

Data 1

This document provides an outline with reproducible R (R Core Team 2023) code to explicitly demonstrate the problems we identified with the data and statistical procedures used by Cassidy et al. (2023). The data used by Cassidy et al. (2023) are available via the open-access policy of Frontiers in Ecology and the Environment and must be downloaded prior to running the code (<https://doi.org/10.5061/dryad.mkkwh713f>).

1 Setup

```
library(tidyverse)
library(knitr)
library(kableExtra)
library(broom.mixed)
library(lme4) # How we are going to implement mixed model

### The NPS data
cassidy.data <- read.csv(file.choose()) # choose 'NPS_wolfmortalitydata.csv'
### Fix the first column name
names(cassidy.data)[1]<-'park'
### Change 'NULL' to NA and force the response and covariates to be numeric
cassidy.data[cassidy.data=="NULL"] <- NA
cassidy.data$persisty1 <- as.numeric(cassidy.data$persisty1)
cassidy.data$mortLEAD <- as.numeric(cassidy.data$mortLEAD)
cassidy.data$mortNONLEAD <- as.numeric(cassidy.data$mortNONLEAD)
cassidy.data$packsizeAug <- as.numeric(cassidy.data$packsizeAug)
### One record for reprody1 is blank (DENA 2017 Bearpaw) so assign that value to be NA
# table(cassidy.data$reprody1)
cassidy.data[cassidy.data$park == 'DENA' & cassidy.data$biolyr==2017&cassidy.data$pack==
  'Bearpaw','reprody1'] <- NA
cassidy.data$reprody1 <- as.numeric(cassidy.data$reprody1)
```

2 Data Issues

2.1 Small number of events

Once the observations with missing covariate values are tabulated (i.e., the dataset used for the models) sample sizes drop to concerning levels.

```
### The data their most complex persistence model uses
cassidy.data.persist <- na.omit(cassidy.data[,c('park','biolyr','pack','mortALL','mortLEAD',
  'persisty1','packsizeAug','mortYN')])

### The data their most complex reproduction model uses
cassidy.data.repro <- na.omit(cassidy.data[,c('park','biolyr','pack','mortALL','mortLEAD',
  'reprody1','packsizeAug','mortYN')])
```

2.1.1 Persistence

For the persistence models, there are only 35 total “events” (i.e., occurrences of pack dissolution following human-caused mortality) (Table 1). When modeling, these observations are then spread across the cells of the regression leaving very small numbers of observations for some covariate combinations.

```
persist.total <- cassidy.data.persist %>% group_by(mortYN, persisty1) %>%  
  summarise( total = n()) %>% arrange(persisty1)  
persist.total
```

Table 1: Tabulation of observations by whether there was human-caused mortality and whether the pack persisted showing there are 35 instances of pack dissolution following human-caused mortality.

mortYN	persisty1	total
0	0	54
1	0	35
0	1	591
1	1	111

The data from Voyageurs are nearly absent from persistence analyses for all models containing the pack size and leader loss covariates. In the dataset from the most complex model, there are $n = 8$ data points from Voyageurs and in each case the pack persisted (Table 2).

```
persist.park <- cassidy.data.persist %>% group_by(park, persisty1) %>%  
  summarise( total = n()) %>% arrange(park)  
persist.park
```

Table 2: Tabulation of observations by park and whether the pack persisted showing there are 8 data points from VNP.

park	persisty1	total
DENA	0	41
DENA	1	291
GNTF	0	5
GNTF	1	64
VNP	1	8
YNP	0	26
YNP	1	213
YUCH	0	17
YUCH	1	126

There is also near complete separation in some covariates. All occurrences of more than 4 wolves lost occurred in YUCH (Table 3) and eight of the twelve cases with loss of 2 or more leaders were from YUCH (Table 4) (most likely due to predator control).

```
persist.loss.park <- cassidy.data.persist %>% group_by(park, mortALL) %>%  
  summarise( total = n()) %>% arrange(mortALL)  
persist.loss.park
```

Table 3: Tabulation of observations by park and a count of all human-caused mortality showing that all occurrences of 4+ wolves lost are in YUCH.

