of a statistical analysis, but also highlights that “{t}he best way to measure an effect is to conduct a census of an entire population but this is seldom feasible in practice. Census-based research may not even be desirable if researchers can identify samples that are representative of broader populations and then use inferential statistics to determine whether sample-based observations reflect population-level parameters.”<sup>581</sup>
- Dr. Algina *et al.* propose the application of an alternative, more “robust version of Cohen’s effect size ... {that} is a better measure of population separate than is Cohen’s

---

<sup>579</sup> *Id.* at 9-10 (emphasis in original); *see also* Draft Redetermination at 22.

<sup>580</sup> *Id.* at 10 (emphasis in original).

<sup>581</sup> *See Ellis* at 5.

effect size” in the context of statistical analysis of sampled data.<sup>582</sup> Nonetheless, they highlight the difference between Cohen’s effect size (*i.e.*, Dr. Cohen’s equation 2.2.1) and estimating the effect size parameter based on sample data.<sup>583</sup>

- Drs. Grissom and Kim also focus on effect size in the context of “Univariate and Multivariate Applications,” *i.e.*, as part of a statistical analysis involving sampled data, yet they also distinguish between the estimated effect size based on sampled data, and the actual effect size based on the full populations of data.<sup>584</sup> Drs. Grissom and Kim specifically equate the “estimator of effect size” (*i.e.*, Drs. Grissom and Kim’s equation 3.5) with Dr. Cohen’s “ $d_s$ ”<sup>585</sup> which “is the standardized mean difference for the sample,” *i.e.*, estimated the effect size of the population.<sup>586</sup>

Thus, the academic literature clearly delineates between effect size as a parameter in a population, and a statistical analysis, including a power analysis, which is based on sampled data. The effect size exists outside the context of a statistical analysis as a parameter of the population, just as the mean and standard deviation are parameters of the population which can also be estimated based on sampled data. An analysis based on sampled data must assess whether the

---

<sup>582</sup> See *Algina* at 317 (“Since at least the 1960s, some methodologists have recommended reporting an effect size (ES) in addition to (or, in some cases, in place of) a hypothesis test.” (internal citations omitted)). In other words, reporting the effect size with, or instead of, the testing of the null hypothesis, such as the t-test for the difference in the means of two sampled datasets, has been a recommended practice in academic research. See *Cohen* at 19-20.

<sup>583</sup> See *Algina* at 318.

<sup>584</sup> See *Grissom* at 63 (equations 3.1 and 3.2), 65 (equations 3.3 and 3.4), and 68 (equations 3.5 and 3.6, (which identifies equations 3.1, 3.3, and 3.5 as the estimators based on sample data of the actual value of the populations’ parameter of effect size in equations 3.2, 3.4, and 3.6, respectively).

<sup>585</sup> See *Cohen* at 66-67 and equation 2.5.1; see also *Grissom* at 69 (“Note first that for an effect size for the purpose of data analysis, which is our purpose (contrasted with the purpose of pre-research power calculations to estimate needed sample size), Cohen (1988, pp. 66-67) used the notation  $d$ , (currently rarely used) where we used in Equation 3.5. Also, *Cohen* used simply  $d$ , without the subscript, to denote the targeted population effect size (a parameter) in the context of pre-research power calculations for estimating needed sample sizes, which was the main purpose of his book. However, for many years, since 1988, many researchers and some writers about effect sizes have come to use simply  $d$  to denote Cohen’s variance-pooling standardized-difference estimator of effect size for data analysis.”) (emphasis added).

<sup>586</sup> See *Grissom* at 68.

estimated parameters based on the sampled data reliably represent the actual parameters of the populations, but that does not imply that the effect size, or mean, or standard deviation only exist in the context of the statistical analysis based on sampled data.

By way of example, if one calculated the mean and standard deviation of the price per pound of all chicken products at noon on a given day between a Safeway store and a Krogers store in the same town, those calculated values are not estimates for each store but are the actual parameter values. Further, a Cohen's  $d$  coefficient comparing the difference in the mean prices of chicken products between the two stores at noon on that day would not be an estimate but would be the actual "effect size" measuring the significance of the difference in the prices of chicken between Safeway and Krogers. However, if chicken prices are sampled over a one-month period at both grocery stores to determine the significance of the price differences over that period, then those calculated values would require a statistical analysis to assess the reliability of the sample data to represent all prices during the month. In that analysis, the mean, standard deviation and effect size will be estimated based on the sampled price data, but that does not mean that the actual value of those parameters do not exist, even if unknown, outside of the analysis of the sampled data. Yet, if the estimated values satisfy the statistical criteria of the analysis, then the estimated values are found to reliably represent the parameter values of all prices of chicken products at both stores during the month.

In Dr. Cohen's presentation of Case 2, where the standard deviations differ, but the sample sizes are equal, Dr. Cohen still references the effect size for the population even though this is in the context of a power analysis. The Canadian Parties highlight Dr. Cohen's "warning 'that power values "may be greatly in error" if both samples sizes and standard deviations are

unequal.”<sup>587</sup> Further, the Canadian Parties state that the CIT rejected Commerce’s explanation<sup>588</sup> that the limitation of equal sample size does not apply to the Cohen’s *d* test because Commerce is not performing a power analysis, but rather is calculating the Cohen’s *d* coefficient based on the full population of prices in both the test and comparison groups.<sup>589</sup> As discussed above, after examination of the academic literature, we find that the sample size limitation, *i.e.*, that the sample sizes must be equal, does not impact or preclude the use of equation 2.3.2.<sup>590</sup> As further explained here, Dr. Cohen’s equation 2.3.2 involves the calculation of the denominator of the effect size of the difference in the means of the populations, and is not an estimated value based on sampled data with sample sizes.

The CIT quotation from Dr. Cohen “that power values ‘may be greatly in error’” is excerpted from Dr. Cohen’s statements concerning the consequences of unequal standard deviations in the two populations, specifically from the third of three consequences. First, Dr. Cohen states that “{t}he unequal variability *need not affect the conception of d* developed in Section 2.2 {*i.e.*, the general formulation of Dr. Cohen’s *d* coefficient}. Given that there is a difference between  $\sigma_A$  and  $\sigma_B$ , we merely are using a kind of average within-population standard deviation to standardize the difference between means.”<sup>591</sup> Dr. Cohen states that unequal standard deviations “*need not affect the conception of {the Cohen’s d coefficient}* developed in Section 2.2,” which is “standardizing the raw effect size as expressed in the measurement unit of the dependent variable by dividing it by the (common) standard deviation of the measures *in their respective populations {i.e., equation 2.2.1}*.”<sup>592</sup> The fact that Dr. Cohen’s formulation of

---

<sup>587</sup> See Canadian Parties Comments on the Draft Redetermination at 43 (quoting *Mid Continent III*, 628 F. Supp. 3d at 1325 (quoting *Cohen* at 44)).

<sup>588</sup> *Id.* (citing *Mid Continent III*, 628 F. Supp. 3d at 1325).

<sup>589</sup> *Id.* (citing Draft Redetermination at 56).

<sup>590</sup> See “Sample Sizes Do Not Limit the Use of a Simple Average” section, *supra*.

<sup>591</sup> See *Cohen* at 44 (emphasis added).

<sup>592</sup> *Id.* at 20 (emphasis added).

the effect size in equation 2.2.1 involves populations and not estimates based on samples is confirmed by the academic literature.<sup>593</sup> Further, the sample size is a characteristic of sampled data: “*n* is the number of independent units in (i.e., the size of) the sample.”<sup>594</sup> Sample size is not a characteristic of a population. It would simply be illogical to ascribe the limitation of equal sample sizes to a calculation involving full populations when sample size is not a characteristic of the populations.

Dr. Cohen’s second consequence of unequal standard deviations is “{i}n interpreting *d* for this case, the *U* (percent nonoverlap) measures can no longer be generally defined and the Table 2.2.1 *U* columns will not {be} obtain{ed}.”<sup>595</sup> As discussed above, Commerce’s analysis is not dependent on Dr. Cohen’s presentations of the *U* measures as a guide to interpreting incremental effect sizes. Accordingly, this consequence of unequal standard deviations does not impact the results of Commerce’s Cohen’s *d* test.

Dr. Cohen’s third consequence notes “that if  $\sigma_A \neq \sigma_B$  and it is also the case that  $n_A \neq n_B$ , the nominal values for *t* and power at a given significance criterion,  $\alpha$ , may differ greatly from the true values. Under these conditions ( $\sigma_A \neq \sigma_B$  and  $n_A \neq n_B$ , simultaneously), the values in Tables 2.3 {i.e., the power tables} *may be greatly in error*.”<sup>596</sup> The “values for *t*” reference the “*t* test for independent means” including “an approximate *t* test on the means of independent samples when  $\sigma_A \neq \sigma_B$ .”<sup>597</sup> As with the nominal values of *t*, the nominal value of power in Tables 2.3<sup>598</sup> “may differ greatly from the true values.”<sup>599</sup> However, Commerce’s Cohen’s *d* test is not based on

---

<sup>593</sup> See *Algina* at 318; see also *Grissom* at 68.

<sup>594</sup> See *Cohen* at 7.

<sup>595</sup> *Id.* at 44.

<sup>596</sup> *Id.* (internal citations omitted).

<sup>597</sup> *Id.* at 19.

<sup>598</sup> *Id.* at 27-39 (“The power tables are used when, in addition to the significance criterion and {effect size}, the sample size is also specified; the tables then yield power values.”).

<sup>599</sup> *Id.* at 44 (internal citations omitted).

independent samples and appropriately does not include the nominal values of  $t$  or power, and the Canadian Parties have not explained how the  $t$ -test or the power analysis are relevant to the result of the Cohen's  $d$  test. Thus, we continue to find that Dr. Cohen's third consequence is not relevant for the Cohen's  $d$  test, and we respectfully disagree with the CIT's understanding that unequal sample sizes may cause the results of the Cohen's  $d$  test to be greatly in error.

The Canadian Parties claim that Commerce misunderstands the academic literature when discussing that Dr. Cohen's equations 2.5.1 and 2.5.2 are used for sampled data.<sup>600</sup> The Canadian Parties conclude that Dr. Cohen "does not substitute equation 2.3.2 with equation 2.5.2 when using sampled data because the two equations perform distinct functions."<sup>601</sup> However, the Canadian Parties' conclusion is based on the false assumption that Dr. Cohen's equation 2.3.2 is based on sampled data rather than on full populations. The Canadian Parties state that Dr. Cohen's equations 2.5.1 and 2.5.2 are used to calculate " $d$ s" with "experimental data to detect 'the palpable characteristics of the sample and their bearing on the null hypothesis'"<sup>602</sup> whereas the Cohen's " $d$ " coefficient "is used to determine one of four variables that are part of {a} power analysis (the other three variables are  $t$ ,  $a$ , and  $n$ )."<sup>603</sup> Dr. Cohen's introduction to "The Use of the {Power} Tables for Significance Testing" notes that "{h}ere our focus shifts from research planning to the appraisal of research results, and from the consideration of the alternate-hypothetical *state of affairs in the population to the palpable characteristics of the sample* and their bearing on the. null hypothesis."<sup>604</sup> This reflects a change from "research planning" where the effect size of the population is one of four parameters in statistical inference

---

<sup>600</sup> See Canadian Parties Comments on the Draft Redetermination at 43 (citing *Cohen* at 66).

<sup>601</sup> *Id.* at 44.

<sup>602</sup> *Id.* at 43-44 (quoting *Cohen* at 66).

<sup>603</sup> *Id.* at 44. We note that "the four parameters of statistical inference {are} power, significance criterion ( $\alpha$ ), sample size ( $n$ ) and effect size (ES)" and does not include the  $t$  statistic for hypothesis testing for the difference in the means. See *Cohen* at 14.

<sup>604</sup> See *Cohen* at 66 (emphasis added).

to that of “appraisal” where the results of research analysis are evaluated using sampled data and statistical inference. Dr. Cohen generally discusses significance testing, where “we can define the effect size in the sample (ESs) using sample statistics in the same way as we define it for the population, and a statistically significant ESs is one which exceeds an appropriate criterion value.”<sup>605</sup> For the difference in the means, the effect size in the sample, “ESs,” is defined as “ $d_s$ ” in equation 2.5.1 which is defined as the difference in “the two sample means” divided by the “pooled within sample estimate of the population standard deviation, that is, {equation 2.5.2}.”<sup>606</sup> Certainly, Dr. Cohen’s equations 2.2.1 and 2.3.2 are not interchangeable with equations 2.5.1 and 2.5.2 as the former are based on full populations A and B and the latter is based on samples A and B.

This division between the calculation of the Cohen’s  $d$  coefficient using full populations and sampled data is supported elsewhere in the academic literature. In particular, Drs. Grissom and Kim describe Dr. Cohen’s two formulations of effect size.<sup>607</sup> Their “estimator of effect size,” equation 3.5, “was Cohen’s  $d_s$ ” which estimates “ $d_{pop}$ ,” the effect size of the population.

Drs. Grissom and Kim note:

that for an effect size for the purpose of data analysis, which is our purpose (contrasted with the purpose of pre-research power calculations to estimate needed sample size), Cohen (1988, pp. 66-67) *{i.e., equation 2.5.1}* used the notation  $d_s$ , ... where we used  $\{d\}$  in Equation 3.5. Also, Cohen used simply  $d$ , without the subscript, to denote the targeted population effect size (a parameter) in the context of pre-research power calculations for estimating needed sample sizes, which was the main purpose of his book.<sup>608</sup>

Thus, Drs. Grissom and Kim recognized the difference in the two formulas for the  $d$  coefficient based on populations and sampled data (even if using different nomenclature as discussed further

---

<sup>605</sup> *Id.* 17.

<sup>606</sup> *Id.* at 66-67.

<sup>607</sup> *See Grissom* at 68.

<sup>608</sup> *Id.* at 69.

below) and for different purposes. Accordingly, Dr. Cohen and other academicians provide different formulas to calculate the  $d$  coefficient whether based on full populations or sampled data; however, this does not demonstrate that Commerce's use of a simple average is limited only to where the sample sizes are equal, *i.e.*, because the weights are equal. Dr. Cohen's equation 2.3.2 also provides for a simple average when the difference in the means is based on the full populations of the two groups being compared.

The Canadian Parties also highlight that neither Professor Coe nor Dr. Ellis discuss using a simple unweighted average.<sup>609</sup> This is beside the point because Dr. Ellis and Professor Coe only provide formulas for calculating the Cohen's  $d$  coefficient, including its denominator, when based on sampled data. Simply because these texts omit formulas for calculating the Cohen's  $d$  coefficient based on full populations does not support that such formulations do not exist or are not relevant to Commerce's Cohen's  $d$  test.

Finally, the Canadian Parties reject Commerce's explanation that equation 2.3.2 relates to the calculation of the effect size based on populations because of the use of Greek letters rather than Latin letters to designate variables.<sup>610</sup> In fact, in their view, the statement from Dr. Algina *et al.* demonstrate "exactly the opposite: Professor Cohen did not follow the convention of using Greek variables to indicate population values, but instead used Latin variables to indicate population values."<sup>611</sup> We agree that the normal convention of using Greek letters to denote population variables and Latin letters to denote sample variables is not consistently followed in the academic literature. For example, compare the formulas for the calculation of the  $d$  coefficient based on populations:

---

<sup>609</sup> See Canadian Parties Comments on the Draft Redetermination at 44.

<sup>610</sup> *Id.* at 44-45.

<sup>611</sup> *Id.* at 45.

Dr. Cohen <sup>612</sup>	$d = \frac{m_A - m_B}{\sigma}$
Dr. Algina <i>et al.</i> <sup>613</sup>	$\delta = \frac{\mu_2 - \mu_1}{\sigma}$
Drs. Grissom and Kim <sup>614</sup>	$d_{pop} = \frac{\mu_a - \mu_b}{\sigma}$

Compare the formulas for the calculation of the  $d$  coefficient based on sampled data:

Dr. Cohen <sup>615</sup>	$d_s = \frac{\bar{X}_A - \bar{X}_B}{s}$
Dr. Ellis <sup>616</sup>	$Cohen's\ d = \frac{M_1 - M_2}{SD_{pooled}}$
Dr. Algina <i>et al.</i> <sup>617</sup>	$d = \frac{\bar{Y}_2 - \bar{Y}_1}{S}$
Drs. Grissom and Kim <sup>618</sup>	$d = \frac{\bar{Y}_a - \bar{Y}_b}{s_p}$
Dr. Li <sup>619</sup>	$d = \frac{(\bar{Y}_1 - \bar{Y}_2)}{s_p}$

Notwithstanding the inconsistencies in these formulas, which each reader must keep in mind when going from one academic text to another, one part that is consistent is that the standard deviation of the population is uniformly represented by the Greek letter “ $\sigma$ ” and the estimated standard deviation of the sampled data is represented by the Latin letters “ $s$ ” or “ $SD$ .” This

further supports the evidence that Dr. Cohen’s equation 2.3.2 (*i.e.*,  $\sigma' = \sqrt{\frac{\sigma_A^2 + \sigma_B^2}{2}}$  when  $\sigma_A \neq \sigma_B$ )

references the denominator of the Cohen’s  $d$  coefficient for the full populations where the perceived limitation of equal sample sizes is not relevant.

