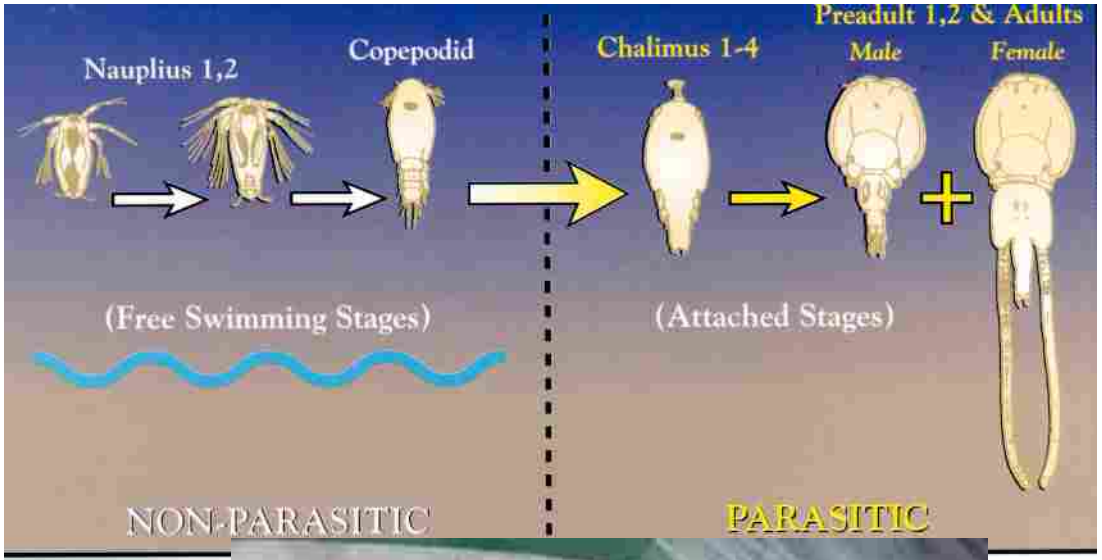
Mediation Analysis: A Case-Study

Ian Dohoo
Javier Sanchez
University of PEI, Charlottetown, Canada

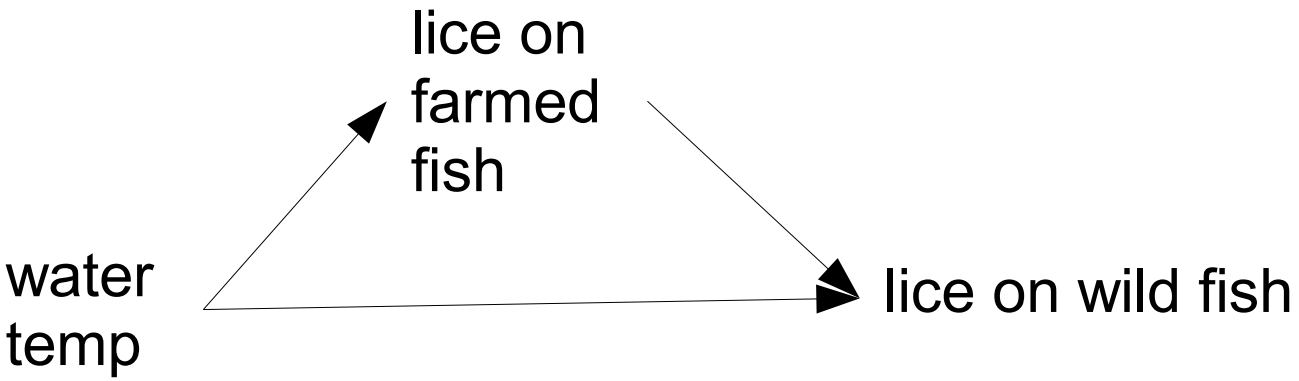
Knut Wiik Vollset
NORCE, Bergen, Norway

Sea Lice on Salmon

- *Lepeophtheirus salmonis*



Motivating Example

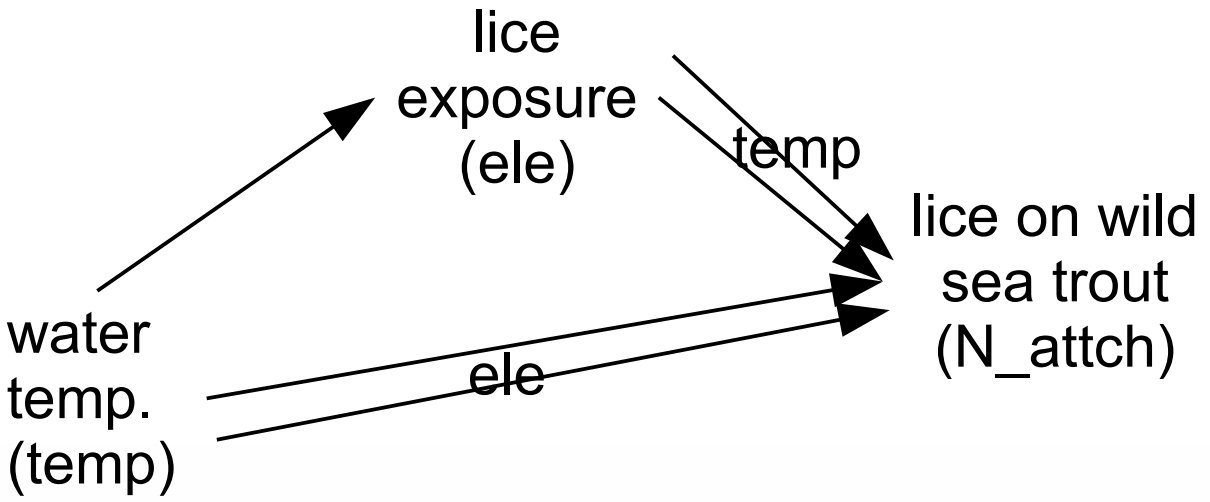


