A mathematical model describes Drosophila sleep behavior in w1118 controls and in a hyposomnolent *insomniac* line

Joshua M Diamond

The conserved nature of sleep in Drosophila has allowed the fruit fly to emerge in the last decade as a powerful model organism in which to study sleep.

Recent sleep studies in Drosophila have focused on the discovery and characterization of hyposomnolent mutants. One common feature of these animals is a change in sleep architecture: sleep bout count tends to be greater, and sleep bout length lower, in hyposomnolent mutants.

I propose a mathematical model, produced by least-squares nonlinear regression to fit the form Y = aX^b, which can explain sleep behavior in the healthy animal as well as previously-reported changes in total sleep and sleep architecture in hyposomnolent mutants.

This model, fit to sleep data, yields coefficient of determination R squared, which describes goodness of fit. R squared is lower in hyposomnolent mutant *insomniac* as compared to control, indicating a poorer fit of the model to the data in *insomniac*. R squared also tends to be lower in daytime sleep as compared to nighttime sleep.

My findings raise the possibility that low R squared is a feature of all hyposomnolent mutants, not just *insomniac*. If this were the case, R squared could emerge as a novel means by which sleep researchers might assess sleep dysfunction.

1 A mathematical model describes Drosophila sleep behavior in w1118
2 controls and in a hyposomnolent *insomniac* line

3 Running title: A Model for Drosophila Sleep

4 **Joshua Diamond[1]***

5 [1]Dickman Lab, University of Southern California, Department of Neuroscience, Los Angeles,
6 CA, USA

7 *** Correspondence:** Joshua Diamond, University of Virginia School of Medicine. 852 W. Main
8 St., Apt. 803c, Charlottesville, VA, 22903, USA
9 jmd2cr@virginia.edu

10 Article contains 3200 words and 11 references.

11 **Conflict of interest statement**

12 My research was conducted in the absence of any commercial or financial relationships that
13 could be construed as a potential conflict of interest.

14

15 **Summary**

16 The conserved nature of sleep in Drosophila has allowed the fruit fly to emerge in the last decade
17 as a powerful model organism in which to study sleep.

18 Recent sleep studies in Drosophila have focused on the discovery and characterization of
19 hyposomnolent mutants. One common feature of these animals is a change in sleep architecture:
20 sleep bout count tends to be greater, and sleep bout length lower, in hyposomnolent mutants.

21 I propose a mathematical model, produced by least-squares nonlinear regression to fit the form Y
22 $= aX^b$, which can explain sleep behavior in the healthy animal as well as previously-reported
23 changes in total sleep and sleep architecture in hyposomnolent mutants.

24 This model, fit to sleep data, yields coefficient of determination R squared, which describes
25 goodness of fit. R squared is lower in hyposomnolent mutant *insomniac* as compared to control,
26 indicating a poorer fit of the model to the data in *insomniac*. R squared also tends to be lower in
27 daytime sleep as compared to nighttime sleep.

28 My findings raise the possibility that low R squared is a feature of all hyposomnolent mutants,
29 not just *insomniac*. If this were the case, R squared could emerge as a novel means by which
30 sleep researchers might assess sleep dysfunction.

31 **Keywords: Drosophila, sleep, activity, waking, insomniac, homeostasis, architecture,**
32 **consolidation, least-squares, nonlinear regression**

33

1.   Introduction

Sleep in Drosophila exhibits many characteristics that are seen also in mammalian sleep, including extended periods of quiescence, reduced arousal threshold, and hyper-consolidation of sleep after sleep deprivation [1]. Further, drugs that alter mammalian sleep have analogous effects in Drosophila, suggesting conserved neural and biochemical mechanisms [2]. The conserved nature of sleep in Drosophila has allowed the animal to emerge in the last decade as a powerful model for the study of sleep.

Much recent work in Drosophila has been focused on the discovery and study of certain hyposomnolent mutants [3] [4] [5] [6] [7].

In addition to reduced total sleep, hyposomnolent mutants also demonstrate altered sleep architecture. Sleep is poorly consolidated: bout length is reduced as compared to control [3] [5] [7] [8]. In some of these cases bout count is also reduced [5], but more frequently it is elevated [3] [7] [8]. One such example is *Insomniac* [3] [8], which is the basis of much of the modeling work in this study.

