

**BEFORE THE
SURFACE TRANSPORTATION BOARD**

RECIPROCAL SWITCHING FOR INADEQUATE SERVICE

Docket No. EP 711 (Sub-No. 2)

**REPLY COMMENTS OF
NORFOLK SOUTHERN RAILWAY COMPANY**

Nabanita C. Nag
Jason M. Morris
Thomas E. Zoeller
Hanna M. Chouest
Julianne C. Freeman
NORFOLK SOUTHERN RAILWAY
COMPANY
650 West Peachtree Street N.W.
Atlanta, GA 30308

Raymond A. Atkins
Morgan B. Lindsay
SIDLEY AUSTIN LLP
1501 K Street N.W.
Washington, D.C. 20005

Counsel for Norfolk Southern Railway Company

Dated: December 20, 2023

TABLE OF CONTENTS

	<u>Page</u>
INTRODUCTION	1
I. THE STB CANNOT REGULATE PRIVATE CONTRACT SERVICE AS REQUESTED BY SHIPPER TRADE ASSOCIATIONS.	3
II. THE TRANSIT TIME METRIC SHOULD BE USED AS ONLY SUPPLEMENTAL EVIDENCE, NOT AS A STAND-ALONE MEASURE OF SERVICE INADEQUACY.	9
III. ONE SIMPLE CHANGE TO THE PROPOSED RULE CAN ADDRESS MULTIPLE CONCERNS AND HELP CONFORM THE PROPOSED RULE TO THE STATUTE.	19
CONCLUSION.....	21

Exhibit A: Reply Verified Statement of Mark A. Israel, Ph.D.

INTRODUCTION

Norfolk Southern Railway Company (“NS”) submits these reply comments on the Notice of Proposed Rulemaking (“NPRM”) in this proceeding¹ and joins those submitted by the Association of American Railroads (“AAR”). NS appreciates and joins the Board’s emphasis on the importance of reliable rail service. NS deeply values its customers and works every day to do right by them. NS is diligently focused on providing its customers with safe, reliable service, as well as customer-centric offerings designed to enhance visibility, communication, private resolutions, and, ultimately, trust between NS and its customers. NS supports remedies permitted by statute to address harmful inadequate service and believes the proposed rule could be adjusted to offer such a remedy.

These reply comments seek to build upon NS’s opening comments and provide constructive and simple ways the Board should alter the proposed rule to conform it to the statute and tailor it to achieve the Board’s stated goals.

First, the Board cannot lawfully apply this program to private, confidential contract movements as requested by various shipper trade associations. Regulating contract traffic by applying performance metrics to that traffic and then threatening a forced switch if those metrics are not satisfied would be a plain violation of the Interstate Commerce Act. Simply put, Congress declared that carriers have no duty in connection with services provided under such contracts other than those duties specified by the terms of the contract. NS urges the STB to continue its long tradition of respecting this plain limit on its authority.

¹ Notice of Proposed Rulemaking and Closure of Subdocket No. 1, *Reciprocal Switching for Inadequate Service*, Docket No. EP 711 (Sub-No. 2) (STB served Sept. 7, 2023) (“NPRM”).

Second, data submitted in the AAR's reply comments reveals that the transit time metric is seriously flawed and requires the STB's full attention. If left unaddressed, it would undermine the entire program. In particular, the data show via clear, empirical evidence that the proposed transit time metric is not, as intended, measuring inadequate service. Instead of creating a service guardrail that isolates potential service problems, this metric is capturing the *majority* of rail traffic—*i.e.*, normal transit time variations that have no empirical tie to inadequate service. If the Board wishes to use the transit time metric, the Board should pause this proceeding to more carefully study transit time data to ensure any proposed transit time metric is empirically supported. However, NS recommends a more expeditious option—permit evidence of transit time variability *only* as supplemental evidence to support a claim of inadequate service and not a stand-alone measure of inadequate service (which, the data shows, it is not). This would retain the metric as potentially valuable evidence while avoiding the harm that would result from imposing the serious remedy of a forced switch as a result of normal railroad operations.

Third, one simple adjustment to the language of proposed Part 1145—namely, permitting any affirmative defenses related to the statutory standard—would address multiple concerns raised on opening at once and better conform the proposed program to the standard in 49 U.S.C. § 11102(c). In particular, permitting affirmative defenses related to the statutory standard would address requests by multiple parties asking for the flexibility to show, *e.g.*, that a prescribed switch is unnecessary, would not remedy an actual harm, or is not in the public interest. For example, Dr. Mark Israel has recommended that the STB use the sophisticated yet simple power of standard statistical tools to assure itself that any perceived violation of performance metrics is not attributable to too few observations or mere

statistical noise. Dr. Israel's Reply Verified Statement, which accompanies these reply comments, details the simple statistical tools that are readily available to the Board. Parties should be free and encouraged to submit this kind of statistical evidence in individual cases as an affirmative defense related to the statutory standard. They should also be permitted to raise other concerns related to the statutory standard, such as those described in various opening comments, including, whether the service problem has been cured, whether the service issues are due to forces—such as heightened demand—that cannot be addressed by a forced switch, whether there is a compelling need for the switch, etc.

The STB has correctly observed that “even temporary access is a serious remedy, given the potentially significant operational, safety, and financial implications for the carriers involved.”² NS fully agrees. Permitting carriers to submit evidence to show no actual necessity or compelling need for a forced switch, or that the switch is not otherwise in the public interest, will safeguard the public interest by ensuring that the program remedies only real service inadequacies that can be remedied by a prescribed switch. This result is consistent with the Board's stated goals for the program. If not so tailored, the program will risk the well-known harms that result from regulating when there is no “compelling need,” which will burden the rail network, rail employees, and ultimately rail customers.

I. THE STB CANNOT REGULATE PRIVATE CONTRACT SERVICE AS REQUESTED BY SHIPPER TRADE ASSOCIATIONS.

Congress has forbidden the STB from regulating contract transportation services. The language Congress used is plain and unambiguous.

² Final Rules, *Expedited Relief for Serv. Inadequacies*, Docket No. EP 628, 3 S.T.B. 968, 977 (STB served Dec. 21, 1998) (“*EP 628 Final Rule*”).

(a) One or more rail carriers providing transportation subject to the jurisdiction of the Board under this part may enter into a contract with one or more purchasers of rail services to provide specified services under specified rates and conditions.

(b) A party to a contract entered into under this section *shall have no duty in connection with services provided under such contract other than those duties specified by the terms of the contract.*

(c) (1) A contract that is authorized by this section, and transportation under such contract, *shall not be subject to this part, and may not be subsequently challenged before the Board or in any court on the grounds that such contract violates a provision of this part.*

49 U.S.C. §§ 10709(a)–(c)(1) (emphasis of key parts added)

This language leaves no wiggle room. The Board may develop service standards to help identify poor service that may justify a prescribed switch under section 11102(c), but it cannot apply those standards, or any other duties and obligations, to contract traffic. A railroad “ha[s] no duty in connection with services provided under such a contract other than those duties specified by the terms of the contract.” 49 U.S.C. § 10709(b). This means what it says. For contract traffic, a railroad would have no duty to meet a federal 60% on-time-performance metric, or a particular transit time, or a regulated first-mile/last-mile service metric, unless those performance standards were negotiated and included in the private contract. If the railroad has no duty to conform its behavior to the federal regulations, then it should be readily apparent that the STB cannot attach a federal remedy to noncompliance. One cannot simultaneously have no obligation to meet federal performance standards yet be punished for not meeting them. The two positions are incompatible.

