METHODS

Weighted and Unweighted Log-Rank Test

The log-rank test is used to test a null hypothesis that two populations have identical survival functions, or equivalently, identical hazard functions. To describe the test in the context of the study under discussion, let t_1, t_2, ..., be the lengths of the bouts; i.e., $t_1 = 4$, $t_2 = 8$, etc. Let d_{i1} be the number of bouts in the control group, and let d_i be the number of bouts in both groups combined whose duration is equal to t_i. Let Y_{i1} be the number of bouts in the control group, and let Y_i be the number of bouts in both groups combined whose duration is greater than or equal to t_i.

When the null hypothesis is true, we would expect that the empirical hazard d_{i1}/Y_{i1} in the control group will generally be similar to the empirical hazard d_i/Y_i in both groups combined.

The log-rank test is based on the differences $d_{i1}/Y_{i1} - d_i/Y_i$, or, equivalently upon multiplying by Y_{i1}, the differences $d_{i1} - Y_{i1}\, d_i/Y_i$. The log-rank test statistic is formed by dividing the sum of these differences by its standard error. The test statistic is

$$Z = \frac{\sum_{i=1}^{D}\left[d_{i1} - Y_{i1}\left(\frac{d_i}{Y_i} \right) \right]}{\sqrt{\sum_{i=1}^{D}\left(\frac{Y_{i1}}{Y_i}\left(1 - \frac{Y_{i1}}{Y_i} \right)\left(\frac{Y_i - d_i}{Y_i - 1} \right) d_i \right)}}$$

where the sums are over all times t_i.

The test statistic has an approximately standard normal distribution for large samples when the null hypothesis is true, so the null hypothesis is rejected when the value of the test statistic is large in absolute value (e.g., > 1.96).

The log-rank test has good power to detect alternatives in which failure rates are consistently higher in one group over time. In these cases the differences in the numerator of the test statistic all tend to have the same sign. However, for alternatives in which failure rates are higher in one group for early times but higher in the other group for later times (i.e., when the survival curves cross), positive and negative differences tend to cancel out, and power is low. Against these alternatives we used a weighted logt rank test (Klein and Moeshberger, Survival Analysis, Springer, 2003). Test weights $W(t_i)$, are defined for each t_i, and the test statistic is

$$Z = \frac{\sum_{i=1}^{D} W(t_i)\left[d_{i1} - Y_{i1}\left(\frac{d_i}{Y_i}\right)\right]}{\sqrt{\sum_{i=1}^{D} W(t_i)^2 \frac{Y_{i1}}{Y_i}\left(1-\frac{Y_{i1}}{Y_i}\right)\left(\frac{Y_i - d_i}{Y_i - 1}\right) d_i}}$$

We used weights $W(t_i) = S(t_i)$ and $W(t_i) = 1 - S(t_i)$, where $\hat{S}(t) = \prod_{t_j \le t}[1 - d_i / Y_i]$ is the Kaplan-Meier (Product-Limit) estimator of the survival function. The weights $W(t_i) = S(t_i)$ put more emphasis on differences that occur early, because $S(t_i)$ decreases with time. Similarly, The weights $W(t_i) = 1 - S(t_i)$ put more emphasis on differences that occur late.

Empirical Hazards

To quantify the differences in hazard rates for each survival curve, we computed the empirical hazard, *h(t)*, for each time *t* as follows:

$$h(t) = \frac{\text{number of bouts } = t}{\text{number of bouts} \ge t}$$

As values of *t* become large, the hazard function can vary widely because the denominators are smaller and values of 0 are sometimes found in the numerator. To construct a plot that would more clearly illustrate the differences in hazard functions associated with different stages of orexin neuron degeneration, we defined a smoothed empirical hazard, *h*(t)*, for each *t* where *h*(t)* is the average of the hazards for all times within 5% of *t*. Since bouts were scored in 4 s epochs, there was no smoothing for bouts < 80s. To further avoid spurious oscillations that occur at very large values of *t* when data are sparse, we restricted the computation of *h*(t)* to values of *t* for which at least 50 bouts were at risk of failure. We tested for statistically significant differences between the hazards with the z-test.

Table S1: In both the dark and light periods, differential effects of orexin on "short" and "long" wake bouts persisted independent of the threshold chosen to differentiate short and long bouts.

