Revisiting the X:BOT Naltrexone Clinical Trial Using a Comprehensive Survival Analysis

Elizabeth M. Ajazi, DrPH, Nabarun Dasgupta, PhD, Stephen W. Marshall, PhD, Jane Monaco, DrPH, Annie Green Howard, PhD, John S. Preisser, PhD, and Todd A. Schwartz, DrPH

Objectives: This paper illustrates survival models for analysis of trials of substance use treatment programs. It uses public release data from a study of extended-release naltrexone (XR-NTX), relative to buprenorphine-naloxone (BUP-NX).

Methods: We used publicly available data from the X:BOT trial (n = 570), which compared XR-NTX to BUP-NX on 2 efficacy outcomes (opioid relapse, use of nonprescribed opioids; positive opioid urine test) and 1 safety outcome (overdose). Intention-to-treat (ITT) and per-protocol approaches were implemented using survival models that included treatment-by-time interactions.

Results: Consistent with the original trial findings, 72% of XR-NTX and 94% of BUP-NX subjects initiated treatment; the ITT hazard ratio for XR-NTX relative to BUP-NX was 1.40 (95% confidence interval: 1.13, 1.73; $P < 0.01$) for opioid relapse and 1.31 (1.07, 1.60; $P = 0.01$) for positive urine test. Using treatment-by-time interactions, we examined the time-dependent effect of XR-NTX and found an elevated ITT overdose hazard ratio of 2.4 (1.1, 5.3; $P = 0.03$) overall and 3.8 (1.2, 11.6; $P = 0.02$) during the study treatment phase. This result (28 overdoses overall; 17 overdoses during the study treatment phase) contrasts with the previous analysis, which reported minimal differences in overdose between XR-NTX and BUP-NX.

Conclusions: An advantage of using time-dependent Cox models is its ability to isolate effects during specific periods. In general, our survival analyses concur with the conclusions of Lee et al (2018) for the efficacy outcomes, which demonstrated superiority of BUP-NX. In contrast to the original report, our analysis indicates a greater risk of overdose for XR-NTX, predominantly during the study treatment phase. Further investigation of this finding is a pressing research priority.

Key Words: intention-to-treat, per-protocol, randomized controlled trial, substance use disorder

(*J Addict Med* 2021;xx: xxx–xxx)

Opioid use disorder (OUD) is a chronic condition that affects at least 3 million people in the United States and 16 million people worldwide.[1] There are 3 medications that have been approved for abstinence-based OUD recovery[2,3]: injectable extended-release naltrexone (XR-NTX), buprenorphine-naloxone (BUP-NX), and methadone. Randomized open-label trials are the primary means of establishing safety and efficacy for OUD recovery. However, there are numerous complexities in conducting trials. These include loss to follow-up that may be differential between treatments, resulting in follow-up time that varies between subjects.

This paper illustrates the use of 2 types of survival analysis models to partially address these issues in the data analysis phase of a trial. Standard Cox models can be used to estimate the treatment effect on average across the entire follow-up period, and Cox models with treatment-by-time interactions can be used to estimate treatment effects specific to defined subintervals of the overall follow-up period.

Example Randomized Controlled Trial: X:BOT Trial

X:BOT was a randomized, multi-center, open-label, clinical trial conducted from 2014 to 2016, comparing XR-NTX and BUP-NX over 24 weeks. X:BOT (Trial CTN-0051) was supported by the National Institute on Drug Abuse. The purpose of X:BOT was to determine the comparative effectiveness and safety of XR-NTX and BUP-NX. A total of 570 patients were 1:1 randomized.[4] The primary outcome of interest was time-to-relapse, where relapse was defined as the use of any nonprescribed opioids, starting 21 days after randomization. Secondary outcomes of opioid use other than treatment medications, and adverse events, including overdose, were also analyzed.[4] Nontreatment opioid use during the study treatment phase was measured by a weekly patient report[5] and urine drug tests.[6] Spontaneous adverse events were reported to study clinicians by study participants.[7]

From the Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC (EMA, JM, AGH, JSP, TAS); Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, NC (ND, SWM).

Received for publication December 31, 2020; accepted June 20, 2021.

Dr. Marshall is partly supported by an award from the National Center for Injury Prevention and Control for an Injury Control Research Center (R49CE003092).

ND is a scientific advisor to the RADARS System of Denver Health and Hospitals Authority, Denver, Colorado.

The authors report no conflicts of interest.

Supplemental digital content is available for this article. Direct URL citation appears in the printed text and is provided in the HTML and PDF versions of this article on the journal's Web site (www.journaladdictionmedicine.com).