<sup>612</sup> See Cohen at 20 (equation 2.2.1).

<sup>613</sup> See Algina at 318.

<sup>614</sup> See Grissom at 68 (equation 3.6).

<sup>615</sup> See Cohen at 66-67 (equation 2.5.1).

<sup>616</sup> See Ellis at 10.

<sup>617</sup> See Algina at 318.

<sup>618</sup> See Grissom at 68 (equation 3.5).

<sup>619</sup> See Li at 1561.

Resolute FP asserts that the “{academic} literature provides no support for {Commerce’s} use of simple averages.”<sup>620</sup> Resolute FP states that the denominator of the Cohen’s *d* coefficient, *i.e.*, equation 2.2.1, “is generally defined as the ‘common within-population standard deviation’ (‘ $\sigma$ ’) assuming equal variances, equal size and normal distribution.”<sup>621</sup> It argues that, “{i}n cases of normally distributed populations of equal size but unequal variances,” Dr. Cohen provides the simple average of the standard deviations, *i.e.*, equation 2.3.2.<sup>622</sup> Dr. Cohen notes that “where variances and sample sizes are different, his table of estimates (Table 2.3) of power values ‘may be greatly in error.’”<sup>623</sup> However, according to Resolute FP, when the two groups “are of unequal size,” the academic literature requires the weighted average.<sup>624</sup> We find that Resolute FP presents similar misrepresentations of Dr. Cohen’s text as the Canadian Parties, which are discussed in detail above.

We do agree with Resolute FP, in part, that the academic literature provides for a weighted average of the estimated standard deviations when the analysis is based on sample data, where those weights in the academic literature are the sample sizes of each group and where the sample sizes are unequal. As discussed above, the sample size, *n*, has no significance in the context of a full population. Nonetheless, we disagree with Resolute FP, as with the Canadian Parties, and find that the simple average is appropriate to calculate the denominator of the Cohen’s *d* coefficient because the test and comparison groups each include the complete universe of relevant sales, and the actual, non-estimated, values of the standard deviations of those two populations differ.

---

<sup>620</sup> See Resolute FP Comments on the Draft Redetermination at 27.

<sup>621</sup> *Id.* at 27 (quoting *Cohen* at 27).

<sup>622</sup> *Id.* at 27 (citing *Cohen* at 43-44).

<sup>623</sup> *Id.* at 28 (quoting *Cohen* at 44).

<sup>624</sup> *Id.* at 28.

Resolute FP also argues that Commerce misreads the academic literature and that the Cohen's *d* test in this case involved "drawing test and control groups from the same overall population, not two separate populations as asserted by {Commerce}."<sup>625</sup> Resolute FP argues that "the standard deviation of the entire population of sales data fulfills Dr. Cohen's criterion for use of a single standard deviation:  $\sigma = \sigma_A = \sigma_B$  (where A is one set of sales compared to B, another set of sales)."<sup>626</sup> Commerce explained above that using as the denominator the single standard deviation of all sale prices of comparable merchandise is not appropriate.<sup>627</sup> Following Professor Coe's explanation, a pooled standard deviation (*i.e.*, based on a weighted average of the estimated standard deviations of the sampled data in each group) includes only the variances in the data within each group.<sup>628</sup> However, the single standard deviation includes not only the variances of the data within each group but also includes the difference in the means between the two groups. Accordingly, the single standard deviation would distort the measure of the effect size.<sup>629</sup>

We disagree with Resolute FP's characterization that all the prices in both the test and comparison groups make up the population of prices. Resolute FP claims that "{Commerce's} own definition of 'population' is internally inconsistent."<sup>630</sup> However, Resolute FP's one-size-fits-all mindset fails to account for the different aspects of Commerce's dumping analysis. We explain above how the prices in each test group and each comparison group include all prices of comparable merchandise during the POI or POR to a given purchaser, region, or time period, and

---

<sup>625</sup> *Id.* at 28.

<sup>626</sup> *Id.* at 28-29.

<sup>627</sup> See "A Single Standard Deviation of the Data in Both Groups" section, *supra*.

<sup>628</sup> *Id.* (citing *Coe* at 6 (with sample data "it is often better to use a 'pooled' standard deviation. The pooled estimate is essentially an average of the standard deviations of the experimental {*i.e.*, the test} and control {*i.e.*, the comparison} groups (Equation 4)."); and *Ellis* at 10).

<sup>629</sup> *Id.*

<sup>630</sup> See Resolute FP Comments on the Draft Redetermination at 28, footnote 131.

all prices of comparable merchandise during the POI or POR to all other purchasers, regions, or time periods.<sup>631</sup> Resolute FP illogically equates this with the fact that Commerce’s dumping analysis includes all U.S. sale prices of subject merchandise during the POI or POR,<sup>632</sup> and judges Commerce’s explanation as “internally inconsistent.” However, because the purpose and circumstances differ between the calculation of the weighted-average dumping margin and the comparison of U.S. prices in the Cohen’s *d* test,<sup>633</sup> a point upon which the CIT has agreed, Resolute FP’s logic is flawed and its conclusion that the combined prices of the test and comparison groups constitute a full population rather than two distinct populations whose means are compared to determine the significance of that difference is without merit.

Resolute FP’s argument—that the combined prices in the test and comparison groups make up the full population contemplated by Dr. Cohen—conflicts with Dr. Cohen’s presentation of his *d* coefficient. Dr. Cohen generally formulates the denominator of the *d* coefficient as “ $\sigma$ ” which he specifically defines as some combination of the standard deviations of the two groups, whether those two groups are full populations or sampled data. This is not the standard deviation of the two groups combined into a single group. As explained above, the standard deviation is calculated based on the difference between each value in the group and the mean of that group.<sup>634</sup> To calculate the single standard deviation of the prices in both groups, the mean price from which each price is measured is the mean prices of all sales in the test and comparison groups, which as explained by Professor Coe, will include not only the variance in prices within each group, but also the difference in prices between the two groups, resulting in an

---

<sup>631</sup> See “Full Population or Sampled Data” section, *supra*.

<sup>632</sup> See Resolute FP Comments on the Draft Redetermination at 28, footnote 131; see also Draft Redetermination at 24.

<sup>633</sup> See *Mid Continent IV*, 680 F. Supp. 3d at 1355-56.

<sup>634</sup> See “A Single Standard Deviation of the Data in Both Groups” section, *supra*; see also *Cohen* at 67 (equation 2.5.2, where the difference in each group, A and B, is measured relative to the mean of each group).

overestimation or valuation of the denominator of the Cohen's  $d$  coefficient.<sup>635</sup> In other words, the values of the single standard deviation will vary with the value of the differences in the means, increasing or decreasing as the difference of the means increases or decreases. In contrast, both a weighted average and a simple average will remain constant as the difference in the means changes.

Further, Resolute FP provides no evidence from the academic literature which supports the suggestion from the Federal Circuit to consider a single standard deviation for all the prices in both the test and comparison groups. The academic literature consistently distinguishes between the standard deviations of the two groups, whether those values are estimated based on sampled data or the actual parameter values based on full populations. When calculating the Cohen's  $d$  coefficient of two populations, when the standard deviations of the two groups are equal, then the denominator of the Cohen's  $d$  coefficient is "the standard deviation of either population."<sup>636</sup> When the standard deviations of the populations differ, then "the formula requires the root mean square {i.e., simple average} of  $\sigma_A$  and  $\sigma_B$ ."<sup>637</sup> When the Cohen's  $d$  coefficient is based on sampled data, then the denominator is the weighted average of estimated standard deviations of the two groups.<sup>638</sup> Resolute FP fails to identify support in the academic literature for the use of a single standard deviation.

Accordingly, we disagree with the Canadian Parties and Resolute FP that equal sample sizes limit the application of Dr. Cohen's equation 2.3.2 to calculate the denominator of the Cohen's  $d$  coefficient when the standard deviations of the two populations differ. We further disagree with Resolute FP that the single standard deviation is an appropriate basis for

---

<sup>635</sup> See "A Single Standard Deviation of the Data in Both Groups" section, *supra*.

<sup>636</sup> See *Cohen* at 20.

<sup>637</sup> *Id.* at 44.

<sup>638</sup> *Id.* at 66-67 (equations 2.5.1 and 2.5.2); see also *Ellis* at 10.

calculating the denominator of the Cohen's  $d$  coefficient. While the academic literature does provide for the use of a weighted average of the estimated standard deviations (which is not the same as a single standard deviation) when sampled data is the basis for the calculation of effect size, in Commerce's Cohen's  $d$  test, the results are based on the full populations of prices in the test and comparison groups, such that the use of a weighted average to pool the estimated standard deviations would be inconsistent with the academic literature.

In response to Commerce's subsequent explanation, the Canadian Parties claim that Commerce's reliability explanation is incorrect and argue that although Dr. Cohen discusses both reliability and equal group sizes, it has nothing to do with the assumption of equal group sizes when calculating the Cohen's  $d$  coefficient.<sup>639</sup> Resolute FP likewise argues that Commerce's reliability explanation is not supported by the academic literature.<sup>640</sup> The Canadian Parties and Resolute FP misunderstand and fail to address Commerce's explanation of reliability and why weight averaging is reasonable when sample sizes are unequal.

As the CIT stated in *Mid Continent II*, "{A}n agency is not duty-bound to follow published literature when, e.g., the literature is inapplicable to the specific problem before the agency."<sup>641</sup> Here, in explaining its alternative basis for why simple averaging is reasonable, Commerce went beyond the academic literature because Commerce uses populations, and the courts' understanding of the academic literature assumes sampling for all options to calculate the denominator of the Cohen's  $d$  coefficient.<sup>642</sup> As noted by the Federal Circuit, "when the sampled

---

<sup>639</sup> See Canadian Parties Comments on the Draft Redetermination at 41-42 and 45-48.

<sup>640</sup> See Resolute FP Comments on the Draft Redetermination at 29-30.

<sup>641</sup> See *Mid Continent II*, 31 F.4th at 1381.

<sup>642</sup> See *Mid Continent III*, 628 F. Supp. 3d at 1325 ("Commerce's assertion that equation (2.5.2) requires estimation from a sample while equation (2.3.2) does not require estimation from a sample, appears inconsistent with the literature. Although Commerce identifies  $\{\sigma_A\}$  and  $\{\sigma_B\}$  in equation (2.3.2) as representing standard deviations of full populations, it fails to consider that the  $\{\sigma\}$  values themselves seem to be used by Cohen as pre-test estimates of the full population value, which will later be calculated with sampling. Thus, Commerce's assertion that sampling is

groups have unequal sizes, the cited literature uniformly teaches use of a pooled standard deviation estimate that involves weighted averaging.”<sup>643</sup> As Commerce explained, the use of a weighted average in that scenario, *i.e.*, with sampled data where the sample sizes are unequal, is consistent with the academic literature,<sup>644</sup> and further it is reasonable as it reflects the relative reliability of the estimated standard deviations used to calculate the pooled standard deviation.

Dr. Cohen demonstrates, through the standard error, that reliability is dependent upon, and indeed directly related to, the sample size. Dr. Cohen explains that the reliability of the results based on sample data “is always dependent upon the size of the sample.”<sup>645</sup> This can be seen in Dr. Cohen’s definition of the standard error<sup>646</sup> of the estimated mean of a group of sampled data, which is inversely related to the sample size:

$$SE_{\bar{x}} = \sqrt{\frac{s^2}{n}}$$

---

not implicated in equation (2.3.2) is unsupported, as Cohen seems to use this equation in calculating statistical power.”) (internal citations and quotations omitted); *see also Cohen* at 67; *Ellis* at 10; *Coe* at 6-7.

<sup>643</sup> *See Mid Continent II*, 31 F.4th at 1378 (citing *Cohen* at 67 (equation 2.5.2, “Note that we have defined *s* quite generally so that it will hold for all cases involving two independent samples, whether or not sample sizes are equal.”); *Ellis* at 26-27 (where two equations for the pooled standard deviation, one restating Dr. Cohen’s equation 2.5.2 and the second, by Dr. Hedges, where the equation is restated in terms of the estimated standard deviations of each group); and *Coe* at 6 (Equation 4 which restates Dr. Hedges equation from *Ellis*)).

<sup>644</sup> Under either of Commerce’s explanations, the use of a weighted average to calculate the denominator of the Cohen’s *d* coefficient with sampled data is consistent with the academic literature when sampled data is the basis for the calculation of the Dr. Cohen’s *d* coefficient. *See Cohen* at 66-67 (equation 2.5.2); *Ellis* at 26-27, footnotes 8 and 9; and *Coe* at 6 (equation 4). Each of these equations applies to the calculation of the *d* coefficient when sampled data is used. The question raised by the Federal Circuit is whether Dr. Cohen’s equation 2.3.2 supports Commerce use of a simple average in its Cohen’s *d* test. Under the first explanation on remand from *Mid Continent II*, Dr. Cohen’s equation 2.3.2 (simple average) applies to populations and Commerce’s Cohen’s *d* test, whereas Dr. Cohen’s equation 2.5.2 *et al.* (weighted average) applies to sampled data (note that when the sample sizes are equal, equation 2.5.2 *et al.* resolve to a simple average). Under the second explanation on remand from *Mid Continent III*, where the understanding is that Dr. Cohen’s equation 2.3.2 also applies to sample data with equal sample sizes, Commerce’s use of a simple average, outside of the teachings of the academic literature, is supported based on the equal reliability of the calculated standard deviations, either based on sampled data with equal sample sizes or, alternatively, using full populations as in Commerce’s Cohen’s *d* test.

<sup>645</sup> *See Cohen* at 6.

<sup>646</sup> *Id.* (“one conventional means for assessing the reliability of a statistic is the standard error (SE) of the statistic”).

where  $s^2$  is the square of the estimated standard deviation of the sampled data and  $n$  is the sample size.<sup>647</sup> As the sample size,  $n$ , increases, the standard error decreases, and the reliability of the estimated statistic increases.

As discussed above, because the reliability of an estimated parameter increases as the sample size increases, it is reasonable that when averaging the estimated parameters that the values of these parameters be weighted to reflect the relative reliability of the two values being averaged. If one value is more reliable than another, then the more reliable value logically warrants more weight than a less reliable value. This is reflected in the equations in the academic literature, as cited by the Federal Circuit, where the estimated values of the standard deviations of each group are weighted by the sample size of each group on which each value is calculated.<sup>648</sup> Based on the reliability of the estimated standard deviations using sampled data, a weighted average of the estimated standard deviations results when the sample sizes are unequal, and a simple average results when the sample sizes are equal. However, because Commerce is using full populations and not samples, Commerce must determine a reasonable method of calculating the denominator of the Cohen's  $d$  coefficient. Based upon reliability, when Dr. Cohen's equation 2.3.2 applies only when sample sizes are equal, Commerce considers that a simple average is reasonable to calculate the denominator for Commerce's Cohen's  $d$  test, when the values of the standard deviations are based on the full populations of the data within each group (*i.e.*, within the test group and within the comparison group) independent of the number of observations in each population. When using a full population for both groups, and regardless of

---

<sup>647</sup> *Id.* at 6-7.

<sup>648</sup> *Id.* at 67 (equation 2.5.2); *see also Ellis* at 26-27, footnotes 8 and 9; *Coe* at 6 (equation 4, where each estimated value of the standard deviation is weighted by the sample size).

the number of observations, the reliability of each group is equally 100 percent reliable, and thus it is appropriate to weight the values equally, *i.e.*, to calculate a simple average.

The Canadian Parties disagree with the CIT's holding that Commerce "provided an explanation that logically connects the relevance of full populations to the use of simple averaging" in its reliability analysis.<sup>649</sup> In their view, "the literature does not support simple averaging in calculating {the} Cohen's *d* {coefficient} for different sample sizes that are equally reliable," and the reliability analysis does not apply to Commerce's use of the Cohen's *d* test.<sup>650</sup> The Canadian Parties' arguments are misplaced and ignore that Commerce's reliability analysis is an explanation in which Commerce went beyond the academic literature given the understanding of the Federal Circuit, and took principles from it in applying it to Commerce's analysis where it uses full populations. As the CIT succinctly described, "Commerce identifies where simple averaging is supported by the literature, extrapolates a rationale for why simple averaging is appropriate, and then applies that rationale to the circumstances before Commerce."<sup>651</sup> Further, "{a}lthough there may be other reasonable alternatives," the CIT found that "Commerce's reliability analysis is reasonable."<sup>652</sup> Commerce provided a reasonable explanation for why the use of a simple average is supported by principles derived from the academic literature, while using a single standard deviation is not. That is all that is required.<sup>653</sup>

---

<sup>649</sup> See Canadian Parties Comments on the Draft Redetermination at 46 (citing *Mid Continent IV*, 680 F. Supp. 3d 1346)

<sup>650</sup> *Id.* at 46-47.