Threshold (sec)	Short, Dark (P value)	Short, Light (P value)	Long, Dark (P value)	Long, Light (P value)
60	4.39E-02	5.40E-01 (ns)	0	1.11E-13
100	1.57E-10	2.46E-06	0	0
200	0	0	0	0
300	0	0	0	2.22E-16
400	0	0	0	2.59E-12
500	0	0	0	3.91E-09
600	0	0	0	4.49E-07
700	0	0	0	6.36E-05
800	0	0	0	5.12E-05
900	0	0	0	4.51E-03
1000	0	0	0	8.16E-02 (ns)

Using the log-rank test, we separately compared the survival of "short" and "long" wake bouts across weeks DOX(-) to assess bout-duration dependence of orexin effects. To avoid any bias related to the choice of threshold dividing "short" and "long" bouts, we performed this analysis with thresholds ranging from 60 sec to 1000 sec. Dark and light period data were analyzed separately. The P values for each log-rank test are reported, and $P < 0.05$ was considered to denote a significant difference in survival of the indicated subset of wake bouts with weeks DOX(-). All $P < 2.22E\text{-}16$ are reported as $P = 0$ and represent significant differences between subsets. All subsets of wake bouts differed significantly with weeks DOX(-) except light period wake bouts ≤ 60 sec ($P = 0.54$) and light period wake bouts > 1000 sec ($P = 0.0816$). The shaded row in the table reports P values for subsets of data corresponding to a threshold of 200 sec (i.e., bouts of duration ≤ 200 sec are considered "short" and bouts of duration > 200 sec are considered "long"), and the corresponding survival curves are presented in Figure 4 (dark period data: Figure 4B, 4C; light period data: Figure 4E, 4F).