Send correspondence to Stephen W. Marshall, PhD, University of North Carolina at Chapel Hill, 725 MLK Jr Blvd., CB 7505, Chapel Hill, NC 27599. E-mail: smarshall@unc.edu.

ISSN: 1932-0620/16/0000-0001

DOI: 10.1097/ADM.0000000000000931

Naltrexone is a pure mu-opioid receptor antagonist. Buprenorphine is a partial mu-opioid receptor agonist with antagonist properties. Therefore, both medications require patients to have metabolized and eliminated exogenous opioid agonists and to effectively be in early stages of withdrawal. In X:BOT, BUP-NX was provided as a daily, self-administered, sublingual film and was initiated once withdrawal symptoms appeared during detoxification. In contrast, patients receiving XR-NTX required complete detoxification, negative opioid urine test results, and a negative naloxone challenge test.[4] XR-NTX injections (4 mL) were scheduled every 28 days.

Induction proportions differed markedly between treatment groups in X:BOT. Induction is defined as a status such that a randomized individual received an injection of XR-NTX or was dispensed BUP-NX.[6] A total of 28% of patients in the XR-NTX arm, compared to 6% in the BUP-NX arm, were not able to be inducted into treatment.[4] The original analysis using an intention-to-treat (ITT) approach indicated the risk of relapse was higher in XR-NTX versus BUP-NX (hazard ratio [HR] = 1.36, 95% confidence interval [CI] 1.10–1.68).[4] The original report[4] noted that this was largely due to the early relapse among XR-NTX participants who did not complete induction.[4] Accordingly, the original report[4] also presented a per-protocol (PP) type analysis, limited to participants successfully inducted, concluding the hazard of relapse was similar for XR-NTX versus BUP-NX (HR 0.92, 95% CI 0.71–1.18).[4]

An acknowledged limitation of the original report[4] was that, in contrast to the primary outcomes, time-to-event (survival) models were not used for urine and overdose outcomes. Thus, time-independent comparative effectiveness assessment for those 2 outcomes reported in the original paper[4] could potentially have been affected by the differences in induction and drop-out between the 2 arms.[8] Furthermore, the ITT and PP time-to-event analyses assumed a constant treatment effect over time for the relapse endpoint. As noted in the original report, the data supported this assumption for the PP models but not the ITT models.[4]

The primary purpose of this paper was to illustrate the usefulness of Cox models, particularly those involving treatment-by-time interactions, for the analysis of trials of treatments of OUD. To illustrate the use of these models, we analyzed the publicly available data from X:BOT,[9] to examine the efficacy (opioid relapse and positive opioid urine test) and safety (overdose) outcomes using time-independent and time-dependent survival models. Specifically, the models used in this paper illustrate simple (time-independent) Cox survival analysis models for time-to-relapse, time-to-first nonnegative opioid urine test, and time-to-overdose. In addition, given evidence for the presence of time interactions in the ITT results presented in the original report,[4] we also utilized Cox models with treatment-by-time interactions with the objective of differentiating effects during the pretreatment, treatment, and posttreatment phases.

METHODS

X:BOT Trial Design

X:BOT was an open-label randomized clinical trial that compared XR-NTX and BUP-NX over a 24 week follow-up period on 2 efficacy outcomes (opioid relapse and positive opioid urine test) and 1 safety outcome (overdose), conducted from 2014 to 2016. Eight study sites were used to recruit patients who were 18 years or older, spoke English, had an OUD (as specified by the Diagnostic and Statistical Manual of Mental Disorders-5), and had any form of nonprescribed opioid use in the last 30 days.

Patients

Of the 570 patients in this study, the majority were white (78%), male (70%), and their average age was 34 years. For the ITT approach, the public use data file contained 283 participants randomized to the XR-NTX group and 287 randomized to the BUP-NX group, identical to the reported treatment allocation numbers[4] Following the approach for the PP approach[4] we included only individuals who were inducted into the study. This yielded 204 individuals in the XR-NTX group and 270 individuals in the BUP-NX group, identical to the numbers reported.[4]

Exclusion Criteria

Patients were excluded if they had any serious medical, psychiatric, or nonopioid substance use disorder. Other exclusion criteria included transaminase concentrations greater than 5 times the upper limit of normal, suicidal, or homicidal idealization, an allergy or sensitivity to XR-NTX or BUP-NX, prior methadone maintenance treatment, chronic pain requiring the use of opioids, legal obligations, inability to safely receive intramuscular injections, and any women who were pregnant, breastfeeding, planning to conceive, or unwilling to use contraception.