park	mortALL	total
DENA	0	281
GNTTP	0	52
VNP	0	8
YNP	0	191
YUCH	0	113
DENA	1	43
GNTTP	1	10
YNP	1	32
YUCH	1	13
DENA	2	7
GNTTP	2	4
YNP	2	10
YUCH	2	8
GNTTP	3	1
YNP	3	5
YUCH	3	1
DENA	4	1
GNTTP	4	2
YNP	4	1
YUCH	4	1
YUCH	5	2
YUCH	8	1
YUCH	10	1
YUCH	12	1
YUCH	14	1
YUCH	24	1

```
persist.lead.loss.park <- cassidy.data.persist %>% group_by(park, mortLEAD) %>%
  summarise(    total = n()) %>% arrange(mortLEAD)
persist.lead.loss.park
```

Table 4: Tabulation of observations by park and a count of all human-caused mortality of leaders showing that 8/12 occurrences of 2+ leaders lost are in YUCH.

park	mortLEAD	total
DENA	0	310
GNTTP	0	59
VNP	0	8
YNP	0	225
YUCH	0	125
DENA	1	20
GNTTP	1	9
YNP	1	13
YUCH	1	10
DENA	2	2
GNTTP	2	1
YUCH	2	7
YNP	3	1
YUCH	3	1

2.1.2 Reproduction

For the reproduction models, there are only 49 total “events” (i.e., occurrences of pack dissolution following human-caused mortality) (Table 5). When modeling, these observations are then spread across the cells of the regression leaving very small numbers of observations for some covariate combinations.

```
repro.total <- cassidy.data.repro %>% group_by(mortYN, reprody1) %>%
  summarise( total = n()) %>% arrange(reprody1)
repro.total
```

Table 5: Tabulation of observations by whether there was human-caused mortality and whether the pack reproduced showing there are 49 instances of pack dissolution following human-caused mortality.

mortYN	reprody1	total
0	0	132
1	0	49
0	1	489
1	1	92

The data from Voyageurs is completely absent for all reproduction models containing pack size and leader loss covariates (Table 6).

```
repro.park <- cassidy.data.repro %>% group_by(park, reprody1) %>%
  summarise( total = n()) %>% arrange(park)
repro.park
```

Table 6: Tabulation of observations by park and whether the pack reproduced showing there are no data points from VNP.

park	reprody1	total
DENA	0	93
DENA	1	227
GNTF	0	17
GNTF	1	49
YNP	0	35
YNP	1	201
YUCH	0	36
YUCH	1	104

2.2 YUCH is unique

YUCH is a National Preserve and, like many preserves in Alaska, hunting and trapping occurs legally within the conservation unit boundary. As a result, the comparison to other units and transboundary comparisons are less clear. Additionally, a large-scale management intervention (i.e., predator control efforts) surrounded YUCH in some years. As stated above, nearly all of the occurrences of losses of large numbers of wolves from a pack and/or both leaders were due to the application of a management action with the express purpose of removing entire packs and reducing wolf populations. Furthermore, mortalities in YUCH wolf control events were not independent (e.g., multiple or all wolves in a pack were removed in the same event). To test the influence of data from YUCH on the overall conclusions, the best persistence model was refit with observations from YUCH excluded. The coefficients for all and leader mortality change markedly and leader mortality was no longer statistically significant ((Table 7); $p=0.14$). Hence the results based on data including the YUCH wolf control events are potentially biased or at a minimum require a caveat stating that the type and magnitude of mortality in YUCH is not likely representative of other areas.

```
best_p <- glmer(persisty1 ~ (1|biolyr)+(1|pack)+packsizeAug+mortALL+mortLEAD,
               family=binomial,
               data=cassidy.data.persist)
noYUCHdat <- cassidy.data.persist[cassidy.data.persist$park != 'YUCH',]
noYUCH <- glmer(persisty1 ~ (1|biolyr)+(1|pack)+packsizeAug+mortALL+mortLEAD,
               family=binomial,
               data=noYUCHdat)
BP <- filter(broom.mixed::tidy(best_p), effect == "fixed") %>%
  mutate(data = "All") %>%
  select(data, term, estimate, std.error, p.value)
NY <- filter(broom.mixed::tidy(noYUCH), effect == "fixed") %>%
  mutate(data = "No YUCH") %>%
  select(data, term, estimate, std.error, p.value)

comp <- cbind(BP, NY)
comp
```