The goal of this study is to produce a mathematical model that describes sleep behavior in control animals. Further, I will examine the extent to which this model also holds true in the hyposomnolent mutant *insomniac*.

My results may establish a new paradigm for analysis of sleep dysfunction in hyposomnolent mutants. These techniques could also be used on higher animals, including humans.

2.   Methods

All animals came from the Bloomington Stock Center at Indiana University. *Insomniac* corresponds to stock number 18307. w1118 was used as control.

Insomniac was outcrossed for 8 generations to an isogenic w1118 line to control for genetic background. Only males were used in this experiment, for mutants and controls. Animals were 1-5 days old.

Sleep was monitored using TriKinetics' DAM2 Drosophila Activity Monitors, as previously described [9]. Briefly, animals were placed inside activity tubes containing food made of 5% sucrose and 2% agarose and then housed in an incubator with 12-hour:12-hour day:night cycles at 25 degrees C and 85% humidity. Animals were given three days to acclimate to the day/night cycle before data collection began. After the acclimation period, data collection lasted 4 full 24-hour periods. Sleep is defined as 5 minutes of inactivity [10]. Animals that died or showed significant loss of health during the course of the experiment were automatically excluded from the results. Data was processed using SleepLab, custom Matlab-based software provided by Dr. William Joiner (UCSD).

All daytime data have been separated from nighttime data, but otherwise all data have been combined together over four days for each genotype. Statistical analysis was handled with GraphPad PRISM 6.

71 In analysis of sleep behavior, ordinary (unweighted) least-squares nonlinear regression is used to
72 produce lines of fit, constrained to the equation $Y = aX^b$. In each line of fit, the independent
73 variable X represents the sleep bout count of the animal-time period pair, while the observed
74 response variable Y represents the mean sleep bout length in that same animal-time period pair.

75 Similar lines of fit are produced using activity bout data. In activity bout data analysis,
76 independent variable X represents activity bout count, and observed response variable Y
77 represents mean activity bout length.

78 Nonlinear regression assumes that the pool of residuals is drawn from a Gaussian distribution.
79 The D'Agostino & Pearson omnibus K2 test is used to test for attainment of this requirement.

80 R^2 is computed based on the data's adherence to the same ordinary least-squares nonlinear
81 regression line discussed previously. Specifically, $R^2 = 1 - (SSres/SStot)$, where SSreg is the
82 sum of the squares of all distances along the y-axis between data points and the best-fit curve,
83 and SStot equals the sum of square of all distances along the y-axis between data points and the
84 horizontal line that runs through the mean of all y-values.

85 **3.   Results**

86 n = 31 per genotype and 64 total. Since each animal slept for four 24-hour periods, including
87 four days and four nights, we consider sets of 124 observations.

88 **3.1   Characterization of wild type sleep**

89 Sleep in wild type animals is consistent with that seen in the literature, in terms of both total time
90 slept and sleep architecture [3] [8].

91 **3.2   Characterization of *insomniac* sleep**

92 *Insomniac* demonstrates a robust phenotype in terms of total time slept. *Insomniac* animals tested
93 in this experiment sleep significantly less than controls in the 24-hour period. According to a
94 two-tailed, two-sample heteroscedastic (allowing for unequal variance) Student's T-test,
95 probability that measures of total sleep per 24 hours in *insomniac* and controls came from the
96 same distribution is given by $p < 0.0001$.

97 *Insomniac* also demonstrates a strong phenotype in sleep architecture. Bout length is shorter and
98 bout count greater in *insomniac* as compared to its control. According to the same Student's t-
99 test performed above, $p < 0.0001$ for both mean sleep bout length per 24 hours and bout count
100 per 24 hours.

101 Sleep in *insomniac* is compared to control in **Fig. 1.** Decreased total sleep is seen in the
102 nighttime only in *insomniac* as compared to control. Decreased mean sleep bout length and
103 increased mean sleep bout count is seen in both daytime and nighttime in *insomniac* as compared
104 to control.

105 **Figure 1 here**

106 That sleep in *insomniac* is poorly consolidated can be observed qualitatively. **Fig. 2** represents
107 activity in *insomniac* and control. We see that, in the case of *insomniac*, activity is distributed
108 throughout periods in which control flies normally sleep.