Survival Analysis Outcomes

Time-to-Relapse

Opioid relapse was defined as ''4 consecutive weeks of any non-study opioid use by urine toxicology or self-report, or 7 consecutive days of self-reported use.''[4] The publicly available data files included a preconstructed binary relapse outcome variable, which we used for our analysis. We analyzed time-to-relapse using Cox regression to compare treatment arms, where time-to-relapse was defined as non-study opioid use measured from date of randomization (day 0). For the ITT approach, individuals who were not successfully inducted were considered to have relapsed on the date of induction failure.

Time-to-Nonnegative Weekly Urine Test

Lee et al[4] analyzed weekly opioid-negative urine samples with a mixed-effects model where missing tests were considered positive. A positive opioid urine test was defined as ''non-study opioids (buprenorphine, methadone, morphine [heroin, codeine, morphine], or oxycodone).''[4] In this reanalysis, we created a new variable from the public-use urine toxicology data file[9] of opioid urine test results. We defined a positive opioid urine test in the same way[4] and analyzed the time-to-first positive opioid urine test for each individual, starting at randomization; occurrences of subsequent positive tests were ignored. In this reanalysis we considered missing

tests as positive,[4] but also considered missing tests as missing to investigate the impact of the strong assumption that all missing tests would have been positive, had they been available.

Time-to-Overdose

Overdose events were precoded on the publicly available data files as a binary variable for any overdose, using date of first overdose as reported spontaneously and supported by medical record review.[7] Overdose was examined as a secondary endpoint in X:BOT because of its considerable public health importance as the adverse event of greatest consequence. Only 8 participants had an opioid specified as the substance involved overdose, but the adverse event reporting system used in X:BOT means substance could not be guaranteed to have been systematically collected by the study procedures. One participant had an unknown substance overdose, 1 participant had a ''narcotic'' overdose, and 1 participant had a cocaine overdose. All other overdoses (n = 17, 61%) in the dataset did not include a type of overdose. Thus, for our analysis, we included any overdose event, irrespective of the substance(s) involved.

Data Analysis

Consistent with the original aims of the X:BOT trial,[4] unadjusted Cox regression was used to compare the 2 treatments for 3 outcomes: opioid relapse, positive opioid urine test, and overdose. For each outcome, patients who did not experience the outcome (opioid relapse, positive opioid urine test, or overdose event) were right censored on the participant's last recorded date of observation. We also conducted a Fisher exact test with the publicly available data to directly compare overdose proportions.[4]

Graphical inspection of the public use dataset demonstrated a strong violation (nonconstant HR over time) in both the ITT and PP analyses for all 3 outcomes (Fig. 1): opioid relapse, positive opioid urine tests, and overdose event outcomes. The log-negative log survival curves represent cumulative incidence over time, and show a clear violation in proportionality due to the nonparallel curves. Thus, in time-dependent Cox models (see Appendix, Supplemental Digital Content 1, http://links.lww.com/JAM/A311, which illustrates regression analysis) time-by-treatment interactions were included to allow HRs to be computed for 3 different periods of follow-up: pretreatment, treatment, and posttreatment. The pretreatment period started at day 0 until day 21, treatment was from day 22 until day 168 (24 weeks), and follow-up was defined as 169+ days.[4] The original report[4] did not include any time-by-treatment interactions, but did describe nonconstant HRs over time in the ITT approach and included a graphical presentation of nonconstant HRs.

Ethics Oversight

This work was reviewed and determined to be exempt by the University of North Carolina at Chapel Hill IRB.

RESULTS

Using publicly available data,[9] we were unable to fully replicate the results in the original report.[4] Specifically, there were small differences in counts of relapse and overdose events between the published results[4] and the counts we computed from the publicly available data files (Table 1). The absolute differences were small, never larger than 5 for any outcome in any arm. However, for the overdose outcome, the number of outcomes was low. Therefore, the difference in overdoses were large as a proportion of the reported counts (about 30% higher).

Time-to-Relapse

For the ITT analysis, the public data file contained 180 relapses in the XR-NTX arm (n = 283) and 159 relapses in the BUP-NX arm (n = 287), compared to 185 and 163 in the Lee et al[4] analysis, respectively. For the PP analysis there were 101 relapses in the XR-NTX arm (n = 204) and 145 relapses in the BUP-NX arm (n = 270), compared to 106 and 150[4] analysis, respectively. Due to these slight differences, this reanalysis produced similar but slightly different HRs (Table 1).