Congress made its intent crystal clear by using a belt-and-suspenders approach in section 10709—*i.e.*, describing *twice*, in two different ways, the fact that contract traffic is not subject to the Act and thus outside the Board’s jurisdiction

and authority. In addition to the “no duty” language, the statute also states that the “transportation under such contract” is “not [] subject to this part.” 49 U.S.C. § 10709(c)(1). This means that transportation provided under a contract is not subject to *any* provision of the Interstate Commerce Act, including section 11102(c) or regulations the STB promulgates under that subsection.³ If contract transportation is not subject to the provisions of 11102(c), it is self-apparent that the service for that same contract transportation cannot be the grounds for relief under 11102(c). Importantly, Congress did not leave customers without remedies for inadequate contract service. It simply removed those remedies from the federal regulator and placed them in the legal forum selected by the parties. The “exclusive remedy” based on any claim of inadequate service is “an action in an appropriate State court or United States district court, unless the parties otherwise agree.” 49 U.S.C. § 10709(c)(2). And the remedies for such inadequate service are those that are negotiated between the parties and decided under state law.

³ In case there is any doubt, both the STB and federal courts have made clear that the reference to “this part” throughout section 10709 refers to “Part A of Subtitle IV of Title 49 of the U.S. Code, i.e., 49 U.S.C. [§§] 10101 through ... 11908.” Decision No. 7, *Can. Nat’l Ry. Co. & Grand Trunk Corp.—Control—Duluth, Missabe & Iron Range Ry. Co. et al.*, Docket No. FD 34424, at 32 (STB served Apr. 9, 2004) (“49 U.S.C. 10709(c)(2) ... also provides that ‘[a] contract that is authorized by this section, and transportation under such contract, shall not be subject to this part [the reference is to Part A of Subtitle IV of Title 49 of the U.S. Code, i.e., 49 U.S.C. 10101 through 49 U.S.C. 11908] ...’”); Decision, *Ameropan Oil Corp. v. Can. Nat’l Ry. Co.*, Docket No. NOR 42161, at 3 (STB served Apr. 17, 2019) (“*Ameropan*”) (stating that “[u]nder § 10709(c)(1), a rail services contract ‘and transportation under such contract, shall not be subject to [49 U.S.C. Subtitle IV, Part A]’ ...”); *Babcock & Wilcox Co. v. Kan. City S. Ry.*, 557 F.3d 134, 136 n.3 (3d Cir. 2009) (explaining that “this part” in 49 U.S.C. § 10709 refers to “i.e., 49 U.S.C. §§ 10101-11908”); *Sompo Japan Ins. Co. of Am. v. Norfolk S. Ry. Co.*, 540 F. Supp. 2d 486, 493, 494 (S.D.N.Y. 2008) (explaining that, under 49 U.S.C. §§ 10709(c)(1) and 10502(a), “[t]his part’ refers to 49 U.S.C. Subtitle IV, Part A,” “which encompasses 49 U.S.C. §§ 10101-11908”).

Various commentators nonetheless encourage the STB to defy this clear Congressional limit and use contract service as grounds for relief under section 11102(c). The principal advocate is the Coalition Associations, which argue that by considering a carrier's performance while the transportation is under contract, "the Board is not regulating contract transportation; nor is the Board interpreting the contract, deciding whether the contract violates the statute, or resolving any contract dispute."⁴ Relying primarily on *Midtec Paper Corp. v. Chicago & North Western Transportation Co. (Use of Terminal Facilities & Reciprocal Switching Agreement)*, 3 I.C.C.2d 171 (1986), and *Vista Chemical Co. v. Atchison, Topeka & Santa Fe Railway Co.*, 5 I.C.C.2d 331 (1989), they assert it is permissible to consider evidence based on contract transportation service that is outside the agency's jurisdiction to grant the serious federal remedy of a forced switch.⁵

This position fails for three reasons. First, the two cases offer no support for the prospect of using perceived inadequate contract services as the grounds for a federal remedy. This is not surprising. The ICC and STB have a long, unbroken tradition of respecting the plain limits Congress placed on their authority.⁶ *Midtec*

⁴ Opening Comments submitted by The Coalition Associations, *Reciprocal Switching for Inadequate Service*, Docket No. EP 711 (Sub-No. 2), at 10 (filed Nov. 7, 2023) ("Coalition Associations Comments").

⁵ Coalition Associations Comments at 11.

⁶ See, e.g., *Ameropan*, at 3–4 (noting that "[i]t is well settled that the Board's authority does not extend to those movements conducted pursuant to transportation contracts" and thus, "if the movement is subject to a contract, the common carrier service terms do not apply to the movement in question") (emphasis added and internal quotation marks omitted); Notice of Proposed Rule, *Rail Transportation Contracts Under 49 U.S.C. 10709*, Docket No. EP 676, at 3–4 (STB served Jan. 6, 2009) (recognizing that "a rail transportation contract [is] outside the Board's jurisdiction under 49 U.S.C. 10709" and explaining "Congress' dual intent to offer regulatory protection to shippers that desire such protection, while encouraging private rail transportation contracts for those shippers that prefer such arrangements"); *Pejepscot Indus. Park, Inc., d/b/a Grimmel Indus.—Petition for*

and *Vista Chemicals* did little more than observe that a carrier's willingness to enter into private rail contracts is one of many factors that might show an absence of abusive anticompetitive conduct. This is a far cry from, and offers no support for, tying a federal remedy to the level and quality of contract transportation service, which Congress expressly placed outside the Board's jurisdiction. The Coalition Associations unearth no precedent for *that* proposition, because none exists.

Second, applying the proposed performance standards to contract traffic—regardless of when a prescription is ultimately awarded—would plainly regulate contract transportation. “Regulation” of transportation services can come in many guises, as the broad body of ICCTA preemption cases reveals. But state or federal regulation of railroad activity all share two common elements: the creation of a duty of care and enforceable sanctions for violating that duty. Combined together, the duty and sanctions are intended to “regulate” the behavior of a railroad—*i.e.*, to force the carrier to conform its conduct to that desired by the regulator. The duty informs the carrier of the expectations of the regulator, and the sanctions are the stick to compel compliance.

Declaratory Order, Docket No. FD 33989, 6 S.T.B. 886, 891 (STB decided May 9, 2003) (recognizing that Congress's “exemption of a commodity under 49 U.S.C. 1050212 generally excuses carriers from virtually all aspects of regulation involving the transportation of that commodity. ... Thus, even if a carrier's conduct would constitute a statutory violation during a period of regulation, the exemption bars regulatory relief during the period when the exemption is in force.”); *Arbitration of Certain Disputes Subject to the Statutory Jurisdiction of the Surf. Transp. Bd.*, Docket No. EP 560, 2 S.T.B. 564, 568 (1997) (“WCTL suggests that we allow unified arbitration of disputes involving jurisdictional and non-jurisdictional elements that are intertwined The example that WCTL offers, however—‘a private contract that incorporates elements of common carriage (such as rates or service levels)’—is one that is *entirely outside of our jurisdiction*.”) (emphasis added); *R.R. Cost Recovery Procs.* 49 C.F.R. Parts 1135 & 1312, 3 I.C.C.2d 60, 68 (1986) (explaining that “rail contracts entered into under §10713 ... are outside of our jurisdiction”).

The proposed program contains both of these hallmarks of regulatory rules. It would impose new minimum federal performance standards on the carriers under section 11102(c) and then use the threat of a two- to four-year forced switch to compel compliance. This is undisguised regulation of railroad transportation. While NS recognizes the Board's authority to regulate the service quality of *regulated* traffic within the confines of the statute, the agency cannot regulate, in any form, directly or indirectly, contract traffic that is not subject to the Act, including, of course, section 11102(c). This conclusion is unassailable. By Congressional mandate, private contract traffic simply is not subject to the requirements of section 11102(c), period.

