A Proposed Poisson model for operating vs. nonoperating effect on turbine-related fatalities.

By Julie Yee
September 3, 2007

Starting Assumptions. Basic model assumptions about the number of mortalities found at a turbine string on a single survey:

1. The number is random and follows a Poisson distribution.
2. The mean varies systematically according to several factors including, but not limited to the following variables:
    - # of days since last survey or clearing search
    - Operating vs. nonoperating status of turbine string
    - String size (i.e. # of turbines, or total MW, in string)
    - Year (year 1 vs. 2)
    - String location (such as north vs. south)

Scavenger Removal. Since mortalities accumulate on a daily basis, I assume the cumulative mean number of mortalities to be proportional to the number of days since the last survey or clearing search. In other words, after an interval of 40 days, there should be about 40 times the mortality as that which would occur after 1 day. The number of observed mortality should be less due to scavenging rates. I used the results from Smallwood (P44) to adjust the expected number of mortalities after a certain number of days accounting for scavenger removal. For surveys that overlapped with portions of both the operating and non-operating periods, then the order in which the periods occurred is important because mortalities occurring in the earlier of the two periods will be subject to higher rates of scavenger removal. As for observer detection error or other errors requiring a proportionality adjustment, I assume that this does not influence relative observed mortality as it relates to the comparison between operating and non-operating turbines.

Data Structure and Variable Names.
Data includes the following variables, broken down on the string and survey level:

nMorts = # of observed mortalities
nDays1 = # of days since last survey or clearing search[1]
ONO1 = operating vs. non-op status in 1st part of period (0 if op, 1 if non-op)
nDay2 = # of days in 2nd part of survey period survey interval[2]
ONO2 = operating vs. non-op status in 2nd part of period[3]
nTurbs = # of turbines in string
nMW = # of MW capacity in string
Year = year 1 or 2 indicator (0 if year 1, 1 if year 2)
NS = north vs. south indicator (0 if north, 1 if south)
StringID = turbine string ID

---

[1] If period is split into both operating and non-op status, then use the # of days in 1st part of survey period
[2] Use zero if period was not split.
[3] In general, should be opposite of ONO1.

Building the Model One Day at a Time. The primary goal is to assess the effect of ONO (ONO1 or ONO2 equally) by expressing the surveyed response variable "nMorts" in terms of the other factors in the data. Since operating and non-operating status can change at any day in the interval preceding each survey, then it's necessary to express the model in terms of daily mortality. This is especially true when considering that the proportion of mortalities observed vary by the number of days since mortality occurred, due to scavenging. Define μ_d = mean number of daily mortalities at an operating string (prior to any scavenger removal). I assume the daily mortality to be proportional to string size, such as nTurbs (or nMW).[4] This implies that the daily mortalities per operating turbine is μ_d/nTurbs, which is a parameter *a*, to be estimated by the model. I also assume that when the string is non-operating, the daily mortality is adjusted by a factor *b*, also to be estimated by the model. The value (1-*b*)100% can be interpreted as the percent reduction in mortality due to shutdown. The model form of the daily mortality rate is:

$$\mu_d = a \times nTurbs \times b^{ONO}$$

(eq. 1)

Because of scavenging, only a proportion (R_i/100 from Smallwood P44) of each daily contribution to the cumulative mortality across a survey period is observed. The mean observed "nMorts" is therefore modeled as

$$\mu_d = \sum_{i=1}^{nDays} R_i \mu_d = \mu_d \sum_{i=1}^{nDays} R_i \ ,$$

(eq. 2)

where R_i = 121.86 – 34.54 ln(i+1) for small raptors[5], and R_i = 106.43-5.16 ln(i+1) for large raptors (Smallwood P44). For situations when the survey interval overlaps an operating and a non-operating period, then use the following generalized expression of the model:

$$\mu_d = \mu_{d1} \sum_{i=nday2+1}^{nDays1+nDays2} R_i \quad + \quad \mu_{d2} \sum_{i=1}^{nDays2} R_i \ ,$$

(eq. 3)

where

$$\mu_{d1} = a \times nTurbs \times b^{ONO_1} \text{ , and } \mu_{d2} = a \times nTurbs \times b^{ONO_2} \ .$$

(eqs. 4a, 4b)

Additional factors such as Year or NS or Year*NS can be incorporated into the model in a similar way.

The model can be fit using PROC NLMIXED in SAS. See sample code "op-nonop using nlin.sas" for an example. Output and code for a simple test dataset are on following pages.

---

[4] Even if the assumption is not quite true, it is convenient for expressing mortality rates on a per turbine or per MW basis.

[5] This function becomes negative as *i* increases (in this case at day 34). I recommend bounding R_i to zero when this occurs.