109 **Figure 2 here**

110 **3.3 Production of a mathematical model**

111 Past work has generally studied *average* total sleep per 24 hours, taken over all animals in a
112 given experimental condition and over the entire duration of the experiment. Instead, I consider
113 daytime and nighttime sleep separately, and I consider the sleep behavior of individual animals
114 during single days or nights. By considering the sum of individual animal-day and animal-night
115 pairs, I can produce from these data a mathematical model that describes sleep behavior.

116 **Fig. 3** shows the individual animal data.

117 Sleep behavior is most regular in the case of control night (**Fig. 3c**). To this set of data, I fit the
118 model

119 $$y = a \cdot x^b \qquad \qquad \textbf{Equation 1}$$

120 where *y* corresponds to mean sleep bout length, for an individual animal, over the course of a
121 single night; and *x* corresponds to sleep bout count for that same individual animal over the
122 course of a single night. Coefficient of determination R^2 is 0.993 in the case of control night. In
123 **Fig. 3**, **Eq. 1** is fit to all experimental conditions. R^2 is not as high in other experimental
124 conditions as it is in control night, indicating a worse fit to the model in these other experimental
125 conditions.

126 **Figure 3 here**

127 **Eq. 1**, the parameters that comprise it, and the R^2 coefficient might provide valuable insight
128 towards the analysis of sleep behavior in Drosophila, even in experimental conditions where R^2
129 is relatively low.

130 The parameter *b* is negative in experimental conditions. This indicates that, as bout count rises,
131 mean sleep bout length falls. Further, *b* tends to reside near −1.

132 In **Eq. 1**, *a* tends to estimate total sleep. For example, in **Fig. 3c**, *a* = 682.9. Consistent with this
133 prediction, measured sleep for this genotype and timeframe is 673 minutes. Given the form of
134 **Eq. 1**, that *a* estimates total sleep should make sense. Suppose, in the regression **Eq. 1**, it so
135 happens that *b* = −1 exactly. Then we can re-express the equation as

136 $$y \cdot x = a \qquad \qquad \textbf{Equation 2}$$

137 **Eq. 2** shows that (in the case *b* = −1) the best regression generates a fixed constant *a* with the
138 special property that the product of any pair of values attained by the variables *x* and *y* tends to
139 fall close to *a*. These values in turn correspond to the bout count and mean bout lengths,
140 respectively, of the animals. And, we know that, in an individual animal-time period pair, mean
141 bout length times bout count equals total sleep for that time period. Thus we see why, when *b*
142 falls close to −1, *a* estimates total sleep.

143 As b deviates from -1, a becomes a worse estimate of mean total sleep. For example, in **Fig. 3b**,
144 $a = 68$. This drastically underestimates total sleep for this genotype and timeframe. For $b > -1$, a
145 is an underestimate of mean total sleep. For $b < -1$, a is an overestimate. The tendency of a to
146 estimate total sleep, as well as the relationship between b and a I have just described, holds in
147 both control animals and in *insomniac*. In *insomniac*, a may not be as good an estimate of total
148 sleep, in part because b may stray further from -1.

149 The coefficient of determination R^2 may be of use. As described earlier, R^2 is greatest in the
150 setting of control sleep behavior at night. R^2 close to 1 indicates that the mathematical model
151 closely fits the data.

152 R^2 is closer to 1 in the nighttime, as compared to the daytime, with genotype controlled for. In
153 other words, control night has greater R^2 than control day; meanwhile, *insomniac* night has
154 greater R^2 than *insomniac* day. Additionally, R^2 is farther from 1 in *insomniac*, as compared to
155 control, with time of day controlled for. *Insomniac* night has lower R^2 than control night;
156 *insomniac* day has lower R^2 than control day.

157 So, in the daytime, and in *insomniac*, the model tends to fit the data less well.

158 Under conditions where R^2 is relatively low, such as *insomniac day*, 95% confidence intervals
159 for parameters a and b tend to be wider relative to the absolute value of these parameters. Also,
160 95% confidence bands tend to be wider as well in conditions with low R^2.