For the Coalition Associations to nonetheless claim that applying these new performance standards to contract service is “just” using the contract service as evidence of the need for a forced switch, but not somehow “regulating” that service, is absurd. If a state were to enact a program that imposed a higher tax on carriers where service levels in the state fell below these same performance standards, no one would seriously dispute that this would be impermissible “regulation” of interstate rail transportation by the state and preempted by federal law. Just as a state cannot regulate interstate rail transportation, the STB cannot regulate private contract traffic because “[a] contract that is authorized by this section, *and transportation under such contract, shall not be subject to this part*, and may not be subsequently challenged before the Board or in any court on the grounds that such contract violates a provision of this part.” 49 U.S.C. § 10709(c)(1) (emphasis added).

Finally, the Coalition Associations' arguments ultimately cannot surmount the plain language of the statute. The STB need not pour through old cases or explore the nuanced parameters of regulatory activity to decide this issue. Its analysis can begin and end with the clear language of Congress. For private

contract transportation, the parties have “no duty in connection with services provided under such contract other than those duties specified by the terms of the contract.” 49 U.S.C. § 10709(b). This includes any performance standards promulgated pursuant to section 11102(c). And the transportation “may not be subsequently challenged before the Board or in any court on the grounds that such contract violates a provision of this part.” 49 U.S.C. § 10709(c)(1). The term “this part” includes section 11102(c). These dual statutory commands are an intractable obstacle to the Coalition Associations’ position that the STB can impose new performance standards on traffic Congress expressly placed outside its jurisdiction. NS therefore strongly urges the STB to reject the arguments raised by the Coalition Associations and conform this rule to the clear jurisdictional limits on its authority.

II. THE TRANSIT TIME METRIC SHOULD BE USED AS ONLY SUPPLEMENTAL EVIDENCE, NOT AS A STAND-ALONE MEASURE OF SERVICE INADEQUACY.

In this section, NS explains why the transit time metric—as proposed—is not measuring what the Board intends it to measure and thus must be altered, in either its use or definition, to be a meaningful and reliable tool in-line with the Board’s stated purpose. However, NS’s feedback with regard to the proposed metric in no way minimizes the value, in general, of transit time data. NS recognizes that transit time data is important to its customers, and it is important to NS—NS monitors transit time and uses it as a tool to diagnose and problem-solve network issues as part of its commitment to providing safe, reliable service to its customers. As such, NS believes transit time data can be valuable for monitoring service. NS also supports the ability of rail customers to use transit time data, where meaningful, to support their complaints alleging inadequate service. As NS explains below, while the transit time metric cannot bear the weight of being a stand-alone measure of service inadequacy, NS supports its use as supplemental evidence.

The centerpiece of this proposal are three performance standards the agency seeks to use as evidence that a forced switch may be in the public interest. As NS understands this proposal, the STB seeks objective service metrics that almost on their face signal a serious service issue that may satisfy the governing legal standard for relief under 49 U.S.C. § 11102(c)'s public interest prong and thus warrant further scrutiny.⁷ These metrics seem to be intended to be guardrails, so to speak—not the ultimate high quality of service customers expect and NS strives to deliver, but rather minimum performance standards that, if not met, reveal a potential serious service deficiency *res ipsa loquitur*: “the thing speaks for itself.”

But if the STB is not careful, the transit time metric will be this proposal's Achilles heel. In opening comments, many commenters expressed opinions about

⁷ More of course is needed beyond objective performance standards to satisfy section 11102(c)'s public interest prong, which, aside from practicability, requires a showing of “compelling need” for the switch and a finding that the prescribed switch is in the public interest based on the totality of circumstances, including the interests of all impacted parties. Opening Comments of Norfolk Southern Railway Company, *Reciprocal Switching for Inadequate Service*, Docket No. EP 711 (Sub-No. 2), at 3–7 (filed Nov. 7, 2023) (“NS Comments”); see *Mfrs. Ass’n of York, Pa. v. Pa. R.R. Co.*, 73 I.C.C. 40, 49 (1922) (when “determining what is ‘in the public interest’ in a given case, as antecedent to the affirmative exercise of this broad grant of power, we must take into consideration not only the interests of the particular shipper[] ... involved but also the interests of the carriers and of the general public.”); *Del. & Hudson Ry. Co. v. Consol. Rail Corp.—Reciprocal Switching Agreement*, 367 I.C.C. 718, 720 (1983) (recognizing that the *Jamestown, N.Y., Chamber of Com. v. Jamestown, Westfield & Nw. R.R. Co.*, 195 I.C.C. 289, 292 (1933) standard of “actual necessity” applies to grants of reciprocal switching under the public interest prong of what is now 49 U.S.C. § 11102(c)); *Cent. States Enters., Inc. v. ICC*, 780 F.2d 664, 678 (7th Cir. 1985) (describing the “pre-Staggers Act cases ... that describe the public interest standard,” for what is now § 11102(c), including that “some actual necessity or some compelling reason must first be shown before [the agency] can find such action in the public interest”); *Midtec Paper Corp. v. Chi. & Nw. Transp. Co. (Use of Terminal Facilities and Reciprocal Switching Agreement)*, 1 I.C.C.2d 362, 364 (1985).

the proposed transit time metric—many pointed out its serious flaws,⁸ while others urged the Board to adopt it but adjust the percentage at which it should be set.⁹ However, no party in support of the metric, including the Board, has supported its

⁸ See, e.g., Opening Comments on Behalf of CN, *Reciprocal Switching for Inadequate Service*, Docket No. EP 711 (Sub-No. 2), at 46 (filed Nov. 7, 2023) (“CN Comments”) (noting that the transit time metric will disproportionately impact traffic with shorter running times and is inappropriate for empty moves in light of how empty private cars are managed); Opening Comments of Canadian Pacific Kansas City, *Reciprocal Switching for Inadequate Service*, Docket No. EP 711 (Sub-No. 2), at 3–6 (filed Nov. 7, 2023) (noting that the transit time metric had never been studied on a lane-by-lane basis and thus could not be reliably tied to service inadequacy, and observing that, based on a preliminary review of its service data, one-fifth of its lanes would trigger the metric despite the fact that one-fifth of its lanes do not experience inadequate service); Comments of Union Pacific Railroad Company, *Reciprocal Switching for Inadequate Service*, Docket No. EP 711 (Sub-No. 2), at 9–10 (filed Nov. 7, 2023) (pointing out that the transit time metric is not supported by data and noting that transit variability can occur for pro-service reasons); Comments of the Alliance for Chemical Distribution, *Reciprocal Switching for Inadequate Service*, Docket No. EP 711 (Sub-No. 2), at 4–5 (filed Nov. 7, 2023) (“ACD Comments”) (noting that the transit time metric would not capture poor transit time that has little variability); Comments of the Portland Cement Association, *Reciprocal Switching for Inadequate Service*, Docket No. EP 711 (Sub-No. 2), at 11 (filed Nov. 7, 2023) (noting that the transit time metric could discourage carriers from improving transit times).