Obs	n Morts	n Days1	ono1	n Days2	ono2	n Turbs	nMW
1	0	10	0	30	1	5	350
2	2	30	0	15	1	4	250
3	3	40	0	0	1	7	400

The NLMIXED Procedure

Specifications

Data Set	WORK.TEST4
Dependent Variable	nMorts
Distribution for Dependent Variable	Poisson
Optimization Technique	Dual Quasi-Newton
Integration Method	None

Dimensions

Observations Used	3
Observations Not Used	0
Total Observations	3
Parameters	2

Parameters

a	b	NegLogLike
0.000137	0.99	27.0747182

Iteration History

Iter	Calls	NegLogLike	Diff	MaxGrad	Slope
1	16	12.3149598	14.75976	1602.816	-1.323E7
2	19	7.84106685	4.473893	481.4711	-0.21926
3	21	6.07127739	1.769789	120.3611	-1.11367
4	22	5.98942681	0.081851	77.96985	-0.11958
5	23	5.97194419	0.017483	54.57138	-0.03041
6	25	5.963636	0.008308	36.23136	-0.00377
7	32	5.04752019	0.916116	98.35202	-0.01214
8	34	4.52691915	0.520601	7.972499	-0.41277
9	38	4.24474387	0.282175	26.29231	-1.54937
10	40	4.19602144	0.048722	10.01367	-0.20593
11	41	4.11574141	0.08028	7.259701	-0.07738
12	43	4.10277502	0.012966	3.628088	-0.02194
13	45	4.09926444	0.003511	0.039891	-0.00779
14	47	4.09911224	0.000152	0.013121	-0.00032
15	49	4.09911191	3.359E-7	0.00036	-6.84E-7
16	51	4.09911191	2.16E-10	0.000014	-423E-12

NOTE: GCONV convergence criterion satisfied.

The NLMIXED Procedure

Fit Statistics

-2 Log Likelihood	8.2
AIC (smaller is better)	12.2
AICC (smaller is better)	24.2
BIC (smaller is better)	10.4

Parameter Estimates

Parameter	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper	Gradient
a	0.05412	0.03124	3	1.73	0.1817	0.05	-0.04532	0.1535	-0.00001
b	0.2183	0.4289	3	0.51	0.6459	0.05	-1.1466	1.5832	-1E-6

output from e:\analyses\altamont\op-nonop using nlin.sas

The NLMIXED Procedure

Specifications

Data Set	WORK.TEST4
Dependent Variable	nMorts
Distribution for Dependent Variable	Poisson
Random Effects	e
Distribution for Random Effects	Normal
Subject Variable	obs
Optimization Technique	Dual Quasi-Newton
Integration Method	Adaptive Gaussian Quadrature

Dimensions

Observations Used	3
Observations Not Used	0
Total Observations	3
Subjects	3
Max Obs Per Subject	1
Parameters	3
Quadrature Points	5

Parameters

logsig	a	b	NegLogLike
0	0.000137	0.99	20.8798705

Iteration History

Iter	Calls	NegLogLike	Diff	MaxGrad	Slope
1	16	8.85430383	12.02557	1119.184	-1.184E7
2	19	6.6366846	2.217619	145.1837	-0.24953
3	22	6.60258759	0.034097	195.7424	-0.00324
4	25	5.64170421	0.960883	53.38455	-0.03775
5	26	5.61159869	0.030106	34.14401	-0.05824
6	27	5.60156546	0.010033	21.67212	-0.00171
7	28	5.59599608	0.005569	15.86415	-0.0056
8	30	5.56147086	0.034525	22.25232	-0.02014
9	32	4.76442491	0.797046	71.36462	-0.0539
10	34	4.48945238	0.274973	16.25508	-0.54305
11	35	4.28069131	0.208761	19.47463	-1.12357
12	36	4.14690988	0.133781	5.68351	-0.76871
13	39	4.10088558	0.046024	1.342591	-0.03855
14	41	4.10000041	0.000885	0.245763	-0.00213

The NLMIXED Procedure

Iteration History

Iter	Calls	NegLogLike	Diff	MaxGrad	Slope
15	43	4.09979635	0.000204	0.028055	-0.00032
16	45	4.09978885	7.504E-6	0.131463	-0.00001
17	47	4.09972502	0.000064	0.054397	-1.98E-6
18	49	4.09928958	0.000435	0.00743	-0.0001
19	50	4.09914513	0.000144	0.062915	-0.00009
20	52	4.099113	0.000032	0.001605	-0.00006
21	54	4.09911293	6.848E-8	0.000809	-9.01E-8
22	56	4.09911223	6.952E-7	0.000068	-4.62E-8
23	57	4.09911192	3.132E-7	0.00069	-1.74E-7
24	59	4.09911191	9.884E-9	0.00002	-1.93E-8

NOTE: GCONV convergence criterion satisfied.