Parasitic sea louse infestations on wild sea trout: separating the roles of fish farms and temperature

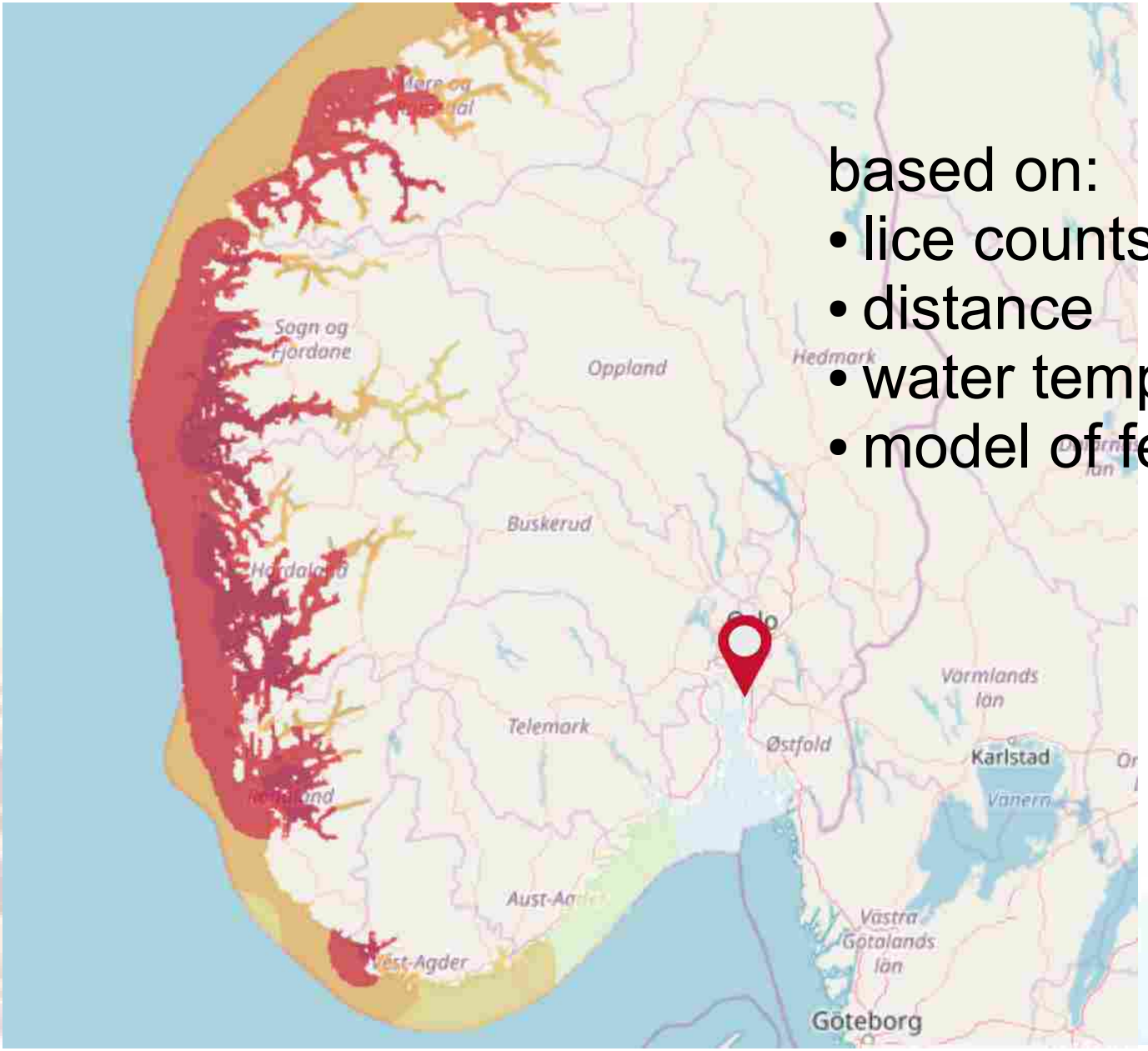
Knut W. Vollset, Lars Qviller, Bjørnar Skår, Bjørn T. Barlaup and Ian Dohoo

Parasites and Vectors, 2018 Nov 29;11(1):609

Motivating Example



ele – Estimated lice exposure