⁹ See, e.g., ACD Comments, at 4–5 (no data supporting requests for 15% threshold for a year-over-year comparison and 25% threshold for comparing the same 12-week period with that of two years prior); Coalition Associations Comments, at 22–43 (filed Nov. 7, 2023) (no data supporting similar request); Comments of the American Fuel & Petrochemical Manufacturers, *Reciprocal Switching for Inadequate Service*, Docket No. EP 711 (Sub-No. 2), at 10–11 (filed Nov. 7, 2023) (“AFPM Comments”) (no data supporting request for threshold to be set at 15% based on a rolling average of three years); Comments of Olin Corporation, *Reciprocal Switching for Inadequate Service*, Docket No. EP 711 (Sub-No. 2), at 6 (filed Nov. 7, 2023) (no data supporting request for threshold to be set at 10–15%); Comments of the Freight Rail Customer Alliance and National Coal Transportation Association, *Reciprocal Switching for Inadequate Service*, Docket No. EP 711 (Sub-No. 2), at 2–3 (filed Nov. 7, 2023) (no data supporting request for thresholds to be more stringent).

preferred threshold with empirical data.¹⁰ In other words, no party provides data-based, empirical evidence to support the critical assumption that underlies the intended function of the transit time metric: that a 20% or 25% (or any %) variation in transit time—on a 12-week rolling average on a lane-by-lane basis—is evidence of poor service that can justify a forced switch under section 11102(c). It is certainly not clear why, for example, a change in transit time from 5 days to 7 days signals an actual necessity or compelling reason for a government-mandated reciprocal switch.

The STB now has that missing data, and it is eye opening. On reply, the AAR presents extensive data with millions of carload observations and thousands of lanes, over four years, which reveal that the proposed transit time metric is wildly overinclusive and, ultimately, unreliable for its intended purpose.¹¹ As shown in **FTI Table 1**, copied below, the Board’s proposed transit time metric would be triggered by the *majority* of Class I railroad movements and, worse, cannot distinguish between changes in transit time for pro-service reasons, for reasons outside a railroad’s control, or that cause no harm to the customer due to, *e.g.*, proactive communication by the carrier or the customer’s unique circumstances. This is not a metric that is functioning as a service guardrail. In fact, it is untailored to its intended purpose—to filter out potential service of such poor quality that it may justify regulatory intervention.

¹⁰ See NPRM at 18 & n.26 (citing as support for the “20 or 25%” threshold mere anecdotal claims from four shippers that testified at the April 2022 hearing in Docket No. EP 770).

¹¹ See *generally* Reply Comments of the Association of American Railroads, *Reciprocal Switching for Inadequate Service*, Docket No. EP 711 (Sub-No. 2) (filed Dec. 20, 2023), Reply Verified Statement of Michael R. Baranowski and Nathaniel S. Zebrowski (“Baranowski & Zebrowski Reply V.S.”).

FTI Table 1:
Likelihood that a Given Carload in the Stated Year Traveled on a Lane
with Year-Over-Year Transit Time Changes Over 20% in at Least One
Rolling 12-Week Period Compared to the Prior Year Average¹²

Year End September 30	20% Transit Time Increases	20% Transit Time Decreases	Transit Time Increases and Decreases	Transit Time Increases and/or Decreases
2020	53.9%	76.7%	38.5%	92.1%
2021	76.6%	53.7%	38.9%	91.5%
2022	82.5%	41.9%	31.3%	93.1%
2023	65.5%	62.3%	38.9%	88.8%

Importantly, the FTI analysis also establishes that transit time variability is the same whether the customer is single-served or dual-served. As shown below in **FTI Table 5**, there is no meaningful difference in transit time variability regardless of competitive options. This should not be surprising, as all cars on the same train move at the same speed from origin to destination, regardless of whether the customer is single-served or dual-served.

¹² Baranowski & Zebrowski Reply V.S. at 7, Table 1.

FTI Table 5:
Likelihood that a Given Carload Traveled on Single- or Multiple-Served Lanes with a Year-Over-Year Transit Time Increase Greater than 20% in at Least One 12-Week Period Compared to Prior Year Average (Traffic from October 1, 2022 to September 30, 2023)¹³

Comparison Group	Likelihood of Triggering Transit Time Metric
Single-Served on One or Both Ends	66.1%
Multi-Served on Both Ends	66.2%

The reason the proposed 20% or 25% transit time metric would capture the majority of all rail traffic is two-fold. First, transit time variation is normal.¹⁴ The

¹³ Baranowski & Zebrowski Reply V.S. at 12, Table 5.

¹⁴ See, e.g., Comments of the Association of American Railroads, *Expedited Relief for Service Inadequacies*, Docket No. EP 628 (filed June 15, 1998), Verified Statement of Deborah Butler, at 2 (“[A]ny determination regarding whether there has been a sufficient deterioration in service to justify an access remedy must first acknowledge the significant element of variability inherent in rail service due to commodity differences, routine operating difficulties, weather, business surges, abnormal periods and customer actions.”); *id.*, Verified Statement of John Orrison, at 6 (“[I]t is critically important to understand that railroad freight service (on CSXT and other railroads) has an inherent and largely unavoidable element of variability, both among different types of freight movements and over time.”); *id.*, Verified Statement of Gregory Stengem, at 2 (“[T]he Board needs carefully to distinguish between changes or variations in service that are part of normal railroad operations and situations where service has actually deteriorated. ... The Board should keep in mind that a railroad is a complex network, and events in one part of the system can have significant effects throughout the network. In this respect, railroads are very different from trucks, which serve their customers by hauling a single load on a single trailer between two points. For a railroad to serve any particular customer, it must necessarily coordinate that service with the service it provides to its other customers. Therefore, to accommodate changed conditions in a different part of the railroad network, service to a particular shipper or in a specific area will necessarily go up and down over time. These fluctuations in service are a normal characteristic of railroading.”); *id.* at 3–4 (“Numerous factors cause railroad service to be inherently variable. One obvious factor is weather. . . . Other factors, such as unscheduled maintenance, contribute to the inherent

reason for this lies not in any absence of skill or commitment on the part of the NS workforce, who work intensively and creatively to find ways to achieve greater service quality and reliability for all of our customers. Rather, transit time variability stems from the fact that these efforts must deal with the nature of railroads—*i.e.*, interconnected and fixed outdoor networks with limited capacity that are subject to ever-changing traffic demands and a host of other effects, all of which are largely outside of a railroad’s control. Every ebb and flow of rail traffic volumes; every introduction of new rail services; every delay in loading or unloading a car at a customer facility or a ship at a port; every derailment, crossing accident, or storm that washes out track, will have impacts that ripple throughout the rail network, sometimes in unpredictable ways.¹⁵ Numerous operating witnesses have

variability of railroad service. . . . [I]t is also common for short-term service disruptions to result from a railroad’s efforts to expand capacity or improve existing facilities. . . . In some markets, seasonal changes in demand make railroad transportation especially subject to variation.”).

¹⁵ See, *e.g.*, Opening Comments of Norfolk Southern Railway Company, *Reciprocal Switching*, Docket No. EP 711 (Sub-No. 1) (filed Oct. 26, 2016), Verified Statement of Jeffrey H. Sliger, at 7, 9 (explaining how unexpected market shifts resulted in regional congestion that “affected the velocity and fluidity of Norfolk Southern’s entire system” and how additional handlings of a shipment negatively impact velocity because handlings “account for the majority” of transit time, and “[e]ach handling event consumes resources and introduces delays in transit time for the traffic.”); Supplemental Comments of Norfolk Southern Railway Company, *Reciprocal Switching*, Docket No. EP 711 (Sub-No. 1) (filed Feb. 14, 2022), Verified Statement of Cynthia M. Sanborn, at 3–5 (describing the complexity, volatility, and “[i]nherent uncertainty” of rail network operations and planning, as well as the vulnerability of the rail network to “[e]ven small changes,” which “can have a rippling effect across the network”); see also U.S. Dep’t of Transp., Off. of Inspector Gen., CR-2008-076, Root Causes of Amtrak Train Delays, at 19 (2008) (“The rail network operates 24 hours a day, 7 days a week. Delays in one part of the network can set off another type of delay later on in the network, which creates a ‘ripple’ effect. . . . It can take up to 5 days, and sometimes up to 1 month, to restore service to normal operations after an unplanned disruption. Unlike the aviation system, which allows planes to be repositioned overnight, there is no ‘down time’ within which trains can be repositioned.”).

shared this well-known feature of railroad operations with the STB over the long history of this proceeding.