Fit Statistics

-2 Log Likelihood	8.2
AIC (smaller is better)	14.2
AICC (smaller is better)	38.2
BIC (smaller is better)	11.5

Parameter Estimates

Parameter	Estimate	Standard Error	DF	t Value	Pr > |t|	Alpha	Lower	Upper	Gradient
logsig	-10.3198	13030	2	-0.00	0.9994	0.05	-56072	56052	3.012E-9
a	0.05412	0.03124	2	1.73	0.2254	0.05	-0.08032	0.1885	-0.00002
b	0.2183	0.4289	2	0.51	0.6614	0.05	-1.6270	2.0636	-1.63E-7

```
footnote 'output from e:\analyses\altamont\op-nonop using nlin.sas';
options pageno=1;

* written by Julie Yee, September 3, 2007 ;

data test;
       input nMorts nDays1 ono1 nDays2 ono2 nTurbs nMW;
              * variables defined as follows:
              nMorts = num fatalities counted [integer]
              ndays1 = num days in 1st part of survey period (if period is split into both operating and non-op status) [integer]
              ono1 = operating/non-operating status of 1st part of survey period [0 if op, 1 if non-op]
              ndays2 = num days in 2nd part of survey period (zero if period was not split) [integer]
              ono2 = operating/non-operating status of 2nd part of survey period [should be opposite of ono1]
              nTurbs = num turbines in string
              nMW = num megawatts in string
              possibly add the following in later :
              stringID = turbine string id
              year = year 1 or 2 indicator [0 if year 1, 1 if year 2]
              ns = north/south indicator [0 if north, 1 if south] ;
       cards;
0 10 0 30 1 5 350
2 30 0 15 1 4 250
3 40 0 0 1 7 400
;
       run;
title 'Simple Test Dataset';
proc print;
       run;

* calculate RC1 and RC2 (RC for period parts 1 and 2, where RC is the numerator of eq 3 from Smallwood document P44) ;
data test2;
       set test;
       obs=_n_;
       do i=1 to ndays2;
              period=2;
              output;
              end;
       do i=ndays2+1 to ndays1+ndays2;
```

```
                period=1;
                output;
                end;
        run;
data test3;
        set test2;
        retain RC1 RC2;
        by obs;
        if first.obs then do;
                RC1=0;
                RC2=0;
                end;
        RiSM=121.86-34.54*log(i+1);                          * Ri defined on p. 3 of Smallwood document P44, for SMall raptors;
        RiSM=max(0,RiSM);                                            * cap values at zero when they start becoming negative ;
        * OR USE: RiLG=106.43-5.16*log(i+1);  * Ri defined on p. 3 of Smallwood document P44, for LarGe raptors;
        if period=1 then RC1=RC1+RiSM;
        else if period=2 then RC2=RC2+RiSM;
        output;
        run;
data test4;
        set test3;
        by obs;
        if last.obs;
        rename i=nDays;
        RC1 = RC1/100;                                               * convert percentages into proportions ;
        RC2 = RC2/100;
        drop period RiSM;
        run;

* without random turbine string effect ;
title 'NLMIXED model results';
proc nlmixed data=test4;
        parms a 0.0001369 b 0.99 ;
                * initial values:
                a=mean mort/turbine/day=0.05/365=0.0001369
                b=mean effect of shutdown=1% reduction (almost no reduction) ;
        mu1 = a*nTurbs*RC2*(b**ono2);
        mu2 = a*nTurbs*RC1*(b**ono1);
```

```
        mu = mu1+mu2;
        model nMorts ~ poisson(mu);
    run;

* without random turbine string effect, with overdispersion ;
title 'NLMIXED model results, with overdispersion';
proc nlmixed data=test4;
        parms logsig 0 a 0.0001369 b 0.99 ;
        mu1 = a*nTurbs*RC2*(b**ono2)*exp(e);
        mu2 = a*nTurbs*RC1*(b**ono1)*exp(e);
        mu = mu1+mu2;
        model nMorts ~ poisson(mu);
        random e ~ normal(0,exp(2*logsig)) subject=obs;
    run;

/*
* with random turbine string effect, with overdispersion ;
proc nlmixed data=test4;
        parms logsigo 0 logsigs 0 a 0.0001369 b 0.99 ;
        mu1 = a*nTurbs*RC2*(b**ono2)*exp(eo+es);
        mu2 = a*nTurbs*RC1*(b**ono1)*exp(eo+es);
        mu = mu1+mu2;
        model y ~ poisson(mu);
        random eo ~ normal(0,exp(2*logsigo)) subject=obs;           * error associated with overdispersion (extra observation
variation) ;
        random es ~ normal(0,exp(2*logsigs)) subject=stringID;      * error associated with turbine string ;
        run;
*/
```