The hazard of opioid relapse was 1.40 times higher (95% CI: 1.13, 1.73; $P < 0.01$) in the XR-NTX arm than the BUP-NX arm on average during the course of the study using an ITT approach (Table 1). Using models with treatment-by-time interaction, during the study treatment phase there was a 45% increase in the hazard of relapse for XR-NTX compared to BUP-NX (HR = 1.45; 95% CI: 1.17, 1.81; $P < 0.01$) using an ITT approach (Table 2).

Using the public data files we found a higher nominal percentage of opioid relapse for BUP-NX (145/270) versus XR-NTX (101/204) for the PP approach. There was no statistically significant difference in opioid relapse over time between the 2 groups using the PP approach, however, the HR suggested a slight protective advantage for relapse prevention for XR-NTX compared to BUP-NX (HR = 0.89; 95% CI: 0.69, 1.14). Lee et al[4] stated that for the PP approach there was a higher percentage of opioid relapse events for the BUP-NX group (150/270) versus the XR-NTX group (106/204), but no difference was found over time for relative hazard of relapse.

Time-to-Nonnegative Urine Test

For the time-to-nonnegative urine test analysis using an ITT approach, there was a 31% increase (HR = 1.31; 95% CI: 1.07, 1.60; $P = 0.01$) in the hazard of a positive opioid urine test with XR-NTX compared to BUP-NX averaged over the entire observation period (Table 1). Treatment-by-time models indicated that this increase was concentrated in pretreatment phase. During pretreatment the risk of a positive opioid urine test was 71% higher in XR-NTX arm relative to the BUP-NX arm (HR = 1.71; 95% CI: 1.30, 2.23; $P < 0.01$).

Using survival analysis, similar results were found when missing urine tests were considered missing and for the PP approach. This contrasts with the reported analysis of the number of weekly opioid negative urine samples.[4]

Time-to-Overdose

The public data file[9] contained 19 individuals in the XR-NTX arm (n = 283) and 9 individuals in BUP-NX (n = 287) who had at least 1 overdose. Fewer individuals with overdoses in the XR-NTX arm were reported[4] in the original analysis (n = 15) than the 19 found in the public data