Second, the proposal is oversampling, which can make rare events appear to be common ones. As proposed, service on a particular lane will fail this performance metric if there is a 20% or 25% change in a rolling 12-week window from the prior year transit time in that same window. There are, however, 365 such windows in a year. Every day is a new rolling 12-week window. Even if a 20% or 25% change in transit time were rare, if tested for 365 times, the mathematical likelihood of detecting just one such occurrence in a year increases dramatically.

For example, many in the D.C. region have spent time on the Bay Bridge on a trip to or from the Eastern Shore. The probability of there being some unusual traffic event that creates gridlock over the Bay Bridge may be quite remote—say one chance in a 100 (or 0.01%). But the likelihood of that remote event happening if you sample the route over the Bay Bridge 365 times in a year rises to 97.4% ($1 - (0.99^{365})$). All rare events will happen once if you wait long enough. In this case, by applying the proposed service metric 365 times a year, numerous corridors would trigger the metric a few times a year. And since it only takes a single “failure” of this performance metric to justify a forced switch under the proposal, the cumulative effect of repeated testing across thousands of difference corridors—of a service metric that fluctuates naturally—results in an unreliable performance metric that is capturing the majority of all corridors over the network.

The effect of oversampling is plainly illustrated in **FTI Table 3**. Reducing the sampling from 365 times year (daily) to 4 times a year (quarterly) dramatically reduces the probability of observing a given change in transit time that is not so common as it appears if all that is needed is one occurrence in a given year.

**FTI Table 3:
Likelihood that October 1, 2022 to September 30, 2023 Carloads Traveled
on a Lane Triggering Transit Time Increase Metric Compared to the Prior
Three Year Average Under Different Parameters¹⁶**

	Rolling 12-Week Comparison Interval	Quarterly Comparison Interval
15% Transit Time Threshold	75.4%	57.0%
20% Transit Time Threshold	69.0%	49.4%
25% Transit Time Threshold	61.9%	41.9%

Ultimately, normal variations in transit time—observed equally for dual-served or sole-served traffic—cannot be the basis for an intrusive competitive access remedy like a prescribed switch. Indeed, there cannot be a “compelling need,” as required by section 11102(c)’s public interest standard, for a prescribed switch based solely on a percentage of variation in transit time—more evidence would be needed to establish the existence of a service inadequacy. To be clear, NS strongly believes that transit time data can be valuable for monitoring network fluidity, and such data can be relevant when determining whether a customer has experienced inadequate service. However, the empirical analyses submitted by AAR’s experts show that the transit time metric, as proposed in the NPRM, simply cannot bear the weight of being a stand-alone measure of service inadequacy.

¹⁶ Baranowski & Zebrowski Reply V.S. at 9, Table 3. Note, the Table 3 copied above omits the column for “Rolling 6-Week Comparison Interval” included in the original as not relevant to the points discussed herein.

The Board must change either the way the metric is measured or the way it will be used. If the Board wishes to use this metric as a stand-alone performance standard, it should pause its work on this rulemaking and undertake a comprehensive data inquiry into industry transit time variability to better calibrate this proposed performance benchmark.¹⁷ On opening, Dr. Israel offered one methodology that uses a competitive benchmark to detect anomalous service levels that might create the kind of guardrails the STB intends. The public may have other suggestions. NS would also welcome the opportunity to work collaboratively with the Board and other stakeholders to craft an alternative metric for measuring transit time variability that meaningfully and reliably reflects service quality. Some parties, for example, have suggested the use of technical conferences or the like, to problem-solve issues with the proposed rule. NS would welcome constructive dialogue to explore how to rehabilitate the proposed transit time metric to develop a meaningful and reliable tool for both customers and the Board.

However, NS recognizes the STB's desire to move this proposal to a final rule to offer rail shippers additional relief options for inadequate service. One path to move this proceeding forward expeditiously, while addressing the serious problems with the transit time metric, would be for the STB to set aside this transit-time metric as a stand-alone measure of service inadequacy and, instead, consider evidence of transit time variability as only supplemental evidence of inadequate

¹⁷ Importantly, changing the period over which transit time is measured, as some parties have suggested, will not fix the serious problems with this metric. *See, e.g.*, AFPM Comments at 10–11 (suggesting that the Board use a “rolling average of three years” instead of a year-over-year period to measure transit time); Coalition Associations Comments at 31 (suggesting that Board include a three-year period in transit time metric). The FTI data make clear that moving to a three-year period results in the exact same flaws, and thus, a transit time metric based on a three-year period would be just as unreliable as a measure of service inadequacy as one based on a one-year period.

service, considered alongside the other performance standards and relevant evidence. NS believes all relevant evidence should be considered on a case-by-case basis.¹⁸ As such, petitioners should be allowed to use transit time variability data to argue that its service is inadequate—or even to argue that the incumbent carrier is intentionally padding its OETA to avoid triggering the OETA metric—so long as that evidence is treated as supplemental to the other performance standards. With this less mechanical approach, the Board could still utilize transit time variability data as valuable evidence, but not place more weight upon it than it can bear.

III. ONE SIMPLE CHANGE TO THE PROPOSED RULE CAN ADDRESS MULTIPLE CONCERNS AND HELP CONFORM THE PROPOSED RULE TO THE STATUTE.

On opening, the STB received a wide array of suggested improvements to its proposal. The number of suggestions should come as no surprise given the intricacy and novelty of the agency's proposal.

In NS's view, there are two categories of improvements that rise above all others. First, nowhere in the language of proposed Part 1145 does the Board mention or acknowledge the statutory standard, and multiple commenters have raised the fact that the Board has not sufficiently tied the NPRM to the requirements of section 11102(c).¹⁹ Second, multiple commentors have observed

¹⁸ See NS Comments at 4–5; *Midtec Paper Corp.*, 1 I.C.C.2d at 364 (“There is no mechanical test and the totality of the circumstances may be considered in these matters.”) (citing *Cent. States Enters., Inc. v. Seaboard Coast Line R.R.*, Docket No. FD 38891 (ICC served May 15, 1984), *aff’d Cent. States Enters., Inc. v. ICC*, 780 F.2d 664 (7th Cir. 1985)).

¹⁹ See, e.g., Opening Comments of CSX Transportation, Inc., *Reciprocal Switching for Inadequate Service*, Docket No. EP 711 (Sub-No. 2), at 9–13, 33–36 (filed Nov. 7, 2023) (“CSXT Comments”); CN Comments at 5–14; Opening Comments of the Association of American Railroads, *Reciprocal Switching for Inadequate Service*, Docket No. EP 711 (Sub-No. 2), at 5–9, 12–18 (filed Nov. 7, 2023) (“AAR Comments”).

that the approach needs more flexibility to account for a range of potential defenses.²⁰ The range of potential defenses not expressly considered by the proposal includes competitive benchmarking, statistical confidence, the curing of the service deficiency, a lack of any harm to the customer, a showing that the remedy will not resolve the service inadequacy or be in the public interest, and others.

The STB may be able to address these multiple concerns with a single stone. One way the Board can both better conform the program to the governing legal standard and build more flexibility into the analysis is to alter the language for proposed section 1145.3 in the following way, which expressly acknowledges section 11102(c)'s statutory standard and provides for a flexible approach to defenses raised by the incumbent carrier.