161 **3.4   Application of the model to activity data**

162 I conducted a similar statistical analysis on the behavior of the animals used in this experiment,
163 except considering activity bouts as opposed to sleep bouts.

164 **Eq. 1** does not fit the activity bout data as well as it fits the sleep bout data. R^2 is 0.608 at
165 maximum.

166 Like in the case of the sleep bout data, R^2 is higher in control than it is in *insomniac*. R^2 values
167 are 0.608 and 0.582 in control, nighttime and daytime, respectively, compared to 0.270 and
168 0.299 in *insomniac*.

169 Note that, in contrast with the sleep bout data, it is not the case in the activity bout data that R^2
170 changes in daytime as compared to nighttime. Within a given genotype, daytime and nighttime
171 R^2 values are nearly identical.

172 **3.5   Statistical tests for the appropriateness of the mathematical model**

173 Dependency between parameters a and b as they fit to the sleep data ranges from 0.822 to 0.984.

174 The sleep data do not pass the D'Agostino & Pearson omnibus K2 test of normalcy test in any
175 genotype or timeframe, including control day, control night, *insomniac* day, and *insomniac* night.

176 Dependency and normalcy in the activity data is not reported, because I do not make the claim
177 that **Eq. 1** is an appropriate model for the activity data.

178 **4.   Discussion**

4.1 Evaluation of sleep behavior

Sleep behavior in the control is similar to that seen in past work. Sleep behavior in control is normal quantitatively (**Fig. 1**) and qualitatively (**Fig. 2**). This indicates that my sleep system is in good working order. Further, the sleep phenotype I have demonstrated in *insomniac* mutants, which is characterized by reduced total sleep and poor consolidation, is consistent with past reports [3] [8].

4.2 The value of consideration of daytime and nighttime sleep

Past work has usually reported only on total sleep and sleep architecture during the 24-hour period. Splitting sleep behavior into daytime and nighttime, as I have, is useful. Consideration of daytime and nighttime sleep provides detail which may be missed if only 24-hour sleep is considered. For example, as seen in **Fig. 1a**, only total nighttime sleep is significantly different in *insomniac* as compared to control; total daytime sleep is not. This is missed when, as has been done in previous work, only 24-hour sleep values are compared.

Perhaps more importantly, splitting daytime and nighttime sleep allows the mathematical model to be fit to each group separately. Sleep behavior is different in daytime versus nighttime, and so combining these two pools of data would reduce signal. Separate consideration of daytime and nighttime allows more information, and more precise information, to be drawn from these data.

4.3 Merits of the mathematical model

Coefficient of determination R^2, which measures goodness of fit to the mathematical model described in **Eq. 1**, is as high as 0.993. This serves to validate the mathematical model: at least in some circumstances, the model describes behavior very well. Even in conditions where R^2 is not as high, the model appears to describe the behavior reasonably well considering the higher degree of variability within those data. Future study should examine whether **Eq. 1** describes sleep behavior in other genotypes besides w1118 and *insomniac*.

As mentioned in the results, R^2 is lower in *insomniac*. Future research might investigate whether or not R^2 is also diminished in other hyposomnolent mutants. If this were the case, R^2 could emerge as a novel means by which sleep dysregulation might be measured. A lower R^2 could indicate a greater degree of dysregulation. R^2 could then be used to assess sleep dysfunction in other Drosophila lines, and in other animals, including humans.

Note that R^2 constitutes a measure of sleep behavior independent of those measures usually studied in Drosophila sleep research, namely, total sleep, mean bout length, and mean bout count. Any of these measures could be changed in a Drosophila line, without change in R^2. Likewise, R^2 could theoretically change without corresponding change in total sleep, mean bout length, or mean bout count. Thus the R^2 measure offers novelty.

4.4 Other measures provided by the mathematical model

As mentioned in the results, where R^2 is relatively low, the 95% confidence intervals for parameters *a* and *b* tends to be wide relative to the absolute value of these parameters. Further research should investigate whether or not the 95% confidence intervals for parameters *a* and *b*

217 are also wider than control in other hyposomnolent mutants besides *insomniac*. **Eq. 1** parameter
218 confidence interval width, like R^2, could serve as a novel measure of sleep dysregulation.

219 In situations with low R^2, confidence bands also tend to be wider. Width of the confidence band
220 could also be evaluated as a potential measure of sleep dysregulation.