Table 7: Comparison of the fixed effects from the best persistence model fit with all data to the best persistence model fit without observations from YUCH.

data	term	estimate	std.error	p.value	data	term	estimate	std.error	p.value
All	(Intercept)	0.868	0.240	0.000	No YUCH	(Intercept)	0.796	0.277	0.004
All	packsizeAug	0.233	0.038	0.000	No YUCH	packsizeAug	0.250	0.043	0.000
All	mortALL	-0.321	0.148	0.030	No YUCH	mortALL	-0.948	0.279	0.001
All	mortLEAD	-1.291	0.377	0.001	No YUCH	mortLEAD	-0.710	0.483	0.142

2.3 Confounded covariates

The two different mortality variables used are heavily confounded as each instance of a leader mortality also gets counted in the covariate for all human-caused mortality. The fact that leader mortality cannot be separated from all mortality makes the estimation and interpretation of each's independent impact difficult.

```
### Testing if the 'mortALL' variable is always equal to the number of leader morts plus the
### number of non-leader morts
mortContamination <- cassidy.data$mortALL == cassidy.data$mortLEAD + cassidy.data$mortNONLEAD
unique(mortContamination) ### Note there are no FALSE indicating mortLEAD is part of mortALL

## [1] TRUE NA

### Note all NAs are due to missing values
# cassidy.data[is.na(cassidy.data$mortContamination),]
```

2.4 Inconsistencies

Dissolution is defined in the paper as fewer than two resident wolves, but there are four observations in the data where the pack size is 1 and the pack persisted (persisty1 = 1).

```
nrow(filter(cassidy.data, packsizeAug < 2 & persisty1 == 1))

## [1] 4
```

There are four packs with only a single year of data where the pack size was zero (Table 8). The Tonzona pack only had one year of data where the pack size was one, that wolf was killed, and the pack did not persist (Table 8).

```
singles <- filter(as.data.frame(table(cassidy.data$pack)), Freq==1)$Var1
singles_sm <- filter(cassidy.data, pack %in% singles & packsizeAug < 2)
singles_sm
```

Table 8: Packs with only a single year of data with pack sizes less than 2.

park	biolyr	pack	packcode	packsizeAug	mortYN	mortALL	mortLEAD	mortNONLEAD	reprod1	persisty1
DENA	2009	Tonzona	116	1	1	1	0	1	0	0
YNP	2009	694Fgroup	8	0	0	0	0	0	0	0
YNP	2021	carnelian	21	0	0	0	0	0	0	0
YNP	1996	lonestar	37	0	0	0	0	0	0	0
YUCH	2001	Snowy Peak1	145	0	0	0	0	0	0	0

There are 3 cases where the pack dissolved but the following year has data that was included in the reproduction models (Table 9).

```
dbd <- cassidy.data %>% group_by(pack) %>% arrange(pack, biolyr) %>%
  mutate(dbd = ifelse(persisty1 == 0 & lead(!is.na(reprod1)) & pack == lead(pack), 1, 0),
         dbd2 = ifelse(lag(persisty1 == 0) & !is.na(reprod1) & pack == lag(pack), 1, 0))

diss_repr <- filter(dbd, dbd == 1 | dbd2 == 1) %>% arrange(pack, biolyr)
diss_repr
```

Table 9: Three cases where the pack dissolved, but then had an observation the following year that was used in the reproduction modeling.

park	biolyr	pack	packcode	packsizeAug	mortYN	mortALL	mortLEAD	mortNONLEAD	reprody1	persisty1
DENA	2003	Castle Rocks2	63	7	1	1	1	0	0	0
DENA	2004	Castle Rocks2	63	1	0	0	0	0	0	NA
YNP	2009	druid	26	12	0	0	0	0	0	0
YNP	2010	druid	26	0	0	0	0	0	0	NA
YNP	2005	nezperce	39	11	0	0	0	0	0	0
YNP	2006	nezperce	39	0	0	0	0	0	0	NA