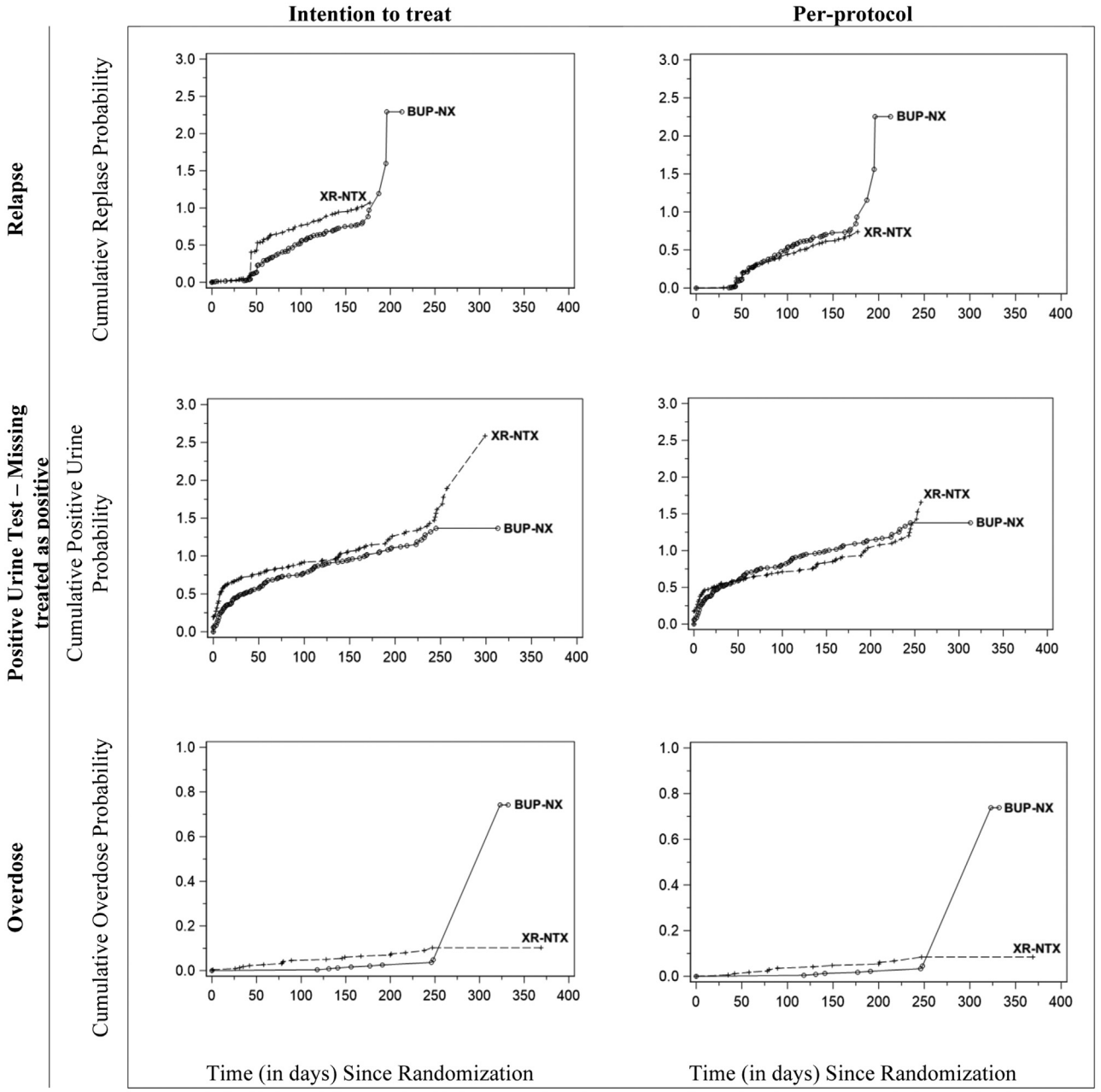

FIGURE 1. Negative log survival curves for time to first opioid relapse event, first positive urine test, and overdose for naltrexone (XR-NTX) and buprenorphine-naloxone (BUP-NX). The plots represent the accumulated number of events (vertical axis) over time (horizontal axis). The plots reveal violation of the proportional hazards assumption in Cox statistical models used in the original trial analysis. The nonlinearity of the lines, both at early and late time points, suggests that time periods need to be modeled separately to not violate basic requirements.

file. However, the total number of fatal overdose events in the public data file[9] was the same as originally reported[4] ($n = 5$ events). In contrast to the conclusion of the original report[4] that ''both treatments are equally safe once initiated,'' we found a statistically significant difference between treatments using a time-to-event analysis. Using a PP approach, we found a protective factor for overdose prevention with BUP-NX compared to XR-NTX (HR: 2.10; 95% CI: 0.86, 5.14). During the study treatment phase the risk of overdose was 3.81 (1.01, 14.36) times higher in XR-NTX compared to BUP-NX when using PP. Similar results were found for an ITT approach (Table 1).

DISCUSSION

Our survival analyses concur with the conclusions of Lee et al (2018) for the efficacy outcomes of relapse and

TABLE 1. Comparison of Reanalysis to the Original Study Findings (Lee et al, 2018) for a RCT Comparing Naltrexone (XR-NTX) to Buprenorphine-naloxone (BUP-NX) With Efficacy (Opioid Relapse; Positive Opioid Urine Test) and Safety (Overdose) Outcomes

		X:BOT Publicly Available Data File					X:BOT Original Study Findings (Lee et al, 2018)		
	Total	XR-NTX	BUP-NX	Avg. HR* (95% CI)†	CLR†	*P*	XR-NTX	BUP-NX	Published Results
Total sample size									
Intention to treat, n	570	283	287	n/a	n/a	n/a	283	287	n/a
Per-protocol, n (% of arm)	474	204 (72%)	270 (94%)	n/a	n/a	n/a	204 (72%)	270 (94%)	n/a
Relapse (n, % of arm)									
Intention to treat	339	180 (64%)	159 (55%)	1.40 (1.13, 1.73)	1.53	<0.01	185 (65%)	163 (57%)	HR* = 1.36 (95%CI†: 1.10, 1.68; *P* = 0.004)
Per-protocol	246	101 (50%)	145 (54%)	0.89 (0.69, 1.14)	1.65	0.36	106 (52%)	150 (56%)	HR* = 0.92 (95%CI†: 0.71, 1.18; *P* = 0.49)
First nonnegative urine test (missing treated as positive) (n, % of arm)									
Intention to treat	381	191 (67%)	190 (66%)	1.31 (1.07, 1.60)	1.50	0.01	n/a	n/a	Median weekly nonnegative urine samples: XR-NTX = 4; BUP-NX = 10 *P* < 0.001‡
Per-protocol	315	131 (64%)	184 (68%)	1.03 (0.82, 1.28)	1.56	0.82	n/a	n/a	Median weekly nonnegative urine samples: XR-NTX = 13; BUP-NX = 11 *P* = 0.81‡
First nonnegative urine test (missing treated as missing) (n, % of arm)									
Intention to treat	377	188 (66%)	189 (66%)	1.30 (1.06, 1.59)	1.50	0.01			Not analyzed
Per-protocol	311	128 (63%)	183 (68%)	1.01 (0.81, 1.27)	1.57	0.92			
Overdose (n, % of arm)									
Intention to treat	28	19 (6.7%)	9 (3.1%)	2.40 (1.09, 5.30)	4.86	0.03	15 (5.3%)	8 (2.8%)	Fisher exact *P* = 0.14
Per-protocol	20	12 (5.9%)	8 (3.0%)	2.10 (0.86, 5.14)	5.98	0.10	9 (4.4%)	7 (2.6%)	Fisher exact *P* = 0.31