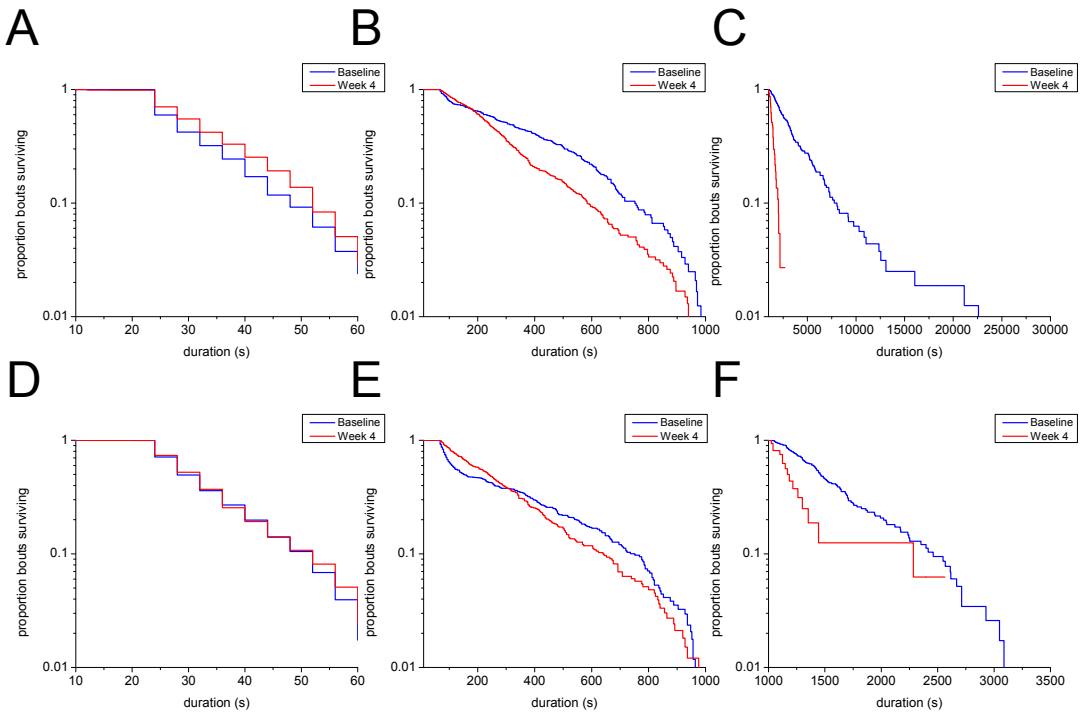

Figure S1. Kaplan-Meier survival curves for subsets of short, intermediate, and long wake bouts show bout-duration-dependent differences in changes in the weeks after DOX discontinuation. (**A**) The survival of short (≤ 60 sec) wake bouts increases at week 4 DOX(-) in the dark period (log-rank test, $P < 0.05$). By contrast, the survival of wake bouts of intermediate duration (durations between 64 and 1000 sec) is reduced at week 4 DOX(-) (log-rank test, $P < 0.01$) (**B**); and survival of long wake bouts (>1000 sec), survival is severely reduced (log-rank test, $P < 0.01$) (**C**). In the light period, there is no difference in the survival of bouts of short (log-rank test, $P < 0.05$) (**D**) or intermediate durations (log-rank test, $P < 0.01$) (**E**), but the survival of long wake bouts is decreased (log-rank test, $P < 0.05$) (**F**).
DOX = doxycycline.