4.5   Application of the model to the activity data

222 Fit of **Eq. 1** to the activity data tends to be poorer than fit of **Eq. 1** to the sleep data. This might
223 suggest that total sleep is more tightly regulated during sleep than total activity is during waking.

224 Whatever the reason, it seems that **Eq. 1** may not be as appropriate a model for waking behavior
225 as it is for sleep behavior.

4.6   Limitations of the model as applied to the sleep data

227 Dependency between parameters can be as high as 0.984, which indicates that *a* and *b* may be
228 redundant. If a simpler model is desired, **Eq. 2** would suffice. However, the inclusion of *b* seems
229 to be merited, because production of a model conforming to **Eq. 1** is not difficult, and *b* still
230 improves goodness of fit.

231 That the sleep data universally fail the D'Agostino & Pearson omnibus K2 test under all
232 circumstances might be cause for concern. Regardless of this finding, though, my model still
233 appears to have merit, discussed in 4.3. Further, failure of this test need not indicate that
234 nonlinear regression is an inappropriate strategy. Especially in large data sets, deviations from
235 normalcy may reach statistical significance without corresponding to real practical meaning [11].
236 So, it appears that my least-squares nonlinear regression procedure may be resistant to violations
237 of the standard that underlying distributions be Gaussian [11]. Nevertheless, future work could
238 look at the use of robust nonlinear regression models, as opposed to the least-squares nonlinear
239 model used here. These are less distorted by data sets whose residuals come from non-Gaussian
240 distributions [11].

241 Note also that, if mean sleep bout length values are weighted by $1/y^2$, performance on the
242 D'Agostino & Pearson omnibus K2 normalcy test is improved but still poor.

5.   Acknowledgments

248
## 6.   References
250
251   1. Cirelli, C. and Bushey, D. Sleep and wakefulness in Drosophila melanogaster. *Ann. N. Y.*
252   *Acad. Sci.*, 2008: 1417.
253   2. Hendricks, J., Finn, S., Panckeri, K., et. al. Rest in Drosophila is a sleep-like state.
254   *Neuron*, 2000: 129-138.

255   3. Stavropoulos, N. and Young, M. W. Insomniac and Cullin-3 Regulate Sleep and
256   Wakefulness in Drosophila. *Neuron*, 2011: 964–976.
257   4. Kume, K., Kume, S., Park, S.K., et. al. Dopamine Is a Regulator of Arousal in the Fruit
258   Fly. *The J. Neurosci.*, 2005: 7377-7384.
259   5. Joiner, W., Koh, K., Wu, M., et. al. Identification of SLEEPLESS, a sleep-promoting
260   factor. *Science*, 2008: 372-376.
261   6. Hendricks, J. C., Willians, J. A., Panckeri, K., et. al. A non-circadian role for cAMP
262   signaling and CREB activity in Drosophila rest homeostasis. *Nat. Neurosci.*, 2001, 4:
263   1108-1115.
264   7. Foltenyi, K., Greenspan, R. J., and Newport, J. W. Activation of EGFR and ERK by
265   rhomboid signaling regulates the consolidation and maintenance of sleep in Drosophila.
266   *Nat. Neurosci.,* 2007, 10: 1160-1167.
267   8. Pfeiffenberger, C. and Allada, R. Cul3 and the BTB Adaptor Insomniac Are Key
268   Regulators of Sleep Homeostasis and a Dopamine Arousal Pathway in Drosophila. *PLOS*
269   *Genet.*, 2012.
270   9. Pfeiffenberger, C., Lear, B., Keegan, K., et. al. Locomotor activity level monitoring using
271   the Drosophila Activity Monitoring (DAM) System. *Cold Spring Harbor Protocol*, 2010.
272   10. Pfeiffenberger, C., Lear, B., Keegan, K., et. al. Processing sleep data created with the
273   Drosophila Activity Monitoring (DAM) System. *Cold Spring Harbor Protocol*, 2010.
274   11. D'Agostino, R. B. Tests for Normal Distribution. In *Goodness-Of-Fit Techniques*. Macel
275   Decker, 1986.