There are 5 cases where the number of human-caused mortalities is larger than the stated pack size (Table 10).

```
tab10 <- filter(cassidy.data, cassidy.data$mortALL > cassidy.data$packsizeAug)
tab10
```

Table 10: Five cases where the number of human-caused mortalities is larger than the stated pack size.

park	biolyr	pack	packcode	packsizeAug	mortYN	mortALL	mortLEAD	mortNONLEAD	reprody1	persisty1
DENA	2003	Death Valley	70	0	1	1	0	1	0	0
YNP	2017	canyon	20	2	1	3	3	0	0	0
YUCH	2002	Hard Luck Creek	135	0	1	5	1	4	0	0
YUCH	2013	Lost Creek	137	11	1	12	3	9	0	0
YUCH	2000	Lower Charley1	138	1	1	2	2	0	0	0

There are cases where all members of the pack were killed but no leader loss listed (Table 11).

```
tab11 <- cassidy.data[cassidy.data$pack=='Yukon Fork' & cassidy.data$biolyr==2013 |
  cassidy.data$pack=='Webber Creek 2' & cassidy.data$biolyr==2009,]
tab11
```

Table 11: Examples of all members of pack being killed, but no leader loss listed.

	park	biolyr	pack	packcode	packsizeAug	mortYN	mortALL	mortLEAD	mortNONLEAD	reprody1	persisty1
743	YUCH	2009	Webber Creek 2	153	NA	1	4	0	4	0	0
746	YUCH	2013	Yukon Fork	156	2	1	2	0	2	0	1

3 Statistical Method Issues

3.1 Incorrect application of AICc

```
### AICc function
AICc <- function(LOG.LIK,num.obs,num.par){(2*num.par-2*LOG.LIK)+
  (((2*num.par)*(num.par+1))/(num.obs-num.par-1))}
```

After recreating the reported models to the best of our ability in R, it appears that the model selection criterion (AICc) was applied incorrectly. Due to missing covariate values, it appears that AICc comparisons were made among models fit from different data sets. For the null model of the persistence models, the authors report a log-likelihood of -302.387 and an AICc of 610.80. Below, the null model we fit using all data (almost) matches those values while the null model fit with the data that was available for more complex models does not (Table 12). The discrepancy between our null model with all data and theirs is likely an artifact of numerical optimization in R vs. STATA.

```

nrow(cassidy.data) # all data
nrow(cassidy.data.persist) # data subsetted to complete covariates for most complex model
null.all <- glmer(persisty1~(1|biolyr)+(1|pack),
  family = binomial,
  data = cassidy.data)
null.sub <- glmer(persisty1~(1|biolyr)+(1|pack),
  family = binomial,
  data = cassidy.data.persist)
### We assume here that Cassidy et al. (2023) added 2 to the AIC penalty term for the 2 random
### intercepts, as they did not state explicitly how they treated random terms in AIC calculations.
### The log-likelihoods suggest this is accurate.
comp2 <- data.frame(Model = rep("Null", 3),
  Source = c("Reported", "All data", "Data for most
    complex model"),
  n = c("?", nrow(cassidy.data), nrow(cassidy.data.persist)),
  log_likelihood = c(-302.387, logLik(null.all), logLik(null.sub)),
  AICc = c( 610.80,
    AICc(as.numeric(logLik(null.all)),
      length(null.all@resp$n),
      length(fixef(null.all))+2),
    AICc(as.numeric(logLik(null.sub)),
      length(null.sub@resp$n),
      length(fixef(null.sub))+2)))
comp2

```

Table 12: The log-likelihood and AICc from the null model of persistence fit with all data more closely matches the reported values than the null model fit when using the data available for the most complex model. This suggests that the authors were comparing AICc among models with different datasets.