§ 1145.3 Affirmative Defenses.

The Board will consider any defense relevant to whether there is a service inadequacy for which there is actual necessity or compelling reason for a prescribed switching agreement, including, but not limited to, the defenses listed in this section. The defenses specified in this section are not exclusive and shall not create a presumption as to the validity of any other defense raised by a rail carrier.

With this change, many of the various fact-specific concerns raised by multiple parties on opening can be addressed through the affirmative defenses under the correct legal standard. This is a simpler solution than trying to add other specific categories of affirmative defenses, which could never be exhaustive in any event. Whatever the list of pre-approved affirmative defenses the Board selects, it is imperative and commonsensical to allow for an affirmative defense that shows that the public interest prong of section 11102(c) has not been satisfied. There is simply no reason to disallow an incumbent railroad from defending itself by establishing

²⁰ See, e.g., AAR Comments at 18; CSXT Comments at 4.

that the statutory standard is not satisfied. And the STB needs to explain the legal standard it plans to apply to affirmative defenses, or it leaves this entire section without an intelligible legal principle.

This single change could address a number of the recommendations NS offered on opening. For example, to show that the section 11102(c)'s public interest prong has not been satisfied, a defendant railroad could raise as a defense the competitive benchmarking analysis proposed by Dr. Israel; or the fact that a carrier has cured the temporary service problem; or that the forced switch will not remedy the service issues; or that the perceived violation of the final performance standards is attributable to few observations and noisy metrics, as demonstrated by the standard, well-accepted, and simple statistical tools described in Dr. Israel's attached Reply Verified Statement. And by unambiguously tying these defenses to the "actual necessity or compelling need" test that is required by section 11102(c)'s public interest prong, the STB will simultaneously offer needed guidance to the parties about how it intends to judge these and other affirmative defenses, and better conform the entire program to the governing legal standard.

CONCLUSION

NS shares the Board's focus on service and continues to work diligently to provide safe, reliable service to its customers, along with customer-centric ways to resolve and redress service issues if they arise. NS thanks the Board for considering its comments in this proceeding and urges the Board to adopt the changes proposed by NS and the AAR to ensure that any final rule conforms to the statute, furthers the Board's stated goals, and minimizes unintended harms.

Respectfully submitted,

/s/ Raymond A. Atkins
Raymond A. Atkins
Morgan B. Lindsay
SIDLEY AUSTIN LLP
1501 K Street, N.W.
Washington, D.C. 20005

*Counsel for Norfolk Southern Railway
Company*

Dated: December 20, 2023

EXHIBIT A

*Reply Verified Statement of
Mark A. Israel, Ph.D.*

**BEFORE THE
SURFACE TRANSPORTATION BOARD**

RECIPROCAL SWITCHING FOR INADEQUATE RAIL SERVICE

Docket No. EP 711 (Sub-No. 2)

REPLY VERIFIED STATEMENT OF

MARK A. ISRAEL, Ph.D.

December 20, 2023

I.	INTRODUCTION.....	1
A.	QUALIFICATIONS.....	1
B.	BACKGROUND AND ASSIGNMENT	1
II.	THE USE OF STATISTICAL METHODS TO IMPROVE THE RELIABILITY OF RECIPROCAL SWITCHING PROPOSALS	2
A.	ACCOUNTING FOR NORMAL VARIATION USING STANDARD ERRORS.....	2
B.	IMPLEMENTATION OF STATISTICAL ANALYSES.....	6
1.	<i>Service Reliability Ratio Analysis (OETA+24 Metric).....</i>	<i>6</i>
2.	<i>Application of Statistical Tests to the Board’s Proposed Analyses</i>	<i>8</i>
(a)	<i>Service Reliability Standard</i>	<i>8</i>
(b)	<i>Service Consistency Standard.....</i>	<i>9</i>
III.	CONCLUSION	11

I. INTRODUCTION

A. QUALIFICATIONS

1. My name is Mark A. Israel. I am the President of Compass Lexecon, an economic consulting firm where I have worked since 2006. From 2000 to 2006, I served as a full-time member of the faculty at Kellogg School of Management, Northwestern University. I received my Ph.D. in economics from Stanford University in 2001.
2. I filed comments in the opening round of this proceeding.¹ Those comments provided more detail about my qualifications and included my full curriculum vitae.

B. BACKGROUND AND ASSIGNMENT

3. In my prior comments in this proceeding, I proposed modifications to the proposal put forth in the Surface Transportation Board's ("the Board") September 7, 2023 Notice of Proposed Rulemaking ("2023 NPRM").² Those modifications included³:
 - Replacing the proposed analysis of OETA+24 with a Service Reliability Ratio ("SRR") that not only identifies whether there was a delay but also provides a measure of the severity of the delay.
 - Defining specific periods of analysis (annual quarters, for example) that eliminate the opportunity for shippers to search for—and base complaints on—a particularly favorable (to their case) 12-week period of their choosing that may not accurately reflect the shipper's typical service quality, much less any persistent service inadequacy.
 - Using a set of competitive benchmark traffic to calculate a threshold SRR—the upper bound of the range of expected outcomes in a competitive market.

¹ *Reciprocal Switching for Inadequate Service*, Docket No. EP 711 (Sub-No. 2, Verified Statement of Mark A. Israel, Ph.D., November 6, 2023 (hereinafter, "*Israel VS*")),

² *Reciprocal Switching for Inadequate Service*, Docket No. EP 711 (Sub-No. 2), Decision, Surface Transportation Board, decided September 5, 2023 (hereinafter, "*2023 NPRM*").

³ *Israel VS*, §§ IV and V.

- Comparing the mean SRR (plus/minus the standard error around the mean) of the traffic at issue to the threshold SRR to assess whether the service received by the complaining traffic was outside the range of expected outcomes based on the competitive benchmark.
4. I also noted that even if the Board decides to implement the proposed standards without any modifications, standard statistical tests should be incorporated into the Board’s proposed analysis to ensure that the analysis identifies service issues that are sufficiently different from normal fluctuations in service, and thus are more than just statistical noise, such that further regulatory consideration is warranted.
 5. I have been asked by Norfolk Southern (“NS”) to provide additional detail on the statistical tests that I suggest the Board incorporate into their proposed analysis. Specifically, I have been asked to explain why simple statistical tools like the ones I have proposed are important to apply to the types of metrics the Board is proposing. In particular, I have been asked to provide additional discussion of the statistical basis for including standard errors when comparing means to benchmarks, both in the case of comparing means from one dataset to a specific benchmark value (such as the competitive benchmark I propose or the Board’s proposed 60% ‘on-time’ threshold), and in the case of comparing means from different datasets to each other (as with the Board’s proposed transit time metric). Finally, I have been asked to provide more specific detail on how to implement the statistical tests I propose.

II. THE USE OF STATISTICAL METHODS TO IMPROVE THE RELIABILITY OF RECIPROCAL SWITCHING PROPOSALS

A. ACCOUNTING FOR NORMAL VARIATION USING STANDARD ERRORS

6. In my prior comments, I recommended modifications to the analysis of OETA+24 data that proposed comparing the petitioner’s mean SRR to a competitive threshold SRR. I noted that proper implementation of this recommendation requires estimating a standard error around the petitioner’s mean SRR to account for the statistical uncertainty that comes from normal variation in data. Put simply, the outcome from observed data needs to be far enough above the benchmark to be sure the difference is not just the “noise” inherent in real-world data, and adding standard errors accomplishes this.