*Avg. HR, average hazard ratio (HR) over time; reference group BUP-NX.
†95% CI, 95% confidence interval; CLR, confidence limit ratio.
‡*P*-value from mixed effect model.

TABLE 2. Comparison of Naltrexone (XR-NTX) to Buprenorphine-naloxone (BUP-NX) Using a Treatment-by-time Interaction Model to Estimate Hazard Ratios During Pretreatment, Treatment, Follow-up, Using X:BOT Publicly Available Data File*

	Pretreatment (0–21 days)						Treatment (22–168 days)						Posttreatment (169+ days)					
	XR-NTX	BUP-NX	HR†	95% CI‡	CLR†	*P*	XR-NTX	BUP-NX	HR†	95% CI‡	CLR†	*P*	XR-NTX	BUP-NX	HR†	95% CI†	CLR†	*P*
Opioid relapse																		
ITT	6	5	1.21	(0.37, 3.97)	10.7	0.75	172	148	1.45	(1.17, 1.81)	1.6	<0.01	2	6	0.31	(0.06, 1.61)	26.8	0.16
Per-protocol	0	0		n/a§			99	139	0.91	(0.71, 1.18)	1.7	0.50	2	6	0.31	(0.06, 1.61)	26.8	0.16
Positive opioid urine – missing tests treated as positive																		
ITT	123	95	1.71	(1.30, 2.23)	1.7	<0.01	44	77	0.78	(0.54, 1.12)	2.1	0.18	24	18	1.48	(0.80, 2.74)	3.4	0.21
Per-protocol	76	93	1.26	(0.93, 1.70)	1.8	0.14	35	76	0.65	(0.43, 0.96)	2.2	0.03	20	15	1.52	(0.78, 2.98)	3.8	0.22
Positive opioid urine – missing tests treated as missing																		
ITT	122	93	1.74	(1.33, 2.28)	1.7	<0.01	41	77	0.74	(0.51, 1.08)	2.1	0.12	25	19	1.35	(0.74, 2.48)	3.4	0.33
Per-protocol	75	91	1.29	(0.95, 1.76)	1.9	0.10	32	76	0.60	(0.40, 0.91)	2.3	0.02	21	16	1.38	(0.71, 2.66)	3.8	0.34
Overdose event																		
ITT	1	0		n/a§			13	4	3.77	(1.23, 11.57)	9.4	0.02	5	5	1.10	(0.32, 3.79)	11.8	0.88
Per-protocol	0	0		n/a§			8	3	3.81	(1.01, 14.36)	14.2	0.05	4	5	1.09	(0.29, 4.06)	14.0	0.90

*Reference group BUP-NX.
†95% CI, 95% confidence interval; CLR, confidence limit ratio; HR, hazard ratio.
‡Individuals who experienced a relapse during pretreatment were not included in the per-protocol approach.
§Unable to compute HR due to zero cell counts.

positive urine test. In contrast to the original report, our analysis indicated a greater risk of overdose for XR-NTX, predominantly during the study treatment phase. The differences in findings is due in part to relatively minor discrepancies in overdose counts (Table 1) between the publicly available data and the original analysis.[4] Replication of analyses using publicly funded, publicly available clinical trial data are a critical tool for ensuring confidence in major trials.

The implications call into question the belief that XR-NTX prevents overdose or carries no elevated overdose risk. Although spontaneous adverse event reporting in clinical trials is normative, there is increasing recognition that RCTs are inadequate for assessing safety endpoints unless they have been designed to do so intentionally.[10] In the case of the X:BOT trial, prespecification of and powering for overdose may have been prohibitively expensive. Yet even with technical caveats, highlighting null associations based solely on spontaneous event data is also problematic.