<sup>651</sup> See *Mid Continent IV*, 680 F. Supp. 3d at 1353.

<sup>652</sup> *Id.*

<sup>653</sup> See *Stupp*, 5 F.4th at 1354 (stating that the standard of review for components of Commerce's differential pricing analysis is reasonableness).

Thus, even if the Canadian Parties presented alternative explanations that in theory may also be reasonable, it does not detract from the reasonableness of Commerce's explanation here.<sup>654</sup>

Moreover, the Canadian Parties' argument that the reliability analysis does not apply to Commerce's use of the Cohen's *d* test is misleading. To be clear, none of the academic texts on the record here discuss Commerce's "Cohen's *d* test." Neither Dr. Cohen, Dr. Ellis, nor Professor Coe opined on the application of the concept of effect size to examine whether prices differ significantly among purchasers, regions, or time periods under the AD statute. Nor could one reasonably expect an academic author to be omniscient and describe all possible applications of his or her concepts, including the situation addressed by Commerce in the use of its Cohen's *d* test. Similarly, these academic authors do not know the myriad situations in which their concepts may be applied. Such expectations are unrealistic that any applications must be preordained by an academic author rather than their concepts being adapted and applied in situations unimagined by the original authors. Nonetheless, these academicians did describe the general principles behind both the concept of effect size and its place in research and data analysis which Commerce has applied in its differential pricing analysis. Commerce has followed these principles in conceptualizing and applying the Cohen's *d* test to the task before the agency, administering AD laws and calculating dumping margins based on the record.

The Canadian Parties argue that the reliability analysis provides no information that would entitle the standard deviations of the test and comparison groups to equal weighting.<sup>655</sup> According to the Canadian Parties, "the weights assigned to the standard deviations in equation 2.3.2 must be based on two considerations: whether the standard deviations are equal and

---

<sup>654</sup> See *Koyo Seiko Co. v. United States*, 36 F.3d 1535, 1570 (Fed. Cir. 1994) ("To survive judicial scrutiny, an agency's construction need not be the *only* reasonable interpretation or even the *most* reasonable interpretation.") (emphasis in original).

<sup>655</sup> See Canadian Parties Comments on the Draft Redetermination at 47-48.

whether the group sizes are equal.”<sup>656</sup> Even given, *arguendo*, the Federal Circuit’s understanding, the Canadian Parties misrepresent what Dr. Cohen states. First, there are no weights in Dr. Cohen’s equation 2.3.2; the equation is the “root mean square” of the standard deviations of each group, such that the square of each standard deviation is summed and divided by “2.” This calculation includes no weights, as, for example, in Dr. Ellis’s or Professor Coe’s equations where the standard deviations are weighted by the sample size minus one (*e.g.*,  $n_A - 1$ ).<sup>657</sup> Second, the “two considerations” for “Case 2” of Dr. Cohen’s illustrative examples for the use of his power tables,<sup>658</sup> are that “ $\sigma_A \neq \sigma_B, n_A = n_B$ .”<sup>659</sup> In other words, the two considerations include the situations where the standard deviations of the two populations differ, and the sample sizes are equal.<sup>660</sup> Note that “ $n_A$ ” and “ $n_B$ ” are the sample sizes of groups A and B which are composed of sampled data; this is not the number of observations in population A or population B.<sup>661</sup> Notwithstanding the misstatement that the standard deviations are equal, the Canadian Parties are wrong to project Dr. Cohen’s requirement that sample sizes be equal to also require that the size of the populations be equal. This is irrespective of whether one considers that equal sample sizes are required in order to use Dr. Cohen’s equation 2.3.2 (*i.e.*, whether equation 2.3.2 applies to populations or sampled data).

In Commerce’s application of the Cohen’s *d* test, Commerce uses the full populations of data, *i.e.*, all prices of comparable merchandise to a given purchaser, region, or time period (*i.e.*,

---

<sup>656</sup> *Id.* (citing *Cohen* at 43-44).

<sup>657</sup> *See Ellis* at 27, footnote 9; *see also Coe* at 6 (equation 6).

<sup>658</sup> *See Cohen* at 27-52.

<sup>659</sup> *Id.* at 43.

<sup>660</sup> The Canadian Parties often conflate the term “sample size” with the term “group size” with no explanation of what “group size” means. The term “group size” is not used by Dr. Cohen. The Canadian parties point to, *e.g.*, *Cohen* at 55-66 or 133-44, allegedly where Dr. Cohen “discusses equal group size in detail.” *See Canadian Parties Comments on the Draft Redetermination* at 42. However, each of the citation point to Dr. Cohen’s discussion of “sample size,” *not* “group size,” and *not* the number of observations in a population.

<sup>661</sup> *See Cohen* at 14 (“Four parameters of statistical inference have been described: power, significance criterion (a), *sample size* (*n*), and effect size (ES).” (emphasis added)).

the test group) and all prices of comparable merchandise to all other purchasers, regions, or time periods (*i.e.*, the comparison group). As a result, the standard deviations calculated for the test and comparison groups each have a reliability of 100 percent, *i.e.*, “the closeness with which {the calculated value} can be expected to approximate the relevant population value.”<sup>662</sup> In other words, the reliability of the calculated standard deviations based on the full population of sale prices to each group is identical. Because the reliability of the standard deviations based on full populations is equal, to calculate the denominator of the Cohen’s *d* coefficient, Commerce found that it is reasonable to weight these standard deviations equally, *i.e.*, a simple average, as presented in Dr. Cohen’s equation 2.3.2, just as when the reliability is equal for standard deviations based on sampled data with equal sample sizes. The academic literature’s limitations on sample size, as understood in *Mid Continent II* and *Mid Continent III*, for the different ways of calculating the denominator of the Cohen’s *d* coefficient are whether the sample sizes are equal or unequal. Because the groups are equally reliable when the sample sizes are equal and because when full populations are used, they too are equally reliable, Commerce finds it appropriate to use Dr. Cohen’s equation 2.3.2 when full populations are used.

### 3. Commerce’s Use of the Alternative Comparison Method Is Consistent with the Act

In its Rule 57(3) reply brief to the Panel, Resolute FP argues that the differential pricing analysis has permitted Commerce to turn the exception into the rule.<sup>663</sup> Resolute FP provided information of Commerce’s 165 determinations during the POI of the *Final Determination* (*i.e.*, October 1, 2015, through September 30, 2016).<sup>664</sup> Resolute FP asserts that “Commerce found a

---

<sup>662</sup> *Id.* at 6.

<sup>663</sup> See Resolute FP Rule 57(3) Reply Brief at 18-20.

<sup>664</sup> *Id.* at Appendix B and at 19, footnote 56 (“Resolute did not include determinations in which Commerce did not calculate a {weighted-average dumping} margin from Respondent’s data and, therefore, did not utilize the

significant pattern of price differences (*i.e.*, targeted dumping) in 145 of those 165 investigations or reviews, or in 87% of the determinations.”<sup>665</sup> Resolute FP concludes that Commerce may not, even though “it is entitled to ‘tremendous deference,’” use the differential pricing analysis when its interpretation of the statute results in the “‘exception’ {becoming} the norm.”<sup>666</sup>

In its comments on the Draft Redetermination, Resolute FP expanded its claim from its Rule 57(3) reply brief, alleging that “Commerce used the {differential pricing analysis} to find targeted dumping in 145 of 165 investigations and reviews (87%) conducted during the period of investigation ... making {the A-to-T method} the norm rather than an exception.”<sup>667</sup> Resolute FP asserts that the statute, supported by the SAA, is “unambiguous” that the A-to-T method is intended “only to address specifically targeted dumping as an uncommon practice when statutory conditions are met.”<sup>668</sup> In their comments on the Draft Redetermination, the Canadian Parties also assert that Commerce “used the {differential pricing analysis} and found targeted dumping, resulting in the application of {an alternative comparison methodology based on the A-to-T method} in 145 or 165 dumping investigation and reviews during the period of review {*sic*} for this investigation.”<sup>669</sup> The commenters conclude that this, in their view, impermissible overuse of the A-to-T method is because Commerce has not adhered to “the strict conditions prescribed by the statute”<sup>670</sup> that has allowed “the ‘exception’ {to} impermissibly swallow {} the rule.”<sup>671</sup>

---

{differential pricing analysis}. Thus, excluded from the 165 determinations are determinations in which Commerce (1) applied AFA (either total AFA or AFA to the {differential pricing analysis}), (2) found no shipments of subject merchandise during a Period of Review or (3) had insufficient sales to conduct a targeted dumping analysis.”)

<sup>665</sup> *Id.* at 19

<sup>666</sup> *Id.* (internal citation omitted).

<sup>667</sup> See Resolute FP Comments on the Draft Redetermination at 26.

<sup>668</sup> *Id.* at 25.

<sup>669</sup> See Canadian Parties Comments on the Draft Redetermination at 36 (citing Resolute FP Rule 57(3) Reply Brief at 18-20 and Appendix B).

<sup>670</sup> See Resolute FP Comments on the Draft Redetermination at 26.

<sup>671</sup> See Canadian Parties Comments on the Draft Redetermination at 37.

We disagree with Resolute FP and the Canadian Parties that Commerce’s use of the differential pricing analysis has led to an impermissible reliance on the A-to-T method. Both commenters misrepresent the statute to reach distorted conclusions concerning Commerce’s differential pricing analysis, including the Cohen’s *d* test. Further, as discussed below, the information provided by Resolute FP in its Rule 57(3) reply brief is flawed such that it fails to support its conclusion. In addition, it is inconsistent with the limitations which Resolute FP claims to have applied in aggregating this information, as well as its criticisms of the information which Commerce included with the Draft Redetermination concerning final determinations (information that was also provided in response to the *Stupp* remand order in that litigation).

We disagree with the conclusions of Resolute FP and the Canadian Parties that the statute, regulations, or legislative history express any expectation on the frequency that Commerce might resort to an alternative comparison methodology pursuant to section 777A(d)(1)(B) of the Act.<sup>672</sup> Section 777A(d)(1)(A) of the Act provides that in an investigation, Commerce will “determine whether the subject merchandise is being sold in the United States at less than fair value” using either A-to-A comparisons or T-to-T comparisons. Section 777A(d)(1)(B) provides the A-to-T method as a possible “exception” to either the A-to-A method or the T-to-T method when the two statutory requirements are met (*i.e.*, the pattern requirement and the meaningful difference requirement). If Commerce finds that these two requirements are satisfied, then it may resort to the “exception,” an alternative comparison methodology based on the A-to-T method.

---

<sup>672</sup> Commerce notes that sections 777A(d)(1)(A) and (B) of the Act specifically address a LTFV investigation. However, beginning in 2012, Commerce elected to follow the same statutory scheme in a review as in a LTFV investigation. See *Antidumping Proceedings: Calculation of the Weighted-Average Dumping Margin and Assessment Rate in Certain Antidumping Duty Proceedings; Final Modification*, 77 FR 8101 (February 14, 2012) (*Final Modification for Reviews*).

Resolute FP and the Canadian Parties argue that an alternative comparison method “must be used with restraint.”<sup>673</sup> Resolute FP relies on the SAA requirement “that Commerce ‘must establish and provide an explanation why it cannot account for such differences through the use of {A-to-A} or {T-to-T} comparisons *sic*’ and that ‘Commerce will proceed on a case-by-case basis, because small differences may be significant for one industry or one type of product, but not for another.’”<sup>674</sup> Neither of these passages support the conclusion that the statute, regulations, or legislative history limit the use of an alternative comparison method beyond the two statutory requirements. Resolute FP’s first citation simply repeats the meaningful difference requirement provided in section 777A(d)(1)(B)(ii) of the Act. Resolute FP’s second citation, in full, addresses the pattern requirement:

the Administration intends that *in determining whether a pattern of significant price differences exist*, Commerce will proceed on a case-by-case basis, because small differences may be significant for one industry or one type of product, but not for another.<sup>675</sup>

This concerns Commerce’s examination of whether there exists a pattern of prices that differ significantly pursuant to section 777A(d)(1)(B)(i) of the Act; further, this is exactly what Commerce does when using the Cohen’s *d* test, *i.e.*, using a measure of effect size, to gauge the significance of the difference in prices which takes into account the variations in prices within the test and comparison groups to define whether the difference in prices are significant. Neither of these statements from the SAA support Resolute FP’s argument that there is some undefined limit on Commerce’s use of an alternative comparison method beyond the two statutory requirements.

---

<sup>673</sup> See Resolute FP Comments on the Draft Redetermination at 25; see also Canadian Parties Comments on the Draft Redetermination at 37 (“{Commerce} should use the A-to-A methodology generally and an alternative method sparingly.”).

<sup>674</sup> See Resolute FP Comments on the Draft Redetermination at 25 (quoting SAA at 843).

<sup>675</sup> See SAA at 843.

In its Rule 57(3) reply brief, Resolute FP distorts the statutory requirement and Commerce’s use of the exception to one of the standard comparison methodologies. The statute provides for an alternative comparison methodology when the pattern requirement and the meaningful difference requirement have been met. Thus, the pattern requirement is but one step in the possible use of the “exception,” the alternative comparison methodology. Nonetheless, Resolute FP asserts that “Commerce calculated margins in 165 determinations in antidumping investigations and reviews during the {*Final Determination* POI} ... {and} found a significant pattern of price differences (*i.e.*, targeting dumping) in 145 of those investigations or reviews, or in 87% of the determinations.”<sup>676</sup>

First, Resolute FP erroneously equates a pattern of prices that differ significantly with “targeted dumping.” As explained above, neither the Cohen’s *d* test nor the ratio test involves comparisons with normal value.<sup>677</sup> The Cohen’s *d* test only involves the comparison of U.S. prices, and the ratio test aggregates the results of the Cohen’s *d* test. Second, Resolute FP erroneously equates finding that a pattern exists with the use of the alternative comparison methodology. As explained above, for Commerce to be permitted to resort to an alternative comparison methodology, both the pattern requirement and the meaningful difference requirement must be satisfied. Resolute FP either simply equates the “exception” with satisfying the pattern requirement or ignores the fact that the meaningful difference requirement must also be met for Commerce to resort to an alternative comparison methodology. Either approach is

---

<sup>676</sup> See Resolute FP Rule 57(3) Reply Brief at 19 (citing Resolute FP Rule 57(3) Reply Brief at Appendix B (Resolute FP Appendix B)). Note, in Resolute FP Appendix B, for a decision in an investigation or review to have used the “exception” (*i.e.*, one of the 145 decisions), Commerce had to have found a pattern of prices for at least one respondent for which a rate had been calculated. From Resolute FP Appendix, 145 of the 165 *Federal Register* notices of decisions included at least one such respondent.

<sup>677</sup> See “The Federal Circuit’s Hypothetical Example” section, *supra* (“The Cohen’s *d* test only examines the relationship of prices of the subject merchandise within the U.S. market and does not examine whether the U.S. price is at less than normal value.”).

wrong, and Resolute FP’s conclusion that Commerce made use of the “exception” “in 145 of {} 165 investigations and reviews” is without merit.<sup>678</sup>

In their comments on the Draft Redetermination, both Resolute FP and the Canadian Parties emphasize that the information in Resolute FP Appendix B demonstrates that Commerce’s use of the differential pricing analysis, and in particular a purportedly flawed Cohen’s *d* test, has caused the exception to swallow the rule.<sup>679</sup> However, in arguing that the exception has become the rule, Resolute FP and the Canadian Parties focus on Commerce’s finding that a pattern of prices exists rather than properly examining whether Commerce ended up using an alternative comparison method to calculate at least one respondent’s weighted average dumping margin. Information concerning the latter is not included in the Resolute FP Appendix B.

Notwithstanding that Resolute FP does not base its analysis on Commerce’s use of an alternative comparison method, the information in the Resolute FP Appendix B is flawed and cannot be the basis for a reasonable analysis. First, Commerce’s consideration of the statutory requirements and decision concerning the appropriate comparison methodology is done on a respondent-by-respondent basis. However, Resolute FP counts a determination as one of the 145 exceptions when at least one company was found to have a pattern of prices, whether there was one company or numerous companies for which an alternative comparison method was used as a result of the differential pricing analysis. For example, if an investigation involved two

---

<sup>678</sup> See Resolute FP Rule 57(3) Reply Brief at 19.