Model	Source	n	log_likelihood	AICc
Null	Reported	?	-302.387	610.800
Null	All data	864	-301.760	609.548
Null	Data for most complex model	791	-276.849	559.729

3.2 Singularity and lack of convergence

We fit the models in WebTable 2 (the persistence model set) in R with the same dataset for each model (Table 13). Models 1 and 2 were singular as they estimated one or both random effects as 0. Model 4 failed to converge. Unless this is solely an artifact of fitting in R vs. STATA, it appears that the authors ignored these potential concerns and the data are inadequate to fit these models.

```

null.p <- glmer(persisty1~(1|biolyr)+(1|pack),
  family = binomial,
  data = cassidy.data.persist)

## boundary (singular) fit: see help('isSingular')

m2.p <- glmer(persisty1~(1|biolyr)+(1|pack)+packsizeAug,
  family=binomial,
  data=cassidy.data.persist)

## boundary (singular) fit: see help('isSingular')

```

```

m3.p <- glmer(persisty1~(1|biolyr)+(1|pack)+packsizeAug+mortALL,
              family=binomial,
              data=cassidy.data.persist)
m4.p <- glmer(persisty1~(1|biolyr)+(1|pack)+packsizeAug+mortLEAD,
              family=binomial,
              data=cassidy.data.persist)

```

```

## Warning in checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv, :
## Model failed to converge with max|grad| = 0.0136122 (tol = 0.002, component 1)

```

```

m5.p <- glmer(persisty1~(1|biolyr)+(1|pack)+packsizeAug+mortALL+mortLEAD,
              family=binomial,
              data=cassidy.data.persist)

```

```

REs <- rbind(filter(broom.mixed::tidy(null.p), effect == "ran_pars") %>%
  mutate(model = 1, terms = "null") %>%
  select(model, group, estimate),
  filter(broom.mixed::tidy(m2.p), effect == "ran_pars") %>%
  mutate(model = 2, terms = "packsize") %>%
  select(model, group, estimate),
  filter(broom.mixed::tidy(m3.p), effect == "ran_pars") %>%
  mutate(model = 3, terms = "packsize + mortALL") %>%
  select(model, group, estimate),
  filter(broom.mixed::tidy(m4.p), effect == "ran_pars") %>%
  mutate(model = 4, terms = "packsize + mortLEAD") %>%
  select(model, group, estimate),
  filter(broom.mixed::tidy(m5.p), effect == "ran_pars") %>%
  mutate(model = 5, terms = "packsize + mortALL + mortLEAD") %>%
  select(model, group, estimate))

```

REs

Table 13: Random effect estimates from the persistence models. Note that these were fit using the same dataset for all.

model	group	estimate
1	pack	0.7345
1	biolyr	0.0000
2	pack	0.0000
2	biolyr	0.0000
3	pack	0.5143
3	biolyr	0.0002
4	pack	0.5028
4	biolyr	0.0041
5	pack	0.5051
5	biolyr	0.0003

3.3 Incorrect random effect structure

The stated random effect structure is incorrect because it implies that year and pack are crossed effects. That treats each year as a single offset, despite the fact that the parks and accordingly packs are spread across thousands of miles and jurisdictional boundaries and unlikely to be experiencing the same biological (e.g., climate) or jurisdictional (e.g., hunting seasons) conditions. In effect a particular year (e.g., 2001) in one park has nothing to do with that same year in another (i.e., the climate, hunting seasons, etc. that “year” represents are unique to each park and not shared across parks). A more accurate structure would be to have year nested within park and those crossed with packs.

4 Literature Cited

- Cassidy, K. A., B. L. Borg, K. J. Klauder, M. S. Sorum, R. Thomas-Kuzilik, S. R. Dewey, J. A. Stephenson, D. R. Stahler, T. D. Gable, J. K. Bump, A. T. Homkes, S. K. Windels, and D. W. Smith. 2023. Human-caused mortality triggers pack instability in gray wolves. *Frontiers in Ecology and the Environment* doi:10.1002/fee.2597.
- R Core Team. 2023. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Schmidt, J. H., J. W. Burch, and M. C. MacCluskie. 2017. Effects of control on the dynamics of an adjacent protected wolf population in interior Alaska. *Wildlife Monographs* 198:1–30.