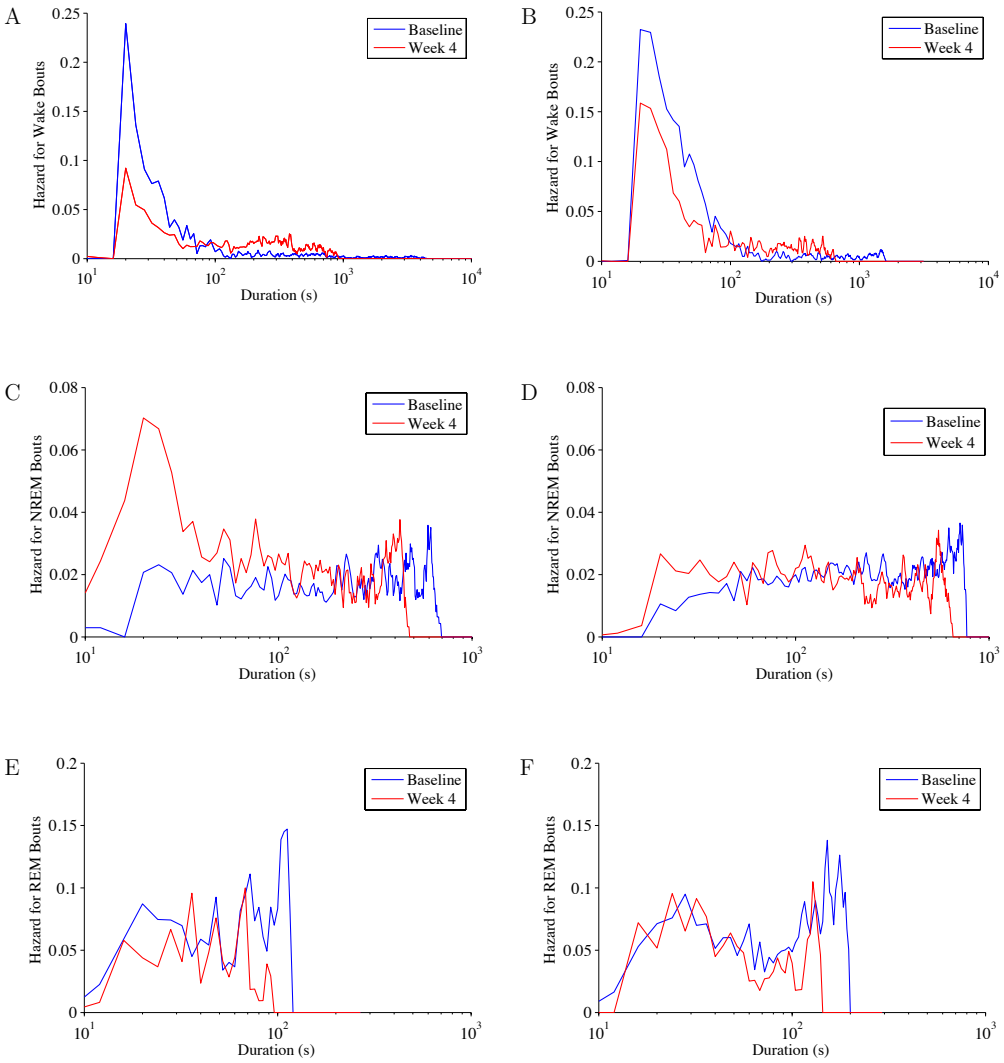

Figure S2. The hazards computed for wake, NREM sleep, and REM sleep survival curves change between baseline and week 4 DOX(-). Hazards for wake survival curves reveal the bout-duration-dependence of the change that causes survival curves for baseline and week 4 to intersect: for brief wake bouts there is a decreased hazard at week 4 compared to baseline in both the dark (**A**) and light (**B**) periods (z-test, $P < 0.05$), but there is a trend to reverse this relationship for longer wake bouts. In both the dark (**C**) and light (**D**) periods, the hazard for shorter NREM bouts is greater at week 4 DOX(-) than the hazard at baseline (z-test, $P < 0.05$). Except for hazards associated with bouts <~30 s in the dark period, the hazards for NREM bouts are relatively constant in both dark (E) and light (**F**) periods. For longer REM bouts, the hazard at week 4 DOX(-) is less than the baseline hazard (z-test, $P < 0.05$). DOX = doxycycline; NREM = nonrapid eye movement; REM = rapid eye movement.

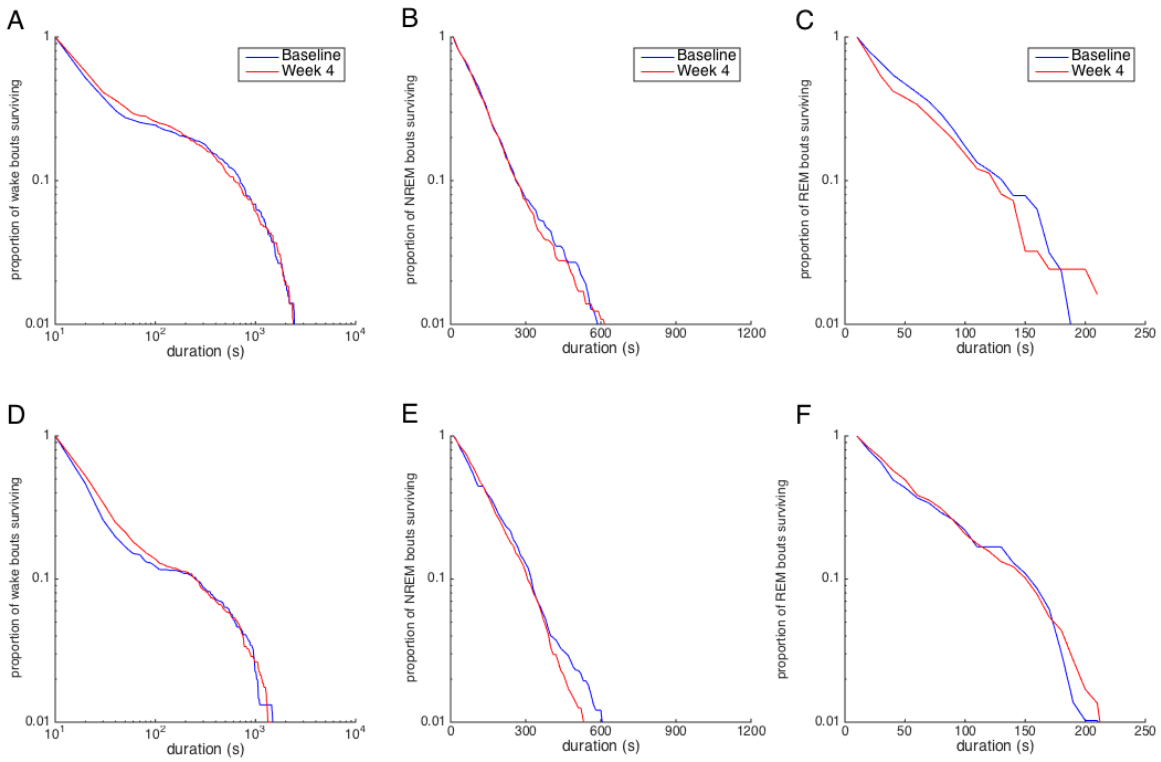

Figure S3. When DTA mice were maintained on DOX(+) chow throughout the four week recording period, Kaplan-Meier survival curves for wake (**A, D**), NREM sleep (**B, E**), and REM sleep (**C, F**) show no changes between baseline and Week 4 conditions in the dark (**A, B, C**) or light periods (**D, E, F**), respectively (n = 5). DTA = orexin-tTA; TetO diphtheria toxin A; DOX = doxycycline; NREM = nonrapid eye movement; REM = rapid eye movement.