7. I also noted that, even if the Board chooses to implement their proposed analyses without modification, it is appropriate—and necessary as a matter of proper statistical analysis—to incorporate standard statistical tests into the analysis of proposed service metrics to account for normal variation in data and the associated statistical uncertainty (as measured by standard errors).
8. I understand that some members have noted that the petitioner traffic at issue with respect to a reciprocal switching request would include all relevant traffic (rather than a sample of relevant traffic) and have asked for additional clarification with respect to the need to include standard errors in the analysis in this case.
9. I agree that it is relatively intuitive to understand the need to account for random variation when relying on samples of data. By definition, data samples contain only a subset of all available observations and different random samples are expected to have different means. Therefore, to account for the uncertainty associated with random sampling, it is standard statistical practice to estimate a range—captured by the standard error—around a sample mean to account for differences between the sample mean and the true population mean.⁴ Treating the mean as a range rather than a single point estimate provides a level of statistical confidence that the true population mean lies within that range.
10. In this case—in which the Board proposes to analyze service measures based on all rail shipments in a given lane over a given time period—the need to account for normal variation by including standard errors in the analysis may be less intuitive, but it is necessary for similar reasons.
11. Even when a dataset includes all observed outcomes over a given time period (minutes of delay for rail shipments over a given lane, for example), each observed outcome is affected by many factors given the state of the world at a specific point in time. However, the factors

⁴ Wooldridge, Jeffrey M. (2012), *Introductory Econometrics: A Modern Approach*, 5th edition, South-Western Cengage Learning (hereinafter, “*Wooldridge (2012)*”), at 344-345 (“Understanding why cross-sectional data should be viewed as random outcomes is fairly straightforward: a different sample drawn from the population will generally yield different values of the independent and dependent variables (such as education, experience, wage, and so on). Therefore, the OLS estimates computed from different random samples will generally differ, and this is why we consider the OLS estimators to be random variables.”).

that make up the state of the world at any given time are, in many ways, random (or stochastic, in formal statistical terms) and may impact those outcomes. That is, the data *could* have come out many ways, and much of what causes a particular set of data to be observed is random variation that has to be adjusted for. A standard statistics textbook explains this as follows⁵:

Formally, a sequence of random variables indexed by time is called a stochastic process or a time series process. (“Stochastic” is a synonym for random.) When we collect a time series data set, we obtain one possible outcome, or realization, of the stochastic process. We can only see a single realization, because we cannot go back in time and start the process over again. (This is analogous to cross-sectional analysis where we can collect only one random sample.) However, if certain conditions in history had been different, we would generally obtain a different realization for the stochastic process, and this is why we think of time series data as the outcome of random variables. The set of all possible realizations of a time series process plays the role of the population in cross-sectional analysis. The sample size for a time series data set is the number of time periods over which we observe the variables of interest.

12. With respect to rail transportation, it is easy to understand how random realizations of events such as heat waves or snow storms (or unexpected passenger train delays or unplanned emergency maintenance or changes in overall economic conditions) can impact rail service. These are stochastic processes that are subject to randomness due to the fact that they cannot be predicted with certainty. We observe outcomes that occurred in one state of the world (fair weather, for example), but if one of the factors that made up that state of the world were different (bad weather, for example), we would likely have observed different outcomes. Because the state of the world under which any given outcome occurred is random and the observed data is a particular realization of that random (or stochastic) process, it is accepted statistical practice to recognize that any estimate derived from a stochastic process comes with some degree of uncertainty.⁶ Put simply, there is “noise” inherent in any real-world set of data; sampling variation is one source of noise, but not the only one, and thus statistical tests are required to deal with the inherent uncertainty in real-world data, even when the data are not, strictly speaking, a sample.

⁵ Wooldridge (2012) at 345.

⁶ *Id.*

13. Therefore, when analyzing this data—calculating summary statistics such as averages, for example—it is accepted practice to account for this uncertainty by estimating a standard error around the mean. The standard error reflects the fact that there are different reasonable and expected average outcomes in different states of the world. Failing to account for this uncertainty is inconsistent with standard statistical practice and will often result in false positives, which occur when the data are different from the benchmark due only to the noise inherent in the data.
14. The formula for calculating the standard error of the mean is as follows, where σ is the standard deviation of the data and n is the number of observations in the data:

$$\textbf{Standard Error} = \frac{\sigma}{\sqrt{n}}$$

15. Calculating standard errors also accounts for the estimate imprecision arising when data contains a relatively small number of observations. For example, the average of shipment-level service metrics for a shipper with daily shipments during a 12-week period would provide a more accurate representation of the railroad’s overall service than the average for a shipper with only a few shipments during that 12-week period. This is true because a higher number of shipments would reflect more states of the world and the range of outcomes associated with all possible states of the world (including shipments affected by bad weather, or a passenger train delay, for example). With a smaller number of observations, there are fewer states of the world represented in the data, meaning there are fewer opportunities for outcomes associated with a particular state (bad weather) to average out with other possible states of the world (good weather) and thus, the average may be a less precise reflection of the overall service the railroad provided.
16. Stated differently, the number of observations reflect the amount of useful variation (or signal) in the data—the higher the number of observations, the higher the confidence in any particular statistical measure. The standard error formula captures this aspect via the denominator: as the number of observations (n) increases, holding σ the same, the standard error decreases. In the context of this case, this means that as the number of petitioner shipments increases, the certainty that the calculated mean statistic accurately reflects overall service levels also increases.

17. In the following section, I explain how to implement the statistical tests I recommend, both in the context of the modified SRR analysis that I propose and in the context of the Board's current proposals.

B. IMPLEMENTATION OF STATISTICAL ANALYSES

1. Service Reliability Ratio Analysis (OETA+24 Metric)

18. With respect to implementing the Service Reliability Ratio ("SRR") analysis that I described in my prior comments, the appropriate test compares the mean SRR of a petitioner's traffic to the competitive threshold SRR, allowing for normal variation as measured by the standard error.

19. More formally, the null hypothesis being tested is whether the mean SRR of the petitioner's traffic is at or below the competitive threshold SRR. If the null hypothesis is rejected—meaning that there is statistical evidence sufficient to conclude that the mean SRR of the petitioner's traffic *exceeds* the competitive threshold SRR—then there would be statistical evidence that the petitioner's service falls outside the bounds of normal service (as measured by the SRR) of the benchmark competitive traffic. Consider the following example:

20. Suppose that in a given quarter, the competitive threshold SRR is found to be 1.3 and that a petitioner made 30 shipments having a mean SRR of 1.33 with a standard deviation of 0.34.⁷

- The standard error of the mean (following the formula above) for petitioner's traffic is $0.35 / \sqrt{30} = 0.06$.⁸
- The hypothesis being tested is whether the petitioner's mean SRR (1.33) is at or below the competitive threshold SRR (1.3).

21. The null hypothesis discussed here leads to a one-sided statistical test (called a '*t test*') that calculates a standardized measure of the petitioner's mean SRR (called the 'test statistic' or *t*) and rejects the null hypothesis if the test statistic is large enough to rule out the conclusion

⁷ For this hypothetical, the petitioner had 10 shipments arrive within the OETA+24 timeframe (SRR=1), 10 shipments arrive 20% later than that timeframe (SRR=1.2), and 10 shipments 80% later than that timeframe (SRR=1.8).

⁸ See ¶ 14.

that the petitioner's SRR levels are at or below the competitive threshold SRR. A *t test* is a straightforward, standard statistical test that accounts for statistical variation by incorporating the standard error into the analysis and is consistent with standard statistical practice when comparing an average to a fixed value or when comparing averages across datasets.⁹ The test statistic (*t*) associated with the null hypothesis in this case is calculated as follows:

$$t = \frac{SRR_p - SRR_{cb}}{SE_{\bar{p}}}$$

Where:

- SRR_p is the mean SRR of the petitioner's traffic,
- SRR_{cb} is the competitive threshold SRR, and
- $SE_{\bar{p}}$ is the standard error of the petitioner's traffic SRRs.