<sup>679</sup> See Resolute FP Comments on the Draft Redetermination at 26 (“{Commerce} used the {differential pricing analysis} to find targeted dumping in 145 or 165 investigations and reviews (87%) conducted during the {POI}, making A-to-T the norm rather than an exception.”); see also Canadian Parties Comments on the Draft Redetermination at 36 (“Resolute observed that {Commerce} has used the {differential pricing analysis} and found targeted dumping, resulting in the application of the A-to-T method or the mixed A-to-A and A-to-T method (the ‘mixed method’) in 145 out of 165 dumping investigations and reviews during the period of review for this investigation.”) and 37 (“{T}he frequency with which {Commerce} applies an alternative {comparison} method demonstrates that the ‘exception’ has impermissibly swallowed the rule.”).

respondents and Commerce found that absence of the pattern of prices that differ significantly for one of the respondents, but not for the other, Resolute FP ignores a negative finding and only counts a positive finding.<sup>680</sup> Second, Resolute FP includes preliminary decisions in its data, which are not final. Not only is the inclusion of preliminary decisions inappropriate, but it also leads to a large majority of the segments included by Resolute FP to be double counted, and one segment is even triple counted.<sup>681</sup> There are also numerous segments with only a preliminary decision, which may be revised for the final decision for any number of reasons. Third, Resolute FP notes that it has excluded certain determinations where Commerce did not calculate a weighted-average dumping margin.<sup>682</sup> Beyond the application of total AFA or a “no shipments” final in a review, Resolute FP does not describe what the application of AFA to the differential pricing analysis might be, or how it determined whether there were insufficient sales to conduct a “targeted dumping analysis” (perhaps a differential pricing analysis). Indeed, Resolute FP even included the final results of a changed circumstances review for the AD order on *Circular Welded Pipe from Korea* which examined a successor-in-interest request and no weighted-

---

<sup>680</sup> For example, in the final determination for the investigation of hot-rolled steel flat products from Japan, Commerce found that a pattern exists for Nippon Steel and Sumitomo Metal, yet no pattern was found to exist for JFE Steel. See Attachment IV. This is reflected in the Resolute FP Appendix B information where “# Companies where {Commerce} calculated rate” is two yet “# Companies’ sales where 33%+ pass Cohen D Test” one and the determination is marked as “Yes.” See Resolute FP Appendix B at page 25. This same pattern is repeated where Resolute FP double-counts Commerce’s preliminary determination in this investigation. See Resolute FP Appendix B at page 15. In general, this occurs for each entry in Resolute FP’s table where the “# Companies where {Commerce} calculated rate” is greater than the “# Companies’ sales where 33%+ pass Cohen D Test.”

<sup>681</sup> See Attachment III, which lists each of the determinations included in Resolute FP Appendix B by segment. Segments where more than one *Federal Register* notice is recorded have been double counted, and in one situation, tripled counted: *Certain Cold-Rolled Steel Flat Products from Brazil: Affirmative Preliminary Determination of Sales at Less Than Fair Value, Postponement of Final Determination, and Extension of Provisional Measures*, 81 FR 11754 (March 7, 2016); *Certain Cold-Rolled Steel Flat Products from Brazil: Amended Preliminary Determination of Sales at Less Than Fair Value*, 81 FR 20366 (April 7, 2016); and *Certain Cold-Rolled Steel Flat Products from Brazil: Final Determination of Sales at Less Than Fair Value*, 81 FR 44946 (July 29, 2016).

<sup>682</sup> See Resolute FP Rule 57(3) Reply Brief at 19, footnote 56 (“Resolute did not include determinations in which Commerce did not calculate a margin from Respondent’s data and, therefore, did not utilize the {differential pricing analysis}. Thus, excluded from the 165 determinations are determinations in which Commerce (1) applied AFA (either total AFA or AFA to the {differential pricing analysis}), (2) found no shipments of subject merchandise during a Period of Review or (3) had insufficient sales to conduct a targeted dumping analysis.”)

average dumping margin was calculated.<sup>683</sup> Fourth, as noted in the analysis of determinations for the same period, certain segments were omitted from the Resolute FP Appendix B data.<sup>684</sup> Thus, the data do not support the parties' conclusion that the exception has become the rule.

As explained above, we have included from the *Stupp* redetermination our analysis of final determinations in investigations during calendar years 2015 (the year of the final determination of the investigation underlying *Stupp*) and 2021 (the year of the *Stupp* opinion and the most recent calendar year for Commerce's redetermination in that litigation).<sup>685</sup> When properly analyzed on a respondent-specific basis, Commerce resorted to an alternative comparison method for 22 percent and 21 percent of the respondents with calculated rates in 2015 and 2021, respectively.<sup>686</sup> Nonetheless, Resolute FP dismisses Commerce's analysis because "{Commerce} included AFA determinations and did not attempt to distinguish and eliminate cases where it found insufficient sales to conduct a targeted dumping analysis. {Commerce} also ignores administrative reviews, making the sample size too small to be statistically significant."<sup>687</sup>

First, Resolute FP is incorrect that Commerce's analysis includes respondents whose rate was determined based on total AFA. Although respondents whose rates were determined based on total AFA are included in the listings for 2015 and 2021, they are not included in the calculation of the proportion of calculated rates where an alternative comparison method was used. Second, as with its only analysis, Resolute FP insists on excluding observations where

---

<sup>683</sup> See Resolute FP Appendix B at page 21 (citing *Notice of Final Results of Antidumping Duty Changed Circumstances Review, Circular Welded Non-Alloy Steel Pipe from the Republic of Korea*, 81 FR 42653 (June 20, 2016) (*Circular Welded Pipe from Korea*)).

<sup>684</sup> See Attachment IV for respondent-specific information for LTFV investigations for which Commerce published a final determination *Federal Register* notice during the POI for the *Final Determination*.

<sup>685</sup> See "Commerce Application of an Alternative Comparison Methodology" section, *supra*; see also Attachments I and II.

<sup>686</sup> *Id.* at 43.

<sup>687</sup> See Resolute FP Comments on the Draft Redetermination at 25, footnote 118.

there are “insufficient sales” without detailing what that means.<sup>688</sup> Third, Resolute FP’s logic that reviews must be included or else “the sample size {would be} too small to be statistically significant” is flawed. There is no sampling and statistical significance has no meaning in this context.

Moreover, this litigation concerns an investigation, and not an administrative review or any other type of segment. While the *Final Modification for Reviews* indicates that Commerce performs the same dumping analysis, including the differential pricing analysis, in a review as in an investigation, there are important differences between the two types of segments. First, in an administrative review, unlike an investigation, the sales of entries are done under the discipline of an antidumping order which impacts the pricing behavior of the respondent, and thus, impacts the results of Commerce’s dumping analysis. Second, respondents in a review are limited by requests for review, and consequently to some level of self-selection, and not all producers and exporters are subject to examination. Third, verifications are always undertaken as part of an investigation to ensure the veracity of the information submitted to Commerce, and the *de minimis* threshold for the meaningful difference test differs between an investigation and a review.<sup>689</sup> Therefore, although Commerce follows a similar approach in a review as in an investigation, comparing the impact of the differential pricing analysis between investigations and reviews is somewhat an apples-to-oranges comparison because of inherent differences in these types of segments. The situation at issue here is an investigation, and, therefore, the analysis Commerce conducted as part of the *Stupp* redetermination, which was limited to

---

<sup>688</sup> As noted above, Resolute FP even included the results of a changed circumstances review where no weighted-average dumping margin was calculated.

<sup>689</sup> Compare 19 CFR 351.106(b)(1) and (c)(1).

investigations, is the appropriate method to gauge the instances in which Commerce has applied the alternative comparison method in the same type of the segment.

In addition to the information included as Attachments I and II,<sup>690</sup> Commerce has provided respondent-specific information on the results of the differential pricing analysis for final determinations published during the POI of the *Final Determination* as Attachment IV.<sup>691</sup> Of the 71 respondents in these LTFV investigations, 50 had calculated rates, of which 14 were based on an alternative comparison method. Thus, only 28 percent of the respondents with calculated rates in the final determinations published during the POI of the *Final Determination* were based on an alternative comparison method. This is comparable with the results found for calendar years 2015 and 2021, when Commerce resorted to an alternative comparison method for 22 percent and 21 percent of the respondents in investigations with calculated rates in 2015 and 2021, respectively.<sup>692</sup> The information for years 2015 and 2021 was first analyzed in a redetermination in response to the *Stupp* remand order, which the CIT sustained.<sup>693</sup> In contrast, Resolute FP's and the Canadian Parties' "analysis" rests on double-counting (and at times even triple counting) as well as other flaws, as described above. When data for investigations, *i.e.*, the relevant type of segment, is properly analyzed on a respondent-specific basis, the data demonstrate that the application of Cohen's *d* test, as part of differential pricing analysis, resulted in the application of the standard A-to-A method to the vast majority of respondents in investigations during the three years that Commerce examined.

---

<sup>690</sup> See Attachments I and II.

<sup>691</sup> See Attachment IV.

<sup>692</sup> *Id.* at 43.

<sup>693</sup> See *Stupp*, 619 F. Supp. 3d 1314.

#### 4. Commerce Is Not Required to Consider Seasonality or to Average Over Shorter Time Periods

The Canadian Parties argue that in identifying significant price differences over periods of time, Commerce should consider alleged “price volatility due to seasonal market conditions or simply price variability over time.”<sup>694</sup> Resolute FP argues that seasonality should be considered because the SAA indicates that small differences may be significant for one industry or type of product, but not another.<sup>695</sup> We disagree that Commerce is required consider seasonal price volatility in looking for significant prices differences over periods of time under the Act or the SAA. Section 777A(d)(1)(B) does not require Commerce to determine the reasons why there is a pattern of export prices for comparable merchandise that differs significantly among purchasers, regions, or time periods. The statute only requires that there exist “significant differences” in prices of comparable merchandise, but there is no obligation for Commerce to consider the underlying reasons that these price differences exist. The Federal Circuit has upheld that when conducting its differential pricing analysis, Commerce is not obligated to investigate a respondent’s intention or subjective reasons for differing prices:

Section {777A(d)(1)(B)} does not require Commerce to determine the reasons why there is a pattern of export prices for comparable merchandise that differs significantly among purchasers, regions, or time periods ... the CIT did not err in finding there is no intent requirement in the statute, and we agree with the CIT that requiring Commerce to determine the intent of a targeted dumping respondent “would create a tremendous burden on Commerce that is not required or suggested by the statute.”<sup>696</sup>

---

<sup>694</sup> See Canadian Parties Comments on the Draft Redetermination at 35-36.

<sup>695</sup> See Resolute FP Comments on the Draft Redetermination at 26 (citing SAA at 843).

<sup>696</sup> See *JBF RAK LLC v. United States*, 790 F.3d 1358, 1368 (Fed. Cir. 2015); see also *Borusan Mannesmann Boru Sanayi ve Ticaret A.S. v. United States*, 608 F. App’x 948, 949 (Fed. Cir. 2015) (agreeing with the holding in *JBF RAK* that under section 777A(d)(1)(B) of the Act, Commerce is not required to take the additional step of considering alternate explanations for why there is a pattern of export prices that differ significantly among purchasers, regions, or time periods).

Thus, there is no requirement for Commerce to consider seasonal market conditions, *i.e.*, the alleged underlying reasons that price differences exist.

The Canadian Parties also argue that the “meaningful difference test, by design, does not consider whether ‘such differences’ can be taken into account using the {A-to-A} method adjusted for time periods” and that Commerce can do so by applying the A-to-A method using shorter time periods.<sup>697</sup> To the extent the Canadian Parties argue that Commerce’s meaningful difference test does not consider the nature of the significant differences identified using the Cohen’s *d* test, as explained above, the statute does not contain this requirement. Moreover, we note that the Federal Circuit has upheld the meaningful difference test as reasonable more than once.<sup>698</sup>

Further, the Canadian Parties’ argument that 19 CFR 351.414(d)(3) gives Commerce the flexibility to average over shorter time periods under the A-to-A method to account for price differences over time is inapposite. The regulation provides that when applying the A-to-A method in an investigation:

Commerce normally will calculate weighted averages for the entire period of investigation. However, when normal values, export prices, or constructed export prices differ significantly over the course of the period of investigation, {Commerce} may calculate weighted averages for such shorter period as the Secretary deems appropriate.<sup>699</sup>

This regulation applies to Commerce’s use of averages when applying the A-to-A comparison methodology to calculate dumping margins and is separate from the differential pricing

---

<sup>697</sup> See Canadian Parties Comments on the Draft Redetermination at 35 (citing 19 CFR 351.414(d)(3)).

<sup>698</sup> See *Stupp*, 5 F.4th at 1355 (affirming Commerce’s use of the meaningful difference test); see also *Apex II*, 862 F.3d at 1348 (“We hold that Commerce’s meaningful difference analysis ... was reasonable”).

<sup>699</sup> See 19 CFR 351.414(d)(3).

analysis.<sup>700</sup> The CIT has distinguished the diagnostic purpose of Commerce’s differential pricing analysis, which has been dictated from Congress’ grant of authority, from the purpose of dumping margin calculations to determine the potential uncollected dumping duty due.<sup>701</sup> Additionally, the regulation does not require that Commerce use shorter periods of time to account for alleged price volatility due to seasonal market conditions. As described above, the Federal Circuit has held that the statute contains no requirement that Commerce consider the underlying reasons that these price differences exist, and therefore, there is no reason to suggest that this regulation should be used as a method for Commerce to consider such reasons.

#### **IV. REJECTION OF RESOLUTE FP AND GOC 12/1 SUBMISSION**

We have considered Resolute FP and the GOC’s request that we revisit our decision to reject certain portions of the factual information attached to their December 1, 2023, submission. We decline to reverse our determination to reject voluminous factual information that was neither solicited by Commerce nor cited by the Federal Circuit in *Stupp, Mid Continent I*, or *Mid Continent II*. As Commerce explained in its memorandum rejecting this information, the Panel remanded to Commerce for further explanation of its use of the Cohen’s *d* test “in light of the {Federal Circuit’s} decisions in *Stupp, Mid Continent I* and *Mid Continent II*.”<sup>702</sup> Commerce’s placement on the record of the information relied on by the Federal Circuit in those opinions effectuated the Binational Panel Order.<sup>703</sup> Furthermore, Commerce’s allowance for Resolute FP

---

<sup>700</sup> See *Apex I*, 144 F. Supp. 3d at 1327-28 (“Plaintiffs’ argument misunderstands the function of the differential pricing analysis. {Section 351.414(d)(3) of Commerce’s regulations} is inapplicable in this context because it refers to Commerce’s use of averages in using the A-to-A comparison methodology to calculate dumping margins. The differential pricing analysis provides Commerce with a method to identify if a respondent’s sales exhibit a pattern of significant price differences, not calculate dumping margins. The regulation in no way restricts the time period over which Commerce calculates the weighted-averages it uses for purposes of finding significant price differences.”).

<sup>701</sup> See *Mid Continent IV*, 680 F. Supp. 3d at 1355-56.

<sup>702</sup> See Rejection of December 1, 2023 Submission at 1 (citations omitted).

<sup>703</sup> See Academic Texts from the Federal Circuit Memorandum at 1-2 (noting that Commerce is placing these texts on the record in light of the Panel’s Order).

and the GOC to submit additional information on the record that was before the Federal Circuit in subsequent remands in the ongoing proceedings also was consistent with the Binational Panel’s Order, as it will allow Commerce to adequately apply this intervening caselaw and consider all the information that was before the Federal Circuit in each phase of litigation in those cases.<sup>704</sup> Conversely, placing other information on the record that was nowhere referenced, cited, or relied upon by the Federal Circuit in *Stupp, Mid Continent I*, or *Mid Continent II*—as Resolute FP and GOC argue for—would not effectuate the Binational Panel Order that Commerce further explain its *Final Determination* in light of these intervening Federal Circuit remand orders.<sup>705</sup>

Moreover, supplementing the already extensive record with the voluminous information that Resolute FP and the GOC propose would significantly protract this segment of the proceeding. As the Federal Circuit has recognized, “{c}onstant reopening and supplementation of the record would lead to inefficiency and delay in finality.”<sup>706</sup> Commerce has a legitimate interest in controlling the bounds of the administrative record so that it can effectively administer the statute and its calculations.

Finally, while Resolute FP and the GOC suggest that Commerce’s declination to supplement the record with unsolicited information that was not before the Federal Circuit in any of the relevant cases raises due process concerns,<sup>707</sup> we disagree. Contrary to what Resolute FP and the GOC suggest, Commerce did not deprive them of the opportunity to respond to the information Commerce placed on the record. Indeed, Commerce specifically provided interested

---

<sup>704</sup> See Rejection of December 1, 2023 Submission at 2 (accepting the additional information submitted by Resolute FP and the GOC that was on the record before the Federal Circuit in *Stupp, Mid Continent I*, and *Mid Continent II*).