1

Characterization of sleep in *insomniac* versus control

Error bars represent the standard error measurement. Each experiment consists of 31 animals, all of which survived for four days. So, each genotype-time period pair represents an average across $n = 124$ measurements. (A) Total sleep in insomniac versus control. Values shown represent mean total sleep, across the four days of the experiment. I have distinguished between insomniac and control, and within these distinctions, I have distinguished again between daytime and nighttime. (B) Sleep bout length in insomniac versus control. Values represent averages across the length of the experiment. (C) Mean number of sleep bouts across the length of the experiment. *p < 0.0001 according to two-tailed, two-sample heteroscedastic Student's T-test.
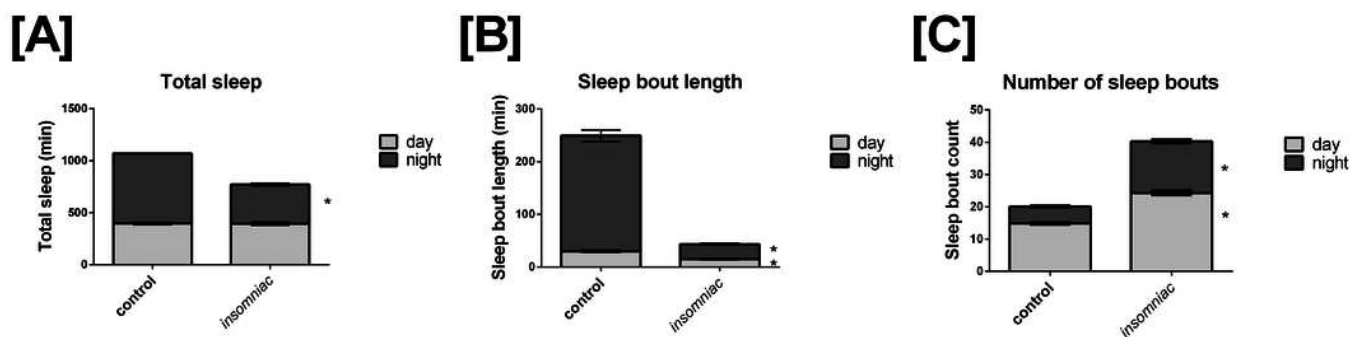
2

Representative actograms for control and *insomniac*

(A) control. (B) insomniac. Each panel represents the sleep/wake activity of a single animal. So, three animals are shown for each genotype, and six are represented in total. Note disorganized sleep/wake behavior in *insomniac*, including extensive activity during lights-off 12-hour periods. Day one (not shown) as well as days two and three (shown in **Fig. 2**) were not considered in data analysis.
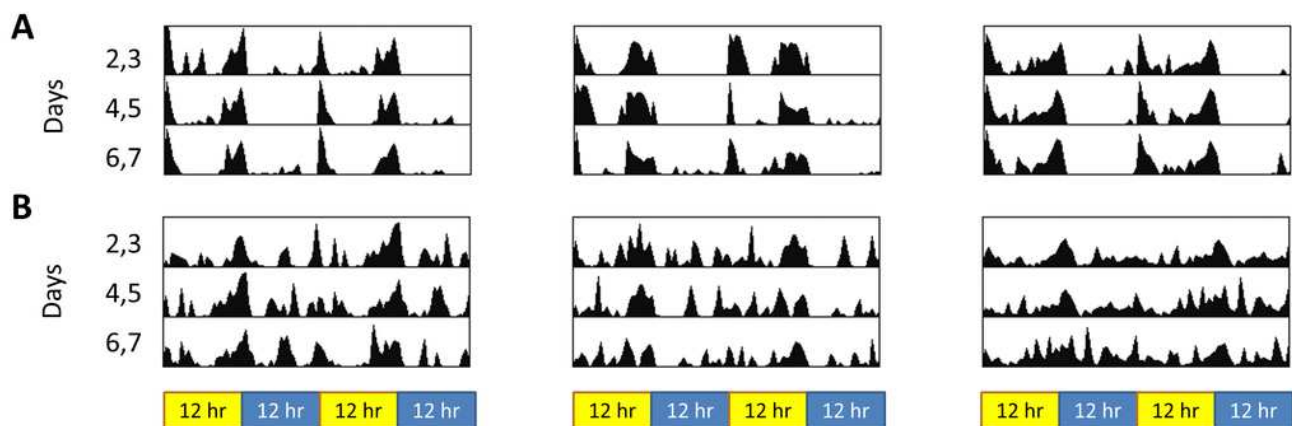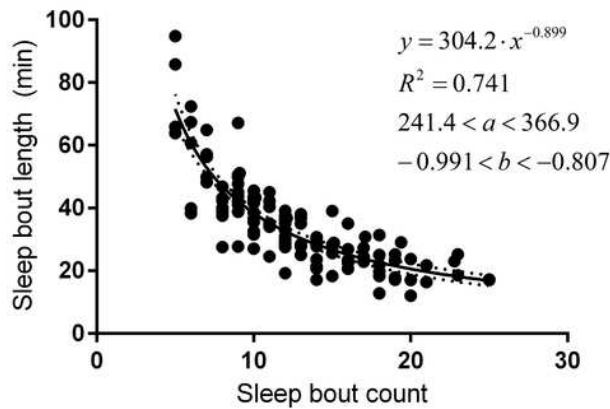
3