Using[10] treatment-by-time interactions, we were able to isolate efficacy and safety effects during the treatment phase from the pretreatment and follow-up phases. This is a major advantage of using time-dependent survival models for all 3 outcomes, however, we acknowledge that the number of relapse and overdose events is too small in the pretreatment and follow-up phases to permit inferences that would be clinically meaningful. Thus, the HRs presented in Table 2 for pretreatment for overdose and relapse, and for all 3 outcomes in the posttreatment phase, should be considered to be illustrative rather than clinically informative.

As in the original report, our analyses using public data files found very different HR estimates for the ITT and PP approaches. Generally, for efficacy outcomes, an ITT approach is more appropriate because it offers conservative estimates of effectiveness.[11] In X:BOT, if we exclude patients not inducted (PP) then the treatment groups are unbalanced, and the use of PP can result in selection bias and nongeneralizability.[12] Since patients who cannot be initiated on treatment are generally considered to be different than those who can be inducted, the results of a PP approach could be argued to have limited practical clinical validity unless induction is assumed to occur completely at random.[13] In this study it is clear that noninduction is not at random, as the initiation of XR-NTX is more difficult than BUP-NX. Lee et al[4] reports that since the results from the PP approach for opioid relapse were nonsignificant, there is no difference between the 2 groups once treatment is established, and they fully acknowledge the differences in induction rates. We concur with the conclusions of the original report[4] that the 2 treatments are similar in terms of relapse once treatment is initiated. Knowing that it is more challenging to initiate XR-NTX than BUP-NX should encourage clinicians to monitor patients on this medication more closely during the initial 3 weeks, provide harm reduction support, and make both agonist therapies available for those who discontinue.

Our time-dependent analyses using an ITT approach showed that during the study treatment phase there was an increased risk of opioid relapse for XR-NTX compared to BUP-NX, but the conclusions found during the study treatment phase do not appear to carry over to the posttreatment phase. This could indicate that although individuals have a more difficult time initiating XR-NTX treatment, long-term results could be promising. Therefore, further investigation could determine whether it is reasonable to make both treatment modalities available to patients, and have a clinical plan to switch to agonist therapy for those who discontinue XR-NTX early. In routine medical care, XR-NTX discontinuation after 30 days was reported in more than half of patients.[14]

In contrast to opioid relapse, our results for positive opioid urine test and overdose events differ very little between ITT and PP approaches. However, results for the time-dependent positive urine test analysis differed between time periods. During posttreatment, there was an increased hazard of positive opioid urine tests for XR-NTX compared to BUP-NX, but during the study treatment phase, there was a decreased hazard. This insight would not available without the use of time-dependent models and is more robust to the effect of differential noninduction (by treatment arm) than the analysis used in the original report.[4]

We found conflicting results for our time-to-overdose analysis compared to the Fisher exact test in the original report.[4] We found statistically significant results for the risk of overdose for XR-NTX compared to BUP-NX during the study treatment phase and marginally statistically significant results across all time periods; however, this analysis reflects small numbers of overdoses. In contrast, no difference in the proportion of overdoses between treatment arms was reported.[4] Additionally, X:BOT did not use overdose as a primary outcome, therefore, we caution that these data may not have been powered to accurately captured differences between the treatment arms.

Finally, since we used Cox models for all 3 outcomes, we were able to observe a difference in the magnitude of the HRs between the opioid relapse and opioid urine models. It might be common for a participant to experiment with opioid use and experience the negative reinforcement of attenuated psychotropic effects of opioids while under naltrexone mu-opioid receptor blockade. XR-NTX pharmacokinetics may also be at play. XR-NTX has an initial plasma concentration peak at 2 hours post administration, and a second peak 2 to 3 days later, with declining levels after 14 days.[15] There may be differences in the completeness of blockade between XR-NTX and BUP-NX, whereby mild intoxicating effects of opioids may be discernible with the latter.[16] During the pretreatment period 43% of the XR-NTX group and only 32% of the BUP-NX group tested positive for opioids. If experimentation during the early stages of treatment is a natural course of treatment, the pharmacological properties of treatment choice may directly influence induction proportions. This has impact on clinical treatment delivery. A harm reduction approach during induction with XR-NTX would be ethically responsible knowing that continued use is highly prevalent and that other treatments afford greater protection from use during this period.

This study has several limitations, primarily that we were unable to completely replicate the analysis variables from the original report published[4] despite communication with that team. Although we found very similar counts for

opioid relapse and overdose events, the final outcome variable and programming code used in the original report were not available to us. Some variables had limited numbers of events overall or by follow-up period in our reanalysis.