<sup>705</sup> We also note that the Binational Panel Order did not require that Commerce reopen the record at all, let alone require that it reopen it in the specific manner that Resolute FP and the GOC propose.

<sup>706</sup> See *Essar Steel Ltd. v. United States*, 678 F.3d 1268, 1277 (Fed. Cir. 2012) (*Essar Steel*).

<sup>707</sup> See Resolute FP and GOC 12/27 Submission at 4.

parties with an opportunity to comment on this information in its memorandum placing the publications cited by the Federal Circuit on the record.<sup>708</sup>

Commerce has “broad discretion regarding the manner in which it develops the record in an antidumping investigation.”<sup>709</sup> For the reasons described above, it was within Commerce’s discretion to refuse to accept a massive volume of information which was never cited by the Federal Circuit and which was never even on the records before the Federal Circuit in any of the cases cited in the Binational Panel Order.

### **Interested Parties’ Comments**

#### *COALITION’s Comments*<sup>710</sup>

Commerce reasonably exercised its discretion to reject from the record certain information submitted by the respondents. The Panel waived the respondents’ failure to exhaust arguments relating to the Federal Circuit’s decisions in *Stupp*, *Mid Continent I*, and *Mid Continent II* based on the exception for intervening judicial precedent, but that waiver does not entitle the respondents to present new information that was not considered in those appeals.

#### *Canadian Parties’ Comments*<sup>711</sup>

The Canadian Parties argue that Commerce’s refusal to allow the Canadian Parties to place on the record and subsequently rely on, analysis and literature relevant to Commerce’s differential pricing methodology is lawful. The materials that the Canadian Parties attempted to place on the record are not “new factual information” according to case law and Commerce’s

---

<sup>708</sup> See Academic Texts from the Federal Circuit Memorandum at 2.

<sup>709</sup> See *Stupp*, 5 F.4th at 1350 (citing *PSC VSMPO-Avisma Corp. v. United States*, 688 F.3d 751, 760 (Fed. Cir. 2012) (“{C}ourts will defer to the judgment of an agency regarding the development of the agency record.”); *Micron Tech.*, 117 F.3d 1386, 1396 (“Congress has implicitly delegated to Commerce the latitude to derive verification procedures ad hoc.”); and *Am. Alloys, Inc. v. United States*, 30 F.3d 1469, 1475 (Fed. Cir. 1994) (“{T}he statute gives Commerce wide latitude in its verification procedures.”).

<sup>710</sup> See COALITION Comments on the Draft Redetermination at 3.

<sup>711</sup> See Canadian Parties Comments on the Draft Redetermination at 7-8.

regulations.<sup>712</sup> Even if, *arguendo*, the materials were new factual information, Commerce's regulations and basic due process require Commerce to provide the Canadian Parties with the opportunity to submit information to clarify, correct, or rebut the information placed on the record by Commerce.<sup>713</sup> Commerce's claim that the Panel's remand order and the Federal Circuit decisions in *Stupp*, *Mid Continent I*, and *Mid Continent II* preclude Commerce from allowing the Canadian Parties to submit the materials is simply inaccurate. Nothing in the relevant decisions limits Commerce's consideration of any materials, and Commerce has not demonstrated otherwise.<sup>714</sup> Similarly, Commerce's claim that allowing the materials onto the record would unnecessarily prolong this proceeding lacks merit. Commerce has emphasized in filings to the Panel the importance of a just decision that considers all the substantive issues. This explanation is incongruent with Commerce's emphasis on the expediency of this remand redetermination. In any event, Commerce has reviewed the materials that the Canadian Parties attempted to place on the record in other segments of this proceeding and is well equipped to address them.

**Commerce's Position:**

We continue to find that our rejection of voluminous materials not on the record before the Federal Circuit or subsequently on the record of remand proceedings in any of the pertinent cases was reasonable and lawful.<sup>715</sup>

As discussed above, the Panel ordered Commerce to further explain its use of the Cohen's *d* test in light of the Federal Circuit's decisions in *Stupp*, *Mid Continent I*, and *Mid*

---

<sup>712</sup> *Id.* at 49.

<sup>713</sup> *Id.* at 49-50 (citing 19 CFR 351.301I(4)).

<sup>714</sup> *Id.* at 50.

<sup>715</sup> We note in this respect that Commerce did accept those materials submitted by Resolute FP and the GOC that were on the records before the Federal Circuit in *Stupp*, *Mid Continent I*, and *Mid Continent II*. See Rejection of December 1, 2023 Submission at 2; see also Resolute FP and GOC 12/27 Submission.

*Continent II*.<sup>716</sup> As part of that intervening precedent, the Federal Circuit cited certain academic literature in its holdings, which Commerce placed on the record in response to the Panel’s remand order to bring this case into alignment with *Stupp, Mid Continent I*, and *Mid Continent II*.<sup>717</sup> Furthermore, Commerce subsequently accepted additional documents on the record of this case which the GOC and Resolute FP provided and which Commerce determined had been on the records of those three cases.<sup>718</sup> However, Commerce also rejected other documents which the GOC and Resolute FP submitted in this case which were never considered by the Federal Circuit or CIT or placed by the parties on the record before the Courts in *Stupp, Mid Continent I*, and *Mid Continent II*.<sup>719</sup>

Supplementation of the record with materials not before the Federal Circuit or on remand before the CIT in those cases would be inconsistent with this order. None of the additional texts submitted by Resolute FP and the GOC were part of the Federal Circuit’s opinions in *Stupp, Mid Continent I*, and *Mid Continent II*, nor considered by the CIT subsequently in those cases on remand. Thus, these texts are beyond the scope of the Binational Panel Order. The supplemental texts were not considered by either the Federal Circuit or the interested parties in the proceedings before the Federal Circuit. Further, these texts have not been, and presently are not, before either the CIT or the Federal Circuit in the ongoing litigation since *Stupp, Mid Continent I*, and *Mid Continent II* were issued. Thus, we reasonably conclude that including these supplemental texts in this North American Free Trade Agreement (NAFTA) remand

---

<sup>716</sup> See Binational Panel Order at 30-32.

<sup>717</sup> See Academic Texts from the Federal Circuit Memorandum at 1-2 (“In *Stupp, Mid Continent I*, and *Mid Continent II*, the {Federal Circuit} quoted certain texts from the academic literature which are not on the administrative record of Commerce’s LTFV investigation. Commerce’s determination must be based on the administrative record before it. Accordingly, Commerce is reopening the record for the limited purpose of placing the publications cited by the {Federal Circuit} in these opinions on the administrative record.”).

<sup>718</sup> See Rejection of December 1, 2023 Submission at 2.

<sup>719</sup> *Id.*

segment would go beyond that Panel's remand order to consider the Federal Circuit's decisions and analyses in *Stupp, Mid Continent I*, and *Mid Continent II*.<sup>720</sup>

Accepting those documents in the first instance on remand in this litigation would also be at odds with the principle of finality.<sup>721</sup> The Binational Panel stated that it wished for Commerce to consider the arguments and findings in *Stupp, Mid Continent I*, and *Mid Continent II*, but it did not order Commerce to look beyond those cases for even further information. The claims advocated by the Canadian Parties would in fact add additional voluminous arguments and documents not considered or cited in those cases, without question. Had Commerce accepted those additional sources, it would have been in direct conflict with the principle of finality. Consequently, Commerce reasonably rejected those documents.

The Canadian Parties' due process argument is inapt.<sup>722</sup> Commerce reasonably addressed due process concerns by providing parties with an opportunity to comment on the materials it placed on the record.<sup>723</sup> Moreover, as explained above, Commerce took additional steps in providing Resolute FP and the GOC an opportunity to complete the record by accepting certain additional documents they submitted which had been before the Federal Circuit or the CIT on remand in the pertinent cases.<sup>724</sup>

Indeed, the Canadian Parties' due process argument is fundamentally flawed. Had Commerce accepted the massive volume of documents which were never seen or considered by the courts in the underlying cases, as the Canadian Parties now advocate, the other parties to this

---

<sup>720</sup> See Binational Panel Order at 29-32.

<sup>721</sup> See *Essar Steel*, 678 F.3d at 1277 (“Constant reopening and supplementation of the record would lead to inefficiency and delay in finality.”).

<sup>722</sup> See, e.g., Canadian Parties Comments on the Draft Redetermination at 49 (“... due process required {Commerce} to place the materials on the record.”).

<sup>723</sup> See Academic Texts from the Federal Circuit Memorandum at 2 (allowing for parties to comment on the factual information Commerce placed on the record).

<sup>724</sup> See Rejection of December 1, 2023 Submission at 2.

litigation would have been faced with information in first instance that was never introduced before either the courts or the interested parties in those cases. In other words, where the Binational Panel in this case seemed to have wanted the parties in this litigation to be similarly situated to the parties in *Stupp, Mid Continent I*, and *Mid Continent II*, had Commerce accepted the voluminous extra documentation submitted by the GOC and Resolute FP, such an acceptance would have been contrary to those wishes. It would have placed all the parties, other than the GOC and Resolute FP, in the unenviable situation of having to read, comprehend, consider, challenge, and respond to all that voluminous information in the first instance in this case on remand. Commerce finds that accepting such data at this point in the case would have been unacceptable, and, far from serving the interest of due process, it would have undermined it. Thus, there is no merit to the Canadian Parties' claim in this regard.

The Canadian Parties' regulatory argument is likewise unavailing.<sup>725</sup> Commerce's regulations do not, by their very text, apply directly to remand redeterminations.<sup>726</sup> Frequently, Commerce looks to those regulations for guidance on certain matters, but for procedural requirements, 19 CFR 351.301 does not apply. Furthermore, neither the Act nor regulations state that the responsive factual information requirements under 19 CFR 351.301(c)(4) apply when Commerce places information on the record pursuant to the order of a court or NAFTA Panel.<sup>727</sup> Notably, the Canadian's Parties cite no authoritative source for their claims in this regard.

---

<sup>725</sup> See, e.g., Canadian Parties Comments on the Draft Redetermination at 49-50 ("... {Commerce's} regulations ... required {Commerce} to place the materials on the record") (citing 19 CFR 351.301(c)(4)).

<sup>726</sup> Indeed, it is an illogical presumption that Commerce's regulations always apply directly to remand redeterminations given the number of regulatory provisions that refer to specifically individual segments, such as investigations and administrative reviews, while no regulation claims to apply to a remand redetermination segment of a proceeding.

<sup>727</sup> It is an odd argument to claim that if Commerce is abiding by a court's or panel's remand order, and pursuant to that order places certain information on the record, that Commerce's regulations would allow parties to go beyond such a remand order, and place information on the record that may be contrary to the court's or panel's order. Commerce's regulations are not a means by which parties may avoid the disciplines of a court's or panel's order and Commerce will not interpret its regulations in such a manner.

Further, even if Commerce’s regulations did apply directly to remand redeterminations, any interpretation of those regulations would need to be read in light of Commerce’s discretion and legitimate interests with respect to the record. As we stated in the Draft Redetermination, Commerce has broad discretion in developing the record,<sup>728</sup> and a legitimate interest in controlling the bounds of that record so that it can effectively administer the statute. The regulation should not be read as permitting voluminous expansion of the record that would hinder Commerce’s ability to administer the statute and, again, be inconsistent with the Panel’s Order.

Commerce reopened the record for the limited purpose of adding certain academic texts on the record before the Federal Circuit and the CIT in *Stupp, Mid Continent I*, and *Mid Continent II*, while also providing the parties an opportunity to comment on those texts, and even add additional text that was before the CIT in those cases upon remand.<sup>729</sup> This approach struck a reasonable balance between effectuating the Binational Panel Order, due process concerns, and the need to avoid wholesale remaking or expansion of the record in the interest of finality. Additional supplementation with voluminous materials not before the Federal Circuit and CIT in the pertinent cases, as Resolute FP and the GOC urged, would have upset this reasonable balance and would have been inconsistent with the Binational Panel Order. Consequently, Commerce reasonably rejected those documents from being presented for the first time on the record of this proceeding.

---

<sup>728</sup> See *Stupp*, 5 F.4th at 1350 (“Commerce is entitled to broad discretion regarding the manner in which it develops the record in an antidumping investigation”) (citations omitted).

<sup>729</sup> See Academic Texts from the Federal Circuit Memorandum; see also Rejection of December 1, 2023 Submission at 2 (accepting the additional information submitted by Resolute FP and the GOC that was on the record before the Federal Circuit and subsequently lower courts on remand in *Stupp, Mid Continent I*, and *Mid Continent II*).

## V. FINAL RESULTS OF REDETERMINATION

In accordance with the Binational Panel Order, Commerce has reconsidered the issues remanded by the Panel, and for the reasons explained above, we have: (1) determined not to deduct 2006 SLA export tax from U.S. price; (2) provided additional support for our use of log inputs to determine when Resolute FP reached commercial production levels at its Atikokan Mill; and (3) addressed the Federal Circuit’s remand orders in *Stupp*, *Mid Continent I*, and *Mid Continent II*. We have also concluded that: (1) Commerce need not observe the three statistical criteria as part of the Cohen’s *d* test; and (2) Commerce’s use of a simple average to calculate the denominator of the Cohen’s *d* coefficient is reasonable.

After making the change concerning the 2006 SLA export tax in the Draft Redetermination, we calculated an estimated weighted-average dumping margin of 6.63 percent for Canfor,<sup>730</sup> 3.08 percent for Resolute,<sup>731</sup> 7.14 percent for Tolko,<sup>732</sup> and 5.18 percent for West Fraser.<sup>733</sup> In the Draft Redetermination, Commerce calculated individual estimated weighted-average dumping margins for Canfor, Resolute, Tolko, and West Fraser, none of which are zero, *de minimis*, or based entirely on facts otherwise available. Thus, here, Commerce calculated the

---

<sup>730</sup> In the *Final Determination*, Commerce determined that Canfor Corporation, Canadian Forest Products Ltd., and Canfor Wood Products Marketing Ltd are a single entity (Canfor). For Canfor’s margin calculation, *see* Memorandum, “Draft Results of Redetermination Pursuant to Court Order: Analysis Memorandum for Canfor,” dated February 20, 2024.

<sup>731</sup> *See* Memorandum, “Draft Results of Redetermination Pursuant to Court Order: Analysis Memorandum for Resolute,” dated February 20, 2024.

<sup>732</sup> In the *Final Determination*, Commerce determined that Tolko Marketing and Sales Ltd., Tolko Industries Ltd., and Gilbert Smith Forest Products Ltd. are a single entity (Tolko). For Tolko’s margin calculation, *see* Memorandum, “Draft Results of Redetermination Pursuant to Court Order: Analysis Memorandum for Tolko,” dated February 20, 2024.

<sup>733</sup> In the *Final Determination*, Commerce determined that West Fraser Mills Ltd., Blue Ridge Lumber Inc., Manning Forest Products Ltd., and Sundre Forest Products Inc. are a single entity (West Fraser). For West Fraser’s margin calculation, *see* Memorandum, “Draft Results of Redetermination Pursuant to Court Order: Analysis Memorandum for Tolko,” dated February 20, 2024.

all-others rate of 5.66 percent using a weighted average of the estimated weighted-average dumping margins calculated for the mandatory respondents.<sup>734</sup>

For these final results of redetermination, we have made no changes to the margin calculations from the Draft Redetermination.

4/30/2024

X 

---

Signed by: RYAN MAJERUS  
Ryan Majerus  
Deputy Assistant Secretary  
for Policy and Negotiations,  
performing the non-exclusive functions and duties  
of the Assistant Secretary for Enforcement and Compliance

---

<sup>734</sup> See Memorandum, "Calculation of the All-Others Rate," dated February 20, 2024.