22. The *t* statistic is then compared to a critical value for a one-sided test assuming a certain significance level.¹⁰ A significance level of 0.05 is common in these types of tests, where the lower the significance number, the more certainty there is in the result.¹¹
23. A p-value of the test can also be calculated. A p-value identifies the largest significance level at which the test could be carried out and still fail to reject the null hypothesis.¹² If the p-value is greater than the accepted significance level, then the null hypothesis cannot be rejected.
24. In our example, then, the test statistic (*t*) is $(1.33 - 1.3) / 0.06 = 0.54$. For a test with 30 observations, the critical *t* value (above which the null hypothesis can be rejected) at a 0.05 significance level is 1.70. The t-statistic in this example (0.54) is lower than these critical values, therefore, the null hypothesis cannot be rejected at the 0.05 significance level. Based on the number of observations and the *t*-statistic, the p-value is 0.30, meaning that only at a

⁹ *Wooldridge (2012)* at 780.

¹⁰ “*t* Tables” identify the critical values of *t* tests for one-sided and two-sided tests given the degrees of freedom in the test, equal to the number of observations minus one. *See, Wooldridge (2012)* at 833 for an example of a *t* Table.

¹¹ *Wooldridge (2012)* at 123-124.

¹² *Wooldridge (2012)* at 784.

significance level of 0.30 could the null hypothesis be rejected. This is above the generally accepted threshold of 0.05 for a one-sided test. As such, the null hypothesis—that the mean SRR of a petitioner’s traffic is at or below the competitive threshold SRR—cannot be rejected, meaning there is not sufficient statistical evidence to conclude that the petitioner’s service falls outside the bounds of normal service.

2. **Application of Statistical Tests to the Board’s Proposed Analyses**

25. In my prior comments, I recommended that, even if the Board decides to implement their proposed standards without any modifications, the Board should incorporate standard statistical tests to determine whether the petitioner’s traffic meets their proposed standards. This recommendation applies to both the Board’s Service Reliability Standard and the Service Consistency Standard.

(a) Service Reliability Standard

26. If the Board chooses to adopt the proposed Service Reliability Standard, I suggested that—rather than relying on a direct comparison of a petitioner’s ‘on-time’ performance to the proposed 60% threshold—the Board should determine whether a petitioner’s traffic meets the 60% ‘on-time’ delivery standard using a *t test*.

27. More formally, I propose that the Board test the null hypothesis that the proportion of a petitioner’s traffic that was delivered within an acceptable service window is greater than or equal to 60%. If the null hypothesis is rejected, one can conclude with confidence that there is statistical evidence that the petitioner’s traffic failed the Board’s 60% ‘on-time’ performance standard and thus that more investigation of the source of this failure is warranted.

28. The appropriate statistical test for this null hypothesis is:

$$t = \frac{60\% - OETA_p}{SE_{\bar{p}}}$$

Where:

- $OETA_p$ is the percentage of petitioner traffic delivered within the Board’s ‘on-time’ service window of OETA+24 hours, and

- $SE_{\bar{p}}$ is the standard error of the petitioner's 'on-time' (i.e., OETA+24) performance.

29. This test is an important and straightforward added step because it is possible, especially with relatively small sample sizes, for a petitioner's average 'on-time' performance to be less than 60% because of statistical noise. As noted above, such noise is present in all real-world data, and is of particular concern when the number of observations is small. Implementing statistical tests based on the standard error accounts for this noise.
30. Suppose, for example, that a petitioner had 10 shipments in a given period, 5 of which met the OETA+24 standard and 5 of which did not. In this case, the 'on-time' performance is 50%, with a standard deviation of 0.50. Using the formula above, the test statistic (t) is $(60\% - 50\%) / 0.158 = 0.63$. The p -value is 0.27, which is not significant at a significance level of 0.05. Therefore the null hypothesis cannot be rejected.
31. If, however, a petitioner had 100 shipments in a given period with an OETA performance of 50%, the test statistic (t) is $(60\% - 50\%) / 0.05 = 2$. The p -value is 0.024, which is significant at a significance level of 0.05. In this case the null hypothesis is rejected, providing statistical evidence that the petitioner's traffic failed the Board's 60% 'on-time' performance standard and more investigation of the source of this failure is warranted.

(b) Service Consistency Standard

32. If the Board chooses to adopt the proposed Service Consistency Standard, the Board should include a t test to determine whether any difference in average transit times is statistically greater than 20%.¹³
33. More specifically, I propose that the Board incorporate a t test to test the null hypothesis that the difference in mean transit times between the two periods is less than or equal to 20%. If the t test shows that the null hypothesis is rejected, we conclude that the difference in transit time is greater than 20%. The appropriate t statistic to test this null hypothesis is as follows:

¹³ I understand the Board has requested comments on whether the 20% threshold proposed in the 2023 NPRM is appropriate. I use 20% here, but the concept is applicable to any threshold the Board adopts.

$$t = \frac{(\bar{X}_1 - \bar{X}_0) - (20\% \times \bar{X}_0)}{\sqrt{\left(\frac{(N_1 - 1)s_1^2 + (N_0 - 1)(1.2 \times s_0)^2}{N_0 + N_1 - 2}\right) \left(\frac{1}{N_0} + \frac{1}{N_1}\right)}}$$

Where:

- $\bar{X}_1$ is the mean transit time of the current period,
- $\bar{X}_0$ is the mean transit time of the prior period being compared,
- N_1 is the number of observations in the current period,
- N_0 is the number of observations in the prior period,
- s_1 is the standard deviation of the observations in the current period, and
- s_0 is the standard deviation of the observations in the prior period.

34. This test is an important and straightforward added step because it is possible, especially with relatively small sample sizes (as explained above), for the difference in two means to be greater than 20% because of statistical noise.

35. Consider an example where a shipper had 5 total shipments in a recent calendar quarter.

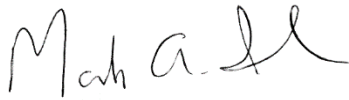
Suppose that the transit times for the 5 shipments were 96 hours, 120 hours, 144 hours, 144 hours, and 240 hours. The average of these transit times is 148.8 hours, with a standard deviation of 49.0. Suppose, then, that in the prior year, that shipper had only 3 shipments with transit times of 96 hours, 120 hours, and 144 hours. The average of these transit times is 120 hours with a standard deviation of 19.6. The increase in average transit time between these sets of shipments is 24%, but given the small number of observations, one should be concerned that the difference is due to statistical noise. Using a t test to assess whether the difference in transit time exceeds the Board's 20% standard can account for these types of issues and lead to more reliable results. Referencing the formula above, the test statistic (t) is calculated as follows:

$$t = \frac{(148.8 - 120) - (20\% \times 120)}{\sqrt{\left(\frac{(5 - 1)49.0^2 + (3 - 1)(1.2 \times 19.6)^2}{5 + 3 - 2}\right) \left(\frac{1}{5} + \frac{1}{3}\right)}}$$

36. The result is a t statistic of 0.16 and a p -value of 0.44. We would therefore fail to reject the null hypothesis that the difference in mean transit times between the two periods is less than or equal to 20% at a significance level of 0.05.

III. CONCLUSION

37. The additional statistical tests that I propose are consistent with standard practice, can be easily implemented, and are necessary to ensure that the Board's findings are accurate and reliable.



Mark A. Israel
December 20, 2023