Relationship between bout count mean bout length in individual animals

Each dot represents a single animal-day pair (panels **A** and **B**) or animal-night pair (panels **C** and **D**). The y axis represents the mean number of bouts slept during each animal-time period pair, and the x axis represents the amount of bouts slept in that same animal-time period pair. Thus n for each figure is equal to 31 * 4 = 124 animal-time period pairs. Each panel contains an inset, which lists, from to bottom: the equation of the line of fit, in the format Y = aX^b; the coefficient of determination R^2; the 95% confidence interval for the a parameter; and the 95% confidence interval for the b parameter. Dotted lines represent the upper and lower margins of the 95% confidence band. The chances are 95% that the true line of fit lies between these upper and lower margins.
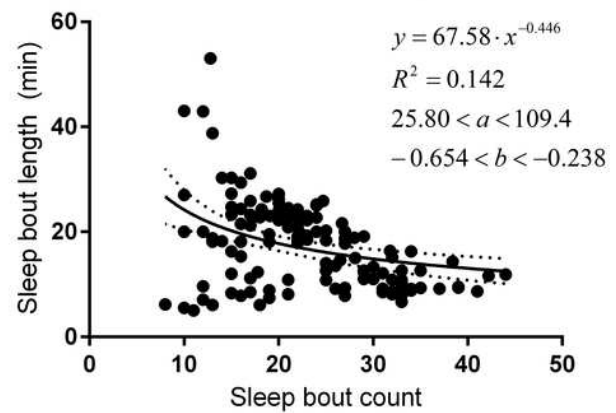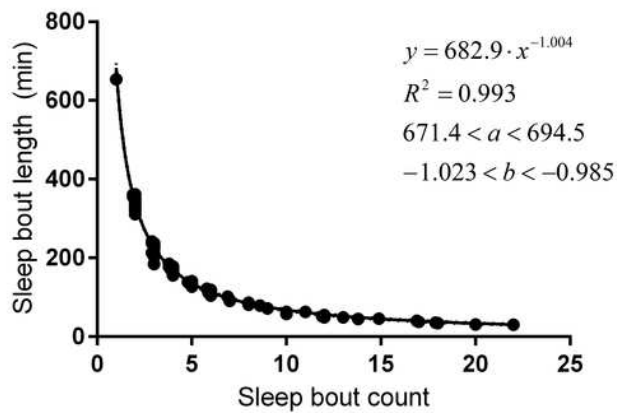
[A]

Control day

$y = 304.2 \cdot x^{-0.899}$
$R^2 = 0.741$
$241.4 < a < 366.9$
$-0.991 < b < -0.807$

[B]

Insomniac day

$y = 67.58 \cdot x^{-0.446}$
$R^2 = 0.142$
$25.80 < a < 109.4$
$-0.654 < b < -0.238$

[C]

Control night

$y = 682.9 \cdot x^{-1.004}$
$R^2 = 0.993$
$671.4 < a < 694.5$
$-1.023 < b < -0.985$

[D]

Insomniac night

$y = 415.9 \cdot x^{-1.011}$
$R^2 = 0.604$
$279.9 < a < 551.9$
$-1.115 < b < -0.867$