# Attachment I

U.S. Less-Than-Fair-Value Investigations  
Final Determinations - Calendar Year 2015

Case Number	Country	Product	Preliminary		Final		Amended Final		Comparison Method	Final Rate	
			FR Pub Date	FR Citation	FR Pub Date	FR Citation	FR Pub Date	FR Citation			Company
A-570-014	China	53-Foot Domestic Dry Containers	11/26/2014	79 FR 70501	4/17/2015	80 FR 21203			Singamas	A-to-A	111.22
A-570-014	China	53-Foot Domestic Dry Containers	11/26/2014	79 FR 70501	4/17/2015	80 FR 21203			China-Wide Entity (CIMC)	A-to-A	107.19
A-580-874	Korea, Rep	Steel Nails	12/29/2014	79 FR 78051	5/20/2015	80 FR 28955			Daejin Steel	A-to-A	11.80
A-580-874	Korea, Rep	Steel Nails	12/29/2014	79 FR 78051	5/20/2015	80 FR 28955			Jinheung Steel	A-to-A	0.00
A-557-816	Malaysia	Steel Nails	12/29/2014	79 FR 78055	5/20/2015	80 FR 28969			Inmax	AFA	39.35
A-557-816	Malaysia	Steel Nails	12/29/2014	79 FR 78055	5/20/2015	80 FR 28969	6/16/2015	80 FR 34370	Region International	A-to-T	2.66
A-557-816	Malaysia	Steel Nails	12/29/2014	79 FR 78055	5/20/2015	80 FR 28969			Tag Fasteners	AFA	39.35
A-523-808	Oman	Steel Nails	12/29/2014	79 FR 78034	5/20/2015	80 FR 28972			Oman Fasteners	A-to-A	9.10
A-583-854	Taiwan	Steel Nails	12/29/2014	79 FR 78053	5/20/2015	80 FR 28959			Quick Advance	A-to-A	0.00
A-583-854	Taiwan	Steel Nails	12/29/2014	79 FR 78053	5/20/2015	80 FR 28959			PT Enterprises	mixed	2.24
A-552-818	Vietnam	Steel Nails	12/29/2014	79 FR 78058	5/20/2015	80 FR 29622			Region International	AFA	323.99
A-552-818	Vietnam	Steel Nails	12/29/2014	79 FR 78058	5/20/2015	80 FR 29622			United Nail Products	AFA	323.99
A-570-016	China	Passenger Vehicle and Light Truck Tires	1/27/2015	80 FR 4250	6/18/2015	80 FR 34893	8/10/2015	80 FR 47902	Giti Tire	A-to-A	30.74
A-570-016	China	Passenger Vehicle and Light Truck Tires	1/27/2015	80 FR 4250	6/18/2015	80 FR 34893			Sailun Group	A-to-A	14.35
A-570-018	China	Boltless Steel Shelving Units	4/1/2015	80 FR 17409	8/26/2015	80 FR 51779			Zhongda	A-to-A	17.55
A-201-845	Mexico	Sugar	11/3/2014	79 FR 65189	9/23/2015	80 FR 57341			FEESA	A-to-A	40.48
A-201-845	Mexico	Sugar	11/3/2014	79 FR 65189	9/23/2015	80 FR 57341			GAM Group	A-to-A	42.14
A-580-876	Korea, Rep	Welded Line Pipe	5/22/2015	80 FR 29620	10/13/2015	80 FR 61366	11/10/2015	80 FR 69637	Hyundai HYSCO	A-to-T	6.23
A-580-876	Korea, Rep	Welded Line Pipe	5/22/2015	80 FR 29620	10/13/2015	80 FR 61366			SeAH Steel	mixed	2.53
A-489-822	Turkey	Welded Line Pipe	5/22/2015	80 FR 29617	10/13/2015	80 FR 61362			Borusan Istikbal	AFA	22.95
A-489-822	Turkey	Welded Line Pipe	5/22/2015	80 FR 29617	10/13/2015	80 FR 61362			Borusan Mannesmann	AFA	22.95
A-489-822	Turkey	Welded Line Pipe	5/22/2015	80 FR 29617	10/13/2015	80 FR 61362			Cayirova/Yucel	A-to-A	22.95
A-489-822	Turkey	Welded Line Pipe	5/22/2015	80 FR 29617	10/13/2015	80 FR 61362			Toscelik	A-to-A	6.66
A-570-020	China	Melamine	6/18/2015	80 FR 34891	11/6/2015	80 FR 68851			Allied	AFA	363.31
A-570-020	China	Melamine	6/18/2015	80 FR 34891	11/6/2015	80 FR 68851			Golden Elephant	AFA	363.31
A-570-020	China	Melamine	6/18/2015	80 FR 34891	11/6/2015	80 FR 68851			Xinji Jiuyuan	AFA	363.31
A-274-806	Trinidad & Tobago	Melamine	6/17/2015	80 FR 34621	11/6/2015	80 FR 68846			MHTL	A-to-A	172.53

## Attachment II

**U.S. Less-Than-Fair-Value Investigations  
Final Determinations - Calendar Year 2021**

Case Number	Country	Product	Preliminary		Final		Amended Final		Company	Comparison Method	Final Rate
			FR Pub Date	FR Citation	FR Pub Date	FR Citation	FR Pub Date	FR Citation			
A-351-853	Brazil	Wood Mouldings and Millwork Products	8/12/2020	85 FR 48667	1/4/2021	86 FR 70			Arupel	A-to-A	0.00
A-570-117	China	Wood Mouldings and Millwork Products	8/12/2020	85 FR 48669	1/4/2021	86 FR 63	2/16/2021	86 FR 9486	Fujian Yinfeng	A-to-A	45.49
A-570-119	China	Vertical Shaft Engines 225cc to 999cc	8/19/2020	85 FR 51015	1/11/2021	86 FR 1936	3/4/2021	86 FR 12623	Loncin Motor	A-to-A	185.65
A-570-119	China	Vertical Shaft Engines 225cc to 999cc	8/19/2020	85 FR 51015	1/11/2021	86 FR 1936			Zongshen	A-to-A	336.26
A-570-121	China	Diffluoromethane (R-32)	8/27/2020	85 FR 52950	1/19/2021	86 FR 5136			Taizhou Qingsong	A-to-A	161.49
A-570-121	China	Diffluoromethane (R-32)	8/27/2020	85 FR 52950	1/19/2021	86 FR 5136			Zibo Feiyuan	A-to-A	221.06
A-570-122	China	Corrosion Inhibitors	9/10/2020	85 FR 55825	1/29/2021	86 FR 7532			Jiangyin Delian	A-to-A	130.52
A-570-122	China	Corrosion Inhibitors	9/10/2020	85 FR 55825	1/29/2021	86 FR 7532			Nantong Botao	A-to-A	139.41
A-570-131	China	Twist Ties	12/10/2020	85 FR 79468	2/22/2021	86 FR 10536			Zhenjiang Hongda	AFA	72.96
A-570-131	China	Twist Ties	12/10/2020	85 FR 79468	2/22/2021	86 FR 10536			Zhenjiang Zhonglian	AFA	72.96
A-580-907	Korea, Rep	Ultra-High Polyethylene	10/6/2020	85 FR 63095	2/25/2021	86 FR 11497			Korea Petrochemical	A-to-A	7.84
A-893-001	Bosnia & Herzegovina	Silicon Metal	12/11/2020	85 FR 80009	2/26/2021	86 FR 11720			R-S Soycop D.O.O.	AFA	21.41
A-400-001	Iceland	Silicon Metal	12/11/2020	85 FR 80009	2/26/2021	86 FR 11720			PCC Bakki Silicon	AFA	47.54
A-851-804	Czech Rep	Seamless Standard, Line and Pressure Pipe	12/21/2021	85 FR 83059	3/5/2021	86 FR 12909			Liberty Ostrava	AFA	51.70
A-851-804	Czech Rep	Seamless Standard, Line and Pressure Pipe	12/21/2021	85 FR 83059	3/5/2021	86 FR 12909			Moravia Steel	AFA	51.70
A-525-001	Bahrain	Aluminum Sheet	10/15/2020	85 FR 65372	3/8/2021	86 FR 13331			Gulf Aluminum	A-to-A	4.83
A-351-854	Brazil	Aluminum Sheet	10/15/2020	85 FR 65363	3/8/2021	86 FR 13302			CBA	AFA	137.06
A-351-854	Brazil	Aluminum Sheet	10/15/2020	85 FR 65363	3/8/2021	86 FR 13302			Novelis	A-to-A	49.61
A-891-001	Croatia	Aluminum Sheet	10/15/2020	85 FR 65384	3/8/2021	86 FR 13312			Impol	A-to-T	3.19
A-729-803	Egypt	Aluminum Sheet	10/15/2020	85 FR 65382	3/8/2021	86 FR 13324			Egypt Alum	A-to-A	12.11
A-428-849	Germany	Aluminum Sheet	10/15/2020	85 FR 65386	3/8/2021	86 FR 13318			Hydro Aluminum	AFA	242.80
A-428-849	Germany	Aluminum Sheet	10/15/2020	85 FR 65386	3/8/2021	86 FR 13318			Novelis	A-to-A	49.40
A-484-804	Greece	Aluminum Sheet	10/15/2020	85 FR 65374	3/8/2021	86 FR 13300			Elval Hellenic	A-to-A	0.00
A-533-895	India	Aluminum Sheet	10/15/2020	85 FR 65377	3/8/2021	86 FR 13282			Hindalco	AFA	47.92
A-533-895	India	Aluminum Sheet	10/15/2020	85 FR 65377	3/8/2021	86 FR 13282			Manaksia	A-to-A	0.00
A-560-835	Indonesia	Aluminum Sheet	10/15/2020	85 FR 65356	3/8/2021	86 FR 13304			PT Alumindo	AFA	32.12
A-475-842	Italy	Aluminum Sheet	10/15/2020	85 FR 65342	3/8/2021	86 FR 13309			Laminazione Sottile	A-to-A	0.00
A-475-842	Italy	Aluminum Sheet	10/15/2020	85 FR 65342	3/8/2021	86 FR 13309			Profilglass	AFA	29.13
A-523-814	Oman	Aluminum Sheet	10/15/2020	85 FR 65340	3/8/2021	86 FR 13328			Oman Aluminum	A-to-A	5.29
A-485-809	Romania	Aluminum Sheet	10/15/2020	85 FR 65358	3/8/2021	86 FR 13320			Alro	AFA	37.26
A-801-001	Serbia	Aluminum Sheet	10/15/2020	85 FR 65386	3/8/2021	86 FR 13295			Impol	A-to-A	11.67
A-801-001	Serbia	Aluminum Sheet	10/15/2020	85 FR 65386	3/8/2021	86 FR 13295			Otocivi Doo	AFA	25.84
A-856-001	Slovenia	Aluminum Sheet	10/15/2020	85 FR 65349	3/8/2021	86 FR 13305			Impol	A-to-A	13.43
A-791-825	South Africa	Aluminum Sheet	10/15/2020	85 FR 65351	3/8/2021	86 FR 13287			Hulamin	A-to-A	8.85
A-580-906	Korea, Rep	Aluminum Sheet	10/15/2020	85 FR 65354	3/8/2021	86 FR 13307			Novelis	A-to-A	0.00
A-469-820	Spain	Aluminum Sheet	10/15/2020	85 FR 65367	3/8/2021	86 FR 13298			Aludium Transformacion	AFA	3.80
A-469-820	Spain	Aluminum Sheet	10/15/2020	85 FR 65367	3/8/2021	86 FR 13298			Valenciana	AFA	24.23
A-583-867	Taiwan	Aluminum Sheet	10/15/2020	85 FR 65361	3/8/2021	86 FR 13293			CS Aluminum	A-to-A	17.50
A-489-839	Turkey	Aluminum Sheet	10/15/2020	85 FR 65346	3/8/2021	86 FR 13326			Assan	A-to-T	2.02
A-489-839	Turkey	Aluminum Sheet	10/15/2020	85 FR 65346	3/8/2021	86 FR 13326			Teknik	A-to-A	13.56
A-570-124	China	Vertical Shaft Engines 99cc to 225cc	10/21/2020	85 FR 66932	3/12/2021	86 FR 14077			Kohler Engines	A-to-A	374.31
A-570-124	China	Vertical Shaft Engines 99cc to 225cc	10/21/2020	85 FR 66932	3/12/2021	86 FR 14077			Zongshen	A-to-A	316.88
A-570-126	China	Non-Refillable Cylinders	10/30/2020	85 FR 68852	3/22/2021	86 FR 15188			Sanjiang	A-to-A	93.09
A-570-126	China	Non-Refillable Cylinders	10/30/2020	85 FR 68852	3/22/2021	86 FR 15188			Wuyi Xilinde	A-to-A	74.33
A-555-001	Cambodia	Mattresses	11/3/2020	85 FR 69594	3/25/2021	86 FR 15894	5/14/2021	86 FR 26460	Best Mattresses	A-to-A	52.41
A-560-836	Indonesia	Mattresses	11/3/2020	85 FR 69597	3/25/2021	86 FR 15899			Zinus Global	mixed	2.22
A-557-818	Malaysia	Mattresses	11/3/2020	85 FR 69574	3/25/2021	86 FR 15901			Delandis	AFA	42.92
A-557-818	Malaysia	Mattresses	11/3/2020	85 FR 69574	3/25/2021	86 FR 15901			Far East Foam	AFA	42.92
A-557-818	Malaysia	Mattresses	11/3/2020	85 FR 69574	3/25/2021	86 FR 15901			Vision Foam	AFA	42.92
A-801-002	Serbia	Mattresses	11/3/2020	85 FR 69589	3/25/2021	86 FR 15892			Healthcare Europe	A-to-A	112.11
A-549-841	Thailand	Mattresses	11/3/2020	85 FR 69568	3/25/2021	86 FR 15928			Nisco (Thailand)	AFA	763.28
A-549-841	Thailand	Mattresses	11/3/2020	85 FR 69568	3/25/2021	86 FR 15928			Saffron Living	A-to-A	37.48
A-489-841	Turkey	Mattresses	11/3/2020	85 FR 69571	3/25/2021	86 FR 15917			BRN Yatak	A-to-A	20.03
A-552-827	Vietnam	Mattresses	11/3/2020	85 FR 69591	3/25/2021	86 FR 15889			Ashley Group	A-to-A	144.92
A-552-827	Vietnam	Mattresses	11/3/2020	85 FR 69591	3/25/2021	86 FR 15889			Vietnam Glory	AFA	668.38
A-560-837	Indonesia	Prestressed Concrete Steel Wire Strand	11/19/2020	85 FR 73676	4/9/2021	86 FR 18495			PT Kingdom Indah	A-to-A	5.76
A-560-837	Indonesia	Prestressed Concrete Steel Wire Strand	11/19/2020	85 FR 73676	4/9/2021	86 FR 18495			PT Bumi Steel	AFA	72.28
A-475-843	Italy	Prestressed Concrete Steel Wire Strand	11/19/2020	85 FR 73679	4/9/2021	86 FR 18505			WBO Italcables	mixed	3.59
A-475-843	Italy	Prestressed Concrete Steel Wire Strand	11/19/2020	85 FR 73679	4/9/2021	86 FR 18505			CB Trafilati	AFA	19.26
A-557-819	Malaysia	Prestressed Concrete Steel Wire Strand	11/19/2020	85 FR 73685	4/9/2021	86 FR 18502			Kiswire	A-to-T	3.94
A-557-819	Malaysia	Prestressed Concrete Steel Wire Strand	11/19/2020	85 FR 73685	4/9/2021	86 FR 18502			Southern PC Steel	AFA	26.95
A-557-819	Malaysia	Prestressed Concrete Steel Wire Strand	11/19/2020	85 FR 73685	4/9/2021	86 FR 18502			Wei Dat Steel Wire	A-to-T	6.42
A-791-826	South Africa	Prestressed Concrete Steel Wire Strand	11/19/2020	85 FR 73674	4/9/2021	86 FR 18497			Scaw Metals	AFA	155.10
A-469-821	Spain	Prestressed Concrete Steel Wire Strand	11/19/2020	85 FR 73683	4/9/2021	86 FR 18512			TYCSA	A-to-A	14.75
A-723-001	Tunisia	Prestressed Concrete Steel Wire Strand	11/19/2020	85 FR 73681	4/9/2021	86 FR 18508			Maklada	A-to-A	30.58
A-823-817	Ukraine	Prestressed Concrete Steel Wire Strand	11/19/2020	85 FR 73688	4/9/2021	86 FR 18498			PISC	A-to-A	19.30
A-427-831	France	Methionine	3/4/2021	86 FR 12627	5/17/2021	86 FR 26697			Adisseo	AFA	43.82
A-570-135	China	Certain Chassis and Subassemblies Thereof	3/4/2021	86 FR 12616	5/17/2021	86 FR 26694			CIMC	AFA	188.05
A-570-135	China	Certain Chassis and Subassemblies Thereof	3/4/2021	86 FR 12616	5/17/2021	86 FR 26694			Fuwa	AFA	188.05
A-570-129	China	Lawn Mowers	12/30/2020	85 FR 86529	5/20/2021	86 FR 27384			Nigbo Daye	A-to-A	98.73
A-552-830	Vietnam	Lawn Mowers	12/30/2020	85 FR 86534	5/20/2021	86 FR 27382			Ducar Technology	A-to-A	148.35
A-580-908	Korea, Rep	Passenger and Light Truck Tires	1/6/2021	86 FR 501	5/27/2021	86 FR 28569			Hankook	A-to-A	27.05
A-580-908	Korea, Rep	Passenger and Light Truck Tires	1/6/2021	86 FR 501	5/27/2021	86 FR 28569			Nexen	mixed	14.72
A-583-869	Taiwan	Passenger and Light Truck Tires	1/6/2021	86 FR 508	5/27/2021	86 FR 28563			Cheng Shin Rubber	A-to-T	20.04
A-583-869	Taiwan	Passenger and Light Truck Tires	1/6/2021	86 FR 508	5/27/2021	86 FR 28563			Nankang Rubber	A-to-A	101.84
A-549-842	Thailand	Passenger and Light Truck Tires	1/6/2021	86 FR 517	5/27/2021	86 FR 28548			LLT	A-to-A	21.09
A-549-842	Thailand	Passenger and Light Truck Tires	1/6/2021	86 FR 517	5/27/2021	86 FR 28548	7/19/2021	86 FR 38011	Sumitomo Rubber	mixed	14.59
A-552-828	Vietnam	Passenger and Light Truck Tires	1/6/2021	86 FR 504	5/27/2021	86 FR 28559			Kenda Rubber	A-to-A	0.00
A-552-828	Vietnam	Passenger and Light Truck Tires	1/6/2021	86 FR 504	5/27/2021	86 FR 28559			Sailun	A-to-A	0.00
A-201-853	Mexico	Standard Steel Welded Wire Mesh	2/1/2021	86 FR 7710	6/23/2021	86 FR 32891			Aceromex	A-to-A	23.04
A-201-853	Mexico	Standard Steel Welded Wire Mesh	2/1/2021	86 FR 7710	6/23/2021	86 FR 32891			Deacero	AFA	110.42
A-557-820	Malaysia	Silicon Metal	2/1/2021	86 FR 7701	6/24/2021	86 FR 33224			PMB Silicon	A-to-A	12.27
A-469-823	Spain	Wind Towers	4/2/2021	86 FR 17354	6/25/2021	86 FR 33656			Vestas Eolica	AFA	73.00
A-552-831	Vietnam	Seamless Refined Copper Pipe and Tube	2/1/2021	86 FR 7698	6/24/2021	86 FR 33228			Haliang Vietnam	mixed	8.35
A-580-909	Korea, Rep	Seamless Standard, Line and Pressure Pipe	2/10/2021	86 FR 8887	7/2/2021	86 FR 35274			ILJIN Steel	A-to-T	4.48
A-821-826	Russia	Seamless Standard, Line and Pressure Pipe	2/10/2021	86 FR 8891	7/2/2021	86 FR 35269			TMK	A-to-A	209.72
A-823-819	Ukraine	Seamless Standard, Line and Pressure Pipe	2/10/2021	86 FR 8889	7/2/2021	86 FR 35272			Interpipe	A-to-A	23.75
A-570-133	China	Certain Metal Lockers and Parts Thereof	2/11/2021	86 FR 9051	7/7/2021	86 FR 35737			Hang		