Additionally, one untestable assumption, and therefore, limitations to proportional hazards models, is the presence of general noninformative censoring due to loss-to-follow-up. This assumption would need to be tested in sensitivity analyses in the future.

CONCLUSIONS

These time-dependent analyses illustrate the use of survival models, particularly models with treatment-by-time interactions, in trials of substance use disorder treatment programs. Results from these models largely concur with reported conclusions,[4] particularly for the efficacy outcomes of relapse and positive urine test. However, for the safety outcome (overdose), these models indicate a greater risk of overdose for XR-NTX compared to BUP-NX during the study treatment phase.

ACKNOWLEDGMENTS

The authors thank the NIDA for sponsoring the X:BOT study (CTN-0051). The authors thank the research team who conducted X:BOT and for answering questions regarding the public data files. The authors also thank NIDA for sharing the public data files (https://datashare.nida.nih.gov/study/nida-ctn-0051).

REFERENCES

1. Huecker MR, Leaming JM. The National Center for Biotechnology Information, U.S. National Library of Medicine. Opioid Addiction. Published January 25, 2020. Available at: https://www.ncbi.nlm.nih.gov/books/NBK448203/. Accessed March 9, 2020.
2. Volkow ND, Frieden TR, Hyde PS, et al. Medication-assisted therapies–tackling the opioid-overdose epidemic. *N Engl J Med.* 2014;370(22): 2063–2066.
3. Randall-Kosich O, Andraka-Christou B, Totaram R, et al. Comparing reasons for starting and stopping methadone, buprenorphine, and naltrexone treatment among a sample of white individuals with opioid use disorder. *J Addict Med*. 2020;14(4):e44–e52.
4. Lee JD, Nunes EV, Novo P, et al. Comparative effectiveness of extended-release naltrexone versus buprenorphine-naloxone for opioid relapse prevention (X:BOT): a multicentre, open-label, randomised controlled trial. *Lancet*. 2018;391(10118):309–318.
5. Sobell LC, Sobell MB. *Alcohol Timeline Followback Users' Manual.* Toronto, Canada: Addiction Research Foundation; 1995.
6. Extended-Release Naltrexone vs. Buprenorphine for Opioid Treatment (X:BOT). Statistical Analysis Plan. April 2017. Available at: https://clinicaltrials.gov/ProvidedDocs/33/NCT02032433/SAP_001.pdf. Accessed October 1, 2020.
7. Rotrosen J, Lee J, Nunes E, et al. NIDA CTN Protocol 0051: Extended-Release Naltrexone vs. Buprenorphine for Opioid Treatment (X:BOT). August 2015. Available at: https://clinicaltrials.gov/ProvidedDocs/33/NCT02032433/Prot_000.pdf. Accessed October 1, 2020.
8. Hernán MA, Robins JM. Per-protocol analyses of pragmatic trials. *N Engl J Med*. 2017;377(14):1391–1398.
9. National Institute on Drug Abuse. Data Share Website. Published January 12, 2018. Available at: https://datashare.nida.nih.gov/study/nida-ctn-0051. Accessed January 4, 2020.
10. Ball G, Reblin T, Buchanan J, et al. A framework for safety evaluation throughout the product development life-cycle. *Ther Innov Regul Sci*. 2020;54(4):821–830.
11. Gupta SK. Intention-to-treat concept: a review. *Perspect Clin Res*. 2011;2(3):109–112.
12. Ranganathan P, Pramesh CS, Aggarwal R. Common pitfalls in statistical analysis: intention-to-treat versus per-protocol analysis. *Perspect Clin Res*. 2016;7(3):144–146.
13. Hernán MA, Hernández-Díaz S. Beyond the intention-to-treat in comparative effectiveness research. *Clin Trials*. 2012;9(1):48–55.
14. Morgan JR, Schackman BR, Leff JA, et al. Injectable naltrexone, oral naltrexone, and buprenorphine utilization and discontinuation among individuals treated for opioid use disorder in a United States commercially insured population. *J Subst Abuse Treat*. 2018;85:90–96.
15. Alkermes Inc. Highlights of Prescribing Information: VIVITROL (Naltrexone for Extended-Release Injectable Suspension). Revised. July 2020. Available at: https://www.vivitrol.com/content/pdfs/prescribing-information.pdf. Accessed December 9, 2020.
16. Sullivan MA, Bisaga A, Mariani JJ, et al. Naltrexone treatment for opioid dependence: does its effectiveness depend on testing the blockade? *Drug Alcohol Depend*. 2013;133(1):80–85.