# Attachment III

Resolute FP Rule 57(3) Reply Brief  
Appendix B

Case Number	Country	Product	Period	Preliminary		Amended Preliminary		Final		Amended Final	
				FR Pub Date	FR Citation	FR Pub Date	FR Citation	FR Pub Date	FR Citation	Fr Pub Date	FR Citation
A-122-853	Canada	Citric Acid and Salts	POR 13/14					10/15/2015	80 FR 62016		
A-122-853	Canada	Citric Acid and Salts	POR 14/15	2/12/2016	81 FR 7500			5/10/2016	81 FR 28827		
A-122-855	Canada	PET Resin	POI	10/15/2015	80 FR 62019			3/14/2016	81 FR 13319		
A-201-830	Mexico	Steel Wire Rod	POR 13/14	11/10/2015	80 FR 69641			5/19/2016	81 FR 31592		
A-201-836	Mexico	Light-Walled Rect Pipe and Tube	POR 13/14					11/12/2015	80 FR 69941		
A-201-838	Mexico	Copper Pipe & Tubing	POR 13/14					11/12/2015	80 FR 69944		
A-201-842	Mexico	Washers	POR 14/15	3/11/2016	81 FR 12873			9/12/2016	81 FR 62714		
A-201-843	Mexico	Prestressed Concrete Steel Rail Tie Wire	POR 13/15	3/9/2016	81 FR 12466			6/23/2016	81 FR 40850		
A-201-847	Mexico	Heavy Walled Rect Pipe & Tube	POI	3/1/2016	81 FR 10587			7/21/2016	81 FR 47352		
A-274-806	Trinidad & Tobago	Melamine	POI					11/6/2015	80 FR 68846		
A-351-825	Brazil	Stainless Steel Bar	POR 14/15	3/9/2016	81 FR 12465			6/22/2016	81 FR 40670		
A-351-842	Brazil	Uncoated Paper	POI					1/20/2016	81 FR 3115		
A-351-843	Brazil	CR Steel Flat Products	POI	3/7/2016	81 FR 11754	4/7/2016	81 FR 20366	7/29/2016	81 FR 44946		
A-351-845	Brazil	HR Steel Flat Products	POI	3/22/2016	81 FR 15235			8/12/2016	81 FR 53424		
A-412-824	UK	CR Steel Flat Products	POI	3/7/2016	81 FR 11744			7/29/2016	81 FR 49929		
A-412-825	UK	HR Steel Flat Products	POI	3/22/2016	81 FR 15244			8/12/2016	81 FR 53436		
A-421-813	Netherlands	HR Steel Flat Products	POI	3/22/2016	81 FR 15225			8/12/2016	81 FR 53421		
A-471-807	Portugal	Uncoated Paper	POI					1/20/2016	81 FR 3105		
A-475-818	Italy	Pasta	POR 13/14					2/17/2016	81 FR 8043	3/7/2016	81 FR 12690
A-475-818	Italy	Pasta	POR 14/15	8/12/2016	81 FR 53404						
A-475-828	Italy	SS Butt Welded Pipe Fittings	POR 14/15	2/25/2016	81 FR 9806			7/5/2016	81 FR 43587		
A-475-832	Italy	Corrosion Resistant Steel Products	POI	1/4/2016	81 FR 69			6/2/2016	81 FR 35320		
A-489-501	Turkey	Standard Pipe & Tube	POR 13/14	6/13/2016	81 FR 38131			12/10/2015	80 FR 76674		
A-489-815	Turkey	Light Walled Rect Pipe & Tube	POR 14/15	2/12/2016	81 FR 7503			5/10/2016	81 FR 28823		
A-489-824	Turkey	Heavy Walled Rect Pipe & Tube	POI	3/1/2006	81 FR 10583			7/21/2016	81 FR 47355		
A-489-826	Turkey	HR Steel Flat Products	POI	3/22/2016	81 FR 15231			8/12/2016	81 FR 53428		
A-520-803	UAE	PET Film	POR 13/14	12/1/2015	80 FR 75052			4/11/2016	81 FR 21314		
A-520-804	UAE	Nails	POR 14/15	6/10/2016	81 FR 37571						
A-520-807	UAE	Circular Welded Pipe	POI	6/8/2016	81 FR 36881						
A-522-801	Vietnam	Frozen Fish Fillets	POR 13/14					3/19/2016	81 FR 17435		
A-523-810	Oman	PET Resin	POI	10/15/2015	80 FR 62021			3/14/2016	81 FR 13336		
A-523-812	Oman	Circular Welded Pipe	POI	6/8/2016	81 FR 36873						
A-533-810	India	Stainless Steel Bar	POR 14/15					9/8/2016	81 FR 62086		
A-533-810	India	SS Bar	POR 14/15	3/10/2016	81 FR 12694						
A-533-813	India	Mushrooms	POR 14/15	3/9/2016	81 FR 12463			9/8/2016	81 FR 62081		
A-533-823	India	Silicomanganese	POR 13/14					12/3/2015	80 FR 75660		
A-533-824	India	PET Film	POR 13/14	8/2/2016	81 FR 50684			2/16/2016	81 FR 7750		
A-533-840	India	Shrimp	POR 14/15	3/10/2016	81 FR 12705			9/13/2016	81 FR 62867		
A-533-843	India	Lined Paper Products	POR 13/14	10/7/2015	80 FR 60628			2/4/2016	81 FR 5986		
A-533-861	India	PET Resin	POI	10/15/2015	80 FR 62029			3/14/2016	81 FR 13327		
A-533-863	India	Corrosion Resistant Steel Products	POI	1/4/2016	81 FR 63			6/2/2016	81 FR 35329		
A-533-865	India	CR Steel Flat Products	POI	3/7/2016	81 FR 11741			7/29/2016	81 FR 49938		
A-533-867	India	Welded Stainless Pressure Pipe	POI	5/10/2016	81 FR 28824			9/29/2016	81 FR 66921		
A-533-869	India	Off-the-Road Tires	POI	8/19/2016	81 FR 55431						
A-549-502	Thailand	Circular Welded Pipe and Tube	POR 13/14					10/2/2015	80 FR 59732		
A-549-822	Thailand	Shrimp	POR 14/15	3/10/2016	81 FR 12696			6/22/2016	81 FR 40671		
A-552-801	Vietnam	Frozen Fish Fillets	POR 14/15	2/1/2016	81 FR 5709			7/7/2016	81 FR 44272		
A-552-802	Vietnam	Shrimp	POR 14/15	3/10/2016	81 FR 12702			9/12/2016	81 FR 62717		
A-552-820	Vietnam	Circular Welded Pipe	POI	6/8/2016	81 FR 36884						
A-557-813	Malaysia	Plastic Bags	POR 14/15	6/24/2016	81 FR 41294						
A-570-001	China	Potassium Permanganate	POR 14/14	2/16/2016	81 FR 7751						
A-570-022	China	Uncoated Paper	POI					1/20/2016	81 FR 3112		
A-570-024	China	PET Resin	POI	10/15/2015	80 FR 62024			3/14/2016	81 FR 13331		
A-570-026	China	Corrosion Resistant Steel Products	POI	1/4/2016	81 FR 75			6/2/2016	81 FR 35316		
A-570-028	China	Hydrofluorocarbon Blends	POI	2/1/2016	81 FR 5098			6/29/2016	81 FR 42314		
A-570-032	China	Mechanical Transfer Drive Components	POI	6/8/2016	81 FR 36876						
A-570-033	China	Washers	POI	7/26/2016	81 FR 48741						
A-570-036	China	Biaxial Integral Biogrid Products	POI	8/22/2016	81 FR 56584						
A-570-038	China	Amorphous Silica Fabric	POI	9/1/2016	81 FR 60341						
A-570-040	China	Truck and Bus Tires	POI	9/6/2016	81 FR 61186						
A-570-601	China	TRBs	POR 13/14	7/14/2016	81 FR 45455			1/12/2016	81 FR 1396		
A-570-827	China	Pencils	POR 14/15	6/10/2016	81 FR 37573						
A-570-831	China	Garlic	POR 13/14	12/7/2015	80 FR 75972			6/20/2016	81 FR 39897		
A-570-848	China	Crawfish	POR 13/14	10/7/2015	80 FR 60624			4/13/2016	81 FR 21840		
A-570-898	China	Chlorinated Isocyanurates	POR 13/14	7/12/2016	81 FR 45128			1/11/2016	81 FR 1167		
A-570-900	China	Diamond Sawblades	POR 13/14	12/4/2015	80 FR 75854			6/8/2016	81 FR 38673		
A-570-904	China	Activated Carbon	POR 13/14					10/9/2015	80 FR 61172		
A-570-904	China	Activated Carbon	POR 14/15	3/4/2016	81 FR 11513			9/8/2016	81 FR 62088		
A-570-909	China	Nails	POR 13/14	9/12/2016	81 FR 62710			3/16/2016	81 FR 14092		
A-570-912	China	Off-the-Road Tires	POR 13/14	10/9/2015	80 FR 61166			4/20/2016	81 FR 23272		
A-570-929	China	Graphite Electrodes	POR 14/15	3/9/2016	81 FR 12468			9/9/2016	81 FR 62474		
A-570-932	China	Steel Threaded Rod	POR 13/14	5/13/2016	81 FR 29843			11/12/2015	80 FR 69938		
A-570-937	China	Citric Acid and Salts	POR 13/14					12/14/2015	80 FR 77323		
A-570-964	China	Copper Pipe & Tubing	POR 13/14	12/7/2015	80 FR 75968			6/20/2016	81 FR 39893		
A-570-967	China	Aluminum Extrusions	POR 13/14					12/1/2015	80 FR 75060		
A-570-970	China	Multilayered Wood Flooring	POR 13/14	1/8/2016	81 FR 903			7/8/2016	81 FR 46899		
A-570-979	China	Solar Cells	POR 13/14	12/28/2015	80 FR 80746			6/20/2016	81 FR 39905		
A-570-983	China	Stainless Steel Sinks	POR 12/14					11/10/2015	80 FR 69644		
A-570-983	China	Stainless Steel Sinks	POR 14/15	5/12/2016	81 FR 29528			8/15/2016	81 FR 54042		
A-570-985	China	Xanthan Gum	POR 14/15	8/15/2016	81 FR 54045						
A-580-809	Korea	Circular Welded Pipe	CCR					6/20/2016	81 FR 42653		
A-580-809	Korea	Circular Welded Pipe	POR 13/14	12/8/2015	80 FR 76267						
A-580-810	Korea	Welded ASTM A-312 SS Pipe	POR 13/14					7/18/2016	81 FR 46647		
A-580-836	Korea	CTL Plate	POR 14/15	3/11/2016	81 FR 12870			9/19/2016	81 FR 62712		
A-580-867	Korea	Large Power Transformers	POR 13/14	9/2/2016	81 FR 60672			3/16/2016	81 FR 14087		
A-580-868	Korea	Washers	POR 14/15	3/11/2016	81 FR 12875			9/12/2016	81 FR 62715		
A-580-876	Korea	Welded Line Pipe	POI					10/13/2015	80 FR 61366		
A-580-878	Korea	Corrosion Resistant Steel Products	POI	1/7/2016	81 FR 78			6/2/2016	81 FR 35303		
A-580-880	Korea	Heavy Walled Rect Pipe & Tube	POI	3/1/2016	81 FR 10585			7/21/2016	81 FR 47347		
A-580-881	Korea	CR Steel Flat Products	POI	3/7/2016	81 FR 11757			7/29/2016	81 FR 49953		
A-580-883	Korea	HR Steel Flat Products	POI	3/22/2016	81 FR 15228			8/12/2016	81 FR 53419		
A-583-837	Taiwan	PET Film	POR 13/14	12/2/2015	80 FR 75451						
A-583-837	Taiwan	PET Film	POR 14/15	8/12/2016	81 FR 53441						
A-583-844	Taiwan	Narrow Woven Ribbons	POR 13/14	10/7/2015	80 FR 60627			4/18/2016	81 FR 22578		
A-583-848	Taiwan	Stilbenic Optical Brightening Agents	POR 13/14					10/13/2015	80 FR 61368		
A-583-850	Taiwan	OCTG	POR 14/15	6/13/2016	81 FR 38135			9/2/2016	81 FR 60671		
A-583-856	Taiwan	Corrosion Resistant Steel Products	POI	1/4/2016	81 FR 72			6/2/2016	81 FR 35313		
A-588-869	Japan	Diff-Annealed Nickel Plated Steel	POR 13/15	6/17/2016	81 FR 39627						
A-588-874	Japan	HR Steel Flat Products	POI	3/22/2016	81 FR 15222			8/12/2016	81 FR 53409		
A-602-808	Australia	Silicomanganese	POI					2/22/2016	81 FR		

# Attachment IV

**U.S. Less-Than-Fair-Value Investigations  
Final Determinations - Oct 1, 2015 through Sept 30, 2016**

Case Number	Country	Product	Preliminary		Amended Preliminary		Final		Amended Final		Company	Comparison Method	Final Rate	Pattern	Meaningful Difference
			FR Pub Date	FR Citation	FR Pub Date	FR Citation	FR Pub Date	FR Citation	FR Pub Date	FR Citation					
A-580-876	Korea	Welded Line Pipe	5/22/2015	80 FR 29620			10/13/2015	80 FR 61366	11/10/2015	80 FR 69637	Hyundai HYSCO	A-to-T	6.23	yes	yes
A-580-876	Korea	Welded Line Pipe	5/22/2015	80 FR 29620			10/13/2015	80 FR 61366			SeAH Steel Corp	mixed	2.53	yes	yes
A-489-822*	Turkey	Welded Line Pipe	5/22/2015	80 FR 29617			10/13/2015	80 FR 61362			Borusan Istikbal	AFA	22.95	n/a	n/a
A-489-822*	Turkey	Welded Line Pipe	5/22/2015	80 FR 29617			10/13/2015	80 FR 61362			Borusan Mannesmann	AFA	22.95	n/a	n/a
A-489-822*	Turkey	Welded Line Pipe	5/22/2015	80 FR 29617			10/13/2015	80 FR 61362			Cayirova/Yucef	A-to-A	22.95	no	n/a
A-489-822*	Turkey	Welded Line Pipe	5/22/2015	80 FR 29617			10/13/2015	80 FR 61362			Toscelik	A-to-A	6.66	yes	no
A-570-020	China	Melamine	6/18/2015	80 FR 34891			11/6/2015	80 FR 68851			Allied	AFA	363.31	n/a	n/a
A-570-020	China	Melamine	6/18/2015	80 FR 34891			11/6/2015	80 FR 68851			Golden Elephant	AFA	363.31	n/a	n/a
A-570-020	China	Melamine	6/18/2015	80 FR 34891			11/6/2015	80 FR 68851			Xinji Jiuyuan	AFA	363.31	n/a	n/a
A-274-806	Trinidad & Tobago	Melamine	6/17/2015	80 FR 34621			11/6/2015	80 FR 68846			MHTL	A-to-A	172.53	yes	no
A-570-022	China	Uncoated Paper	8/26/2015	80 FR 51768			1/20/2016	81 FR 3112			Asia Symbol	A-to-A	84.05	yes	no
A-570-022	China	Uncoated Paper	8/26/2015	80 FR 51768			1/20/2016	81 FR 3112			Sun Paper	AFA	149.00	n/a	n/a
A-570-022	China	Uncoated Paper	8/26/2015	80 FR 51768			1/20/2016	81 FR 3112			UPM (China)	AFA	149.00	n/a	n/a
A-471-807	Portugal	Uncoated Paper	8/26/2015	80 FR 51777			1/20/2016	81 FR 3105			Portucel S.A.	A-to-A	7.80	yes	no
A-602-807	Australia	Uncoated Paper	8/26/2015	80 FR 51783			1/20/2016	81 FR 3108			Paper Australia	AFA	222.46	n/a	n/a
A-351-842	Brazil	Uncoated Paper	8/27/2015	80 FR 52029			1/20/2016	81 FR 3115			International Paper	A-to-A	41.39	no	n/a
A-351-842	Brazil	Uncoated Paper	8/27/2015	80 FR 52029			1/20/2016	81 FR 3115	3/3/2016	81 FR 11174	Suzano Papel e Celulose	A-to-A	22.37	no	n/a
A-560-828*	Indonesia	Uncoated Paper	8/26/2015	80 FR 51771			1/20/2016	81 FR 3101			Great Champ Trading	AFA	17.39	n/a	n/a
A-560-828*	Indonesia	Uncoated Paper	8/26/2015	80 FR 51771			1/20/2016	81 FR 3101			Indah Kiat Pulp & Paper	AFA	17.39	n/a	n/a
A-560-828*	Indonesia	Uncoated Paper	8/26/2015	80 FR 51771			1/20/2016	81 FR 3101	3/3/2016	81 FR 11174	APRIL	mixed	2.10	yes	yes
A-602-808	Australia	Silicomanganese	9/25/2015	80 FR 57787			2/22/2016	81 FR 8882			Tasmanian Electro	A-to-A	12.03	yes	no
A-523-810	Oman	PET Resin	10/15/2015	80 FR 62021			3/14/2016	81 FR 13336	5/6/2016	81 FR 27979	OCTAL SAOC	A-to-A	7.62	yes	no
A-570-024	China	PET Resin	10/15/2015	80 FR 62024			3/14/2016	81 FR 13331			Far Eastern Industries	A-to-A	104.98	yes	no
A-570-024	China	PET Resin	10/15/2015	80 FR 62024			3/14/2016	81 FR 13331			Jiangyin Xingyu	A-to-A	118.32	yes	no
A-122-855	Canada	PET Resin	10/15/2015	80 FR 62019			3/14/2016	81 FR 13319			Selenis Canada	A-to-A	13.60	yes	no
A-533-861	India	PET Resin	10/15/2015	80 FR 62029			3/14/2016	81 FR 13327			Dhunseri Petrochem	AFA	19.41	n/a	n/a
A-533-861	India	PET Resin	10/15/2015	80 FR 62029			3/14/2016	81 FR 13327			Ester Industries	A-to-A	14.23	no	n/a
A-533-861	India	PET Resin	10/15/2015	80 FR 62029			3/14/2016	81 FR 13327			JBF Industries	AFA	19.41	n/a	n/a
A-533-861	India	PET Resin	10/15/2015	80 FR 62029			3/14/2016	81 FR 13327			Reliance Industries	A-to-A	8.03	yes	no
A-588-873	Japan	CR Steel Flat Products	3/7/2016	81 FR 11747			5/24/2016	81 FR 32721			JFE Steel	AFA	71.35	n/a	n/a
A-588-873	Japan	CR Steel Flat Products	3/7/2016	81 FR 11747			5/24/2016	81 FR 32721			Nippon Steel/Sumitomo	AFA	71.35	n/a	n/a
A-570-029	China	CR Steel Flat Products	3/7/2016	81 FR 11751			5/24/2016	81 FR 32725			China-wide entity	AFA	265.79	n/a	n/a
A-570-026	China	CORE Steel Products	1/4/2016	81 FR 75			6/2/2016	81 FR 35316			Hebei Iron & Steel	AFA	209.97	n/a	n/a
A-570-026	China	CORE Steel Products	1/4/2016	81 FR 75			6/2/2016	81 FR 35316			Yeih Phi	A-to-A	209.97	yes	no
A-570-026	China	CORE Steel Products	1/4/2016	81 FR 75			6/2/2016	81 FR 35316			Baoshan Iron & Steel	AFA	209.97	n/a	n/a
A-475-832	Italy	CORE Steel Products	1/4/2016	81 FR 69			6/2/2016	81 FR 35326			Acciaieria Arvedi	A-to-A	12.63	yes	no
A-475-832	Italy	CORE Steel Products	1/4/2016	81 FR 69			6/2/2016	81 FR 35326			Marcegaglia	AFA	92.12	n/a	n/a
A-533-863	India	CORE Steel Products	1/4/2016	81 FR 63			6/2/2016	81 FR 35329	7/25/2016	81 FR 48390	JSW	A-to-T	4.43	yes	yes
A-533-863	India	CORE Steel Products	1/4/2016	81 FR 63			6/2/2016	81 FR 35329			Uttam Galva	A-to-A	3.05	yes	no
A-580-878	Korea	CORE Steel Products	1/4/2016	81 FR 78			6/2/2016	81 FR 35305			Dongkuk Steel	A-to-A	8.75	yes	no
A-580-878	Korea	CORE Steel Products	1/4/2016	81 FR 78			6/2/2016	81 FR 35305			Hyundai Steel	A-to-A	47.79	yes	no
A-583-856	Taiwan	CORE Steel Products	1/4/2016	81 FR 72			6/2/2016	81 FR 35313	7/25/2016	81 FR 48390	Prosperity Tieh	A-to-A	10.34	yes	no
A-570-028	China	Hydrofluorocarbon Blends	2/1/2016	81 FR 5098			6/29/2016	81 FR 42314	7/25/2016	81 FR 48390	T. T. International	A-to-A	101.82	yes	no
A-201-847	Mexico	Heavy-Walled Rect Pipe & Tube	3/1/2016	81 FR 10587			7/21/2016	81 FR 47352			Maquilacero	A-to-T	3.83	yes	yes
A-201-847	Mexico	Heavy-Walled Rect Pipe & Tube	3/1/2016	81 FR 10587	3/16/2016	81 FR 14090	7/21/2016	81 FR 47352			Prolamsa	A-to-A	5.21	yes	no
A-489-824	Turkey	Heavy-Walled Rect Pipe & Tube	3/1/2016	81 FR 10583			7/21/2016	81 FR 47355			MMZ Boru Profil	AFA	35.66	n/a	n/a
A-489-824	Turkey	Heavy-Walled Rect Pipe & Tube	3/1/2016	81 FR 10583			7/21/2016	81 FR 47355			Ozdemir Boru Profil	A-to-A	0.00	yes	no
A-580-880	Korea	Heavy-Walled Rect Pipe & Tube	3/1/2016	81 FR 10585			7/21/2016	81 FR 47347			Dong-A Steel	A-to-T	2.34	yes	yes
A-580-880	Korea	Heavy-Walled Rect Pipe & Tube	3/1/2016	81 FR 10585			7/21/2016	81 FR 47347			HiSteel	A-to-T	3.82	yes	yes
A-351-843	Brazil	CR Steel Flat Products	3/7/2016	81 FR 11754	4/7/2016	81 FR 20366	7/29/2016	81 FR 49946	9/20/2016	81 FR 64432	Siderugicas Nacional	A-to-A	19.58	yes	no
A-351-843	Brazil	CR Steel Flat Products	3/7/2016	81 FR 11754			7/29/2016	81 FR 49946			Usiminas	AFA	35.43	n/a	n/a
A-412-824	United Kingdom	CR Steel Flat Products	3/7/2016	81 FR 11744			7/29/2016	81 FR 49929			Caparo Precision Strip	mixed	5.40	yes	yes
A-412-824	United Kingdom	CR Steel Flat Products	3/7/2016	81 FR 11744			7/29/2016	81 FR 49929	9/20/2016	81 FR 64432	Tata Steel UK	A-to-A	25.17	yes	no
A-533-865	India	CR Steel Flat Products	3/7/2016	81 FR 11741			7/29/2016	81 FR 49938			JSW Steel	A-to-A	7.60	yes	no
A-821-822	Russia	CR Steel Flat Products	3/8/2016	81 FR 12072			7/29/2016	81 FR 49950			Severstal	A-to-T	13.36	yes	yes
A-821-822	Russia	CR Steel Flat Products	3/8/2016	81 FR 12072			7/29/2016	81 FR 49950			Novolipetsk Steel	A-to-A	1.04	no	n/a
A-580-881	Korea	CR Steel Flat Products	3/7/2016	81 FR 11757			7/29/2016	81 FR 49953			Hyundai Steel	A-to-A	34.33	yes	no
A-580-881	Korea	CR Steel Flat Products	3/7/2016	81 FR 11757			7/29/2016	81 FR 49953			POSCO	A-to-T	6.32	yes	yes
A-588-874	Japan	HR Steel Flat Products	3/22/2016	81 FR 15222			8/12/2016	81 FR 53409			Nippon Steel/Sumitomo	mixed	4.99	yes	yes
A-588-874	Japan	HR Steel Flat Products	3/22/2016	81 FR 15222			8/12/2016	81 FR 53409			JFE Steel	A-to-A	7.23	no	no
A-421-813	Netherlands	HR Steel Flat Products	3/22/2016	81 FR 15225			8/12/2016	81 FR 53421			Tata Steel IJmuiden	A-to-T	3.73	yes	yes
A-602-809	Australia	HR Steel Flat Products	3/22/2016	81 FR 15241			8/12/2016	81 FR 53406	10/3/2016	81 FR 67692	Bluescope Steel	A-to-A	29.58	yes	no
A-351-845	Brazil	HR Steel Flat Products	3/22/2016	81 FR 15235			8/12/2016	81 FR 53424			Siderugicas Nacional	A-to-A	33.14	yes	no
A-351-845	Brazil	HR Steel Flat Products	3/22/2016	81 FR 15235			8/12/2016	81 FR 53424			Usiminas	AFA	34.28	n/a	n/a
A-412-825	United Kingdom	HR Steel Flat Products	3/22/2016	81 FR 15244			8/12/2016	81 FR 53436			Tata Steel UK	A-to-A	33.06	yes	no
A-580-883	Korea	HR Steel Flat Products	3/22/2016	81 FR 15228			8/12/2016	81 FR 53419			Hyundai Steel	A-to-A	9.49	yes	no
A-580-883	Korea	HR Steel Flat Products	3/22/2016	81 FR 15228			8/12/2016	81 FR 53419	10/3/2016	81 FR 67692	POSCO	mixed	4.61	yes	yes
A-489-826	Turkey	HR Steel Flat Products	3/22/2016	81 FR 15231			8/12/2016	81 FR 53428	10/3/2016	81 FR 67692	Colakoglu	A-to-A	6.77	yes	no
A-489-826	Turkey	HR Steel Flat Products	3/22/2016	81 FR 15231			8/12/2016	81 FR 53428	10/3/2016	81 FR 67692	Erdemir	A-to-T	4.15	yes	yes
A-533-867	India	Welded Stainless Pressure Pipe	5/10/2016	81 FR 28824			9/29/2016	81 FR 66921			Streamline	A-to-A	12.66	yes	no
A-533-867	India	Welded Stainless Pressure Pipe	5/10/2016	81 FR 28824			9/29/2016	81 FR 66921			Sunrise Group	A-to-A	0.00	yes	no

\* Not included in Resolute FP Rule 57(3) Reply Brief at Appendix B.

## CERTIFICATE OF SERVICE

I hereby certify that, on this 30<sup>th</sup> day of April, 2024, I served or caused to be served copies of the preceding Final Results of Redetermination upon the following parties in the manner indicated below:

### VIA ELECTRONIC MAIL:

On behalf of the British Columbia Lumber  
Trade Council et al.:

Stephanie W. Wang  
STEPTOE LLP  
1330 Connecticut Ave. NW  
Washington, DC 20036  
Telephone: (202) 429-6496  
Facsimile: (202) 429-3902  
steptoelumber@steptoe.com

Duncan Reid  
Farris, Vaughan, Wills & Murphy LLP  
P.O. Box 10026, Pacific Centre South  
25<sup>th</sup> Floor, 700 West Georgia Street  
Vancouver, B.C. V7Y 1B3  
Telephone: (604) 661-1736  
Facsimile: (604) 661-9349  
dreid@farris.com

On behalf of the Committee Overseeing  
Action for Lumber International Trade  
Investigations or Negotiations  
(COALITION)

Whitney M. Rolig  
Picard Kentz & Rowe LLP  
1750 K Street, N.W., Suite 800  
Washington, DC 20006  
Telephone: (202) 331-4040  
Facsimile: (202) 331-4011  
trade@pkrlp.com

On behalf of the Government of Ontario:

H. Deen Kaplan

On behalf of Canfor Corporation

Donald B. Cameron, Esq.  
Morris, Manning & Martin, LLP  
1401 I Street, NW, Suite 600  
Washington, DC 20005  
Telephone: (202) 216-4811  
Facsimile: (202) 408-5146  
tradeservice@mmmlaw.com

On behalf of the Government of Canada:

Joanne E. Osendarp  
Blank Rome LLP  
1825 Eye Street NW  
Washington, DC 20006  
Telephone: (202) 420-2587  
joanne.osendarp@blankrome.com  
tradeservice@blankrome.com

On behalf of the Government of Quebec:

Nancy A. Noonan  
ArentFox Schiff LLP  
1717 K Street, NW  
Washington, DC 20006-5344  
Telephone: (202) 857-6066  
Facsimile: (202) 857-6395  
Nancy.noonan@afslaw.com

On behalf of Resolute FP Canada Inc., the  
Conseil de l'Industrie forestière du Québec,  
the Ontario Forest Industries Association, and  
René Bernard Inc.:

Hogan Lovells US LLP  
555 13<sup>th</sup> Street, NW  
Washington, DC 20004  
Telephone: (202) 637-5799  
Facsimile: (202) 637-5910  
Deen.Kaplan@hoganlovells.com

On behalf of J.D. Irving, Limited:

Walter J. Spak  
White & Case LLP  
701 13<sup>th</sup> Street, NW  
Washington, DC 20005-3807  
Telephone: (202) 626-3606  
Facsimile: (202) 639-9355  
wspak@whitecase.com  
apotrade@whitecase.com

On behalf of Tolko Marketing and Sales Ltd.,  
and Tolko Industries, Ltd.:

Lynn Fischer Fox  
Arnold & Porter Kay Scholer LLP  
601 Massachusetts Avenue, NW  
Washington, DC 20001-3743  
Telephone: (202) 942-5601  
Facsimile: (202) 942-5999  
lynn.fischerfox@apks.com  
xlumber@apks.com

Elliot J. Feldman  
Baker Hostetler LLP  
1050 Connecticut Avenue, NW, Suite 1100  
Washington, DC 20036  
Telephone: (202) 861-1679  
Facsimile: (202) 861-1783  
efeldman@bakerlaw.com  
trade@bakerlaw.com

On behalf of West Fraser Mills Ltd.

Donald Harrison  
Gibson, Dunn & Crutcher LLP  
1050 Connecticut Avenue, NW  
Washington, DC 20036-5306  
Telephone: (202) 955-8560  
Facsimile: (202) 530-9554  
DHarrison@gibsondunn.com

/s/ Jason Miller

Jason Miller  
Attorney  
Office of the Chief Counsel  
for Trade Enforcement and Compliance  
Washington, D.C. 20230  
(202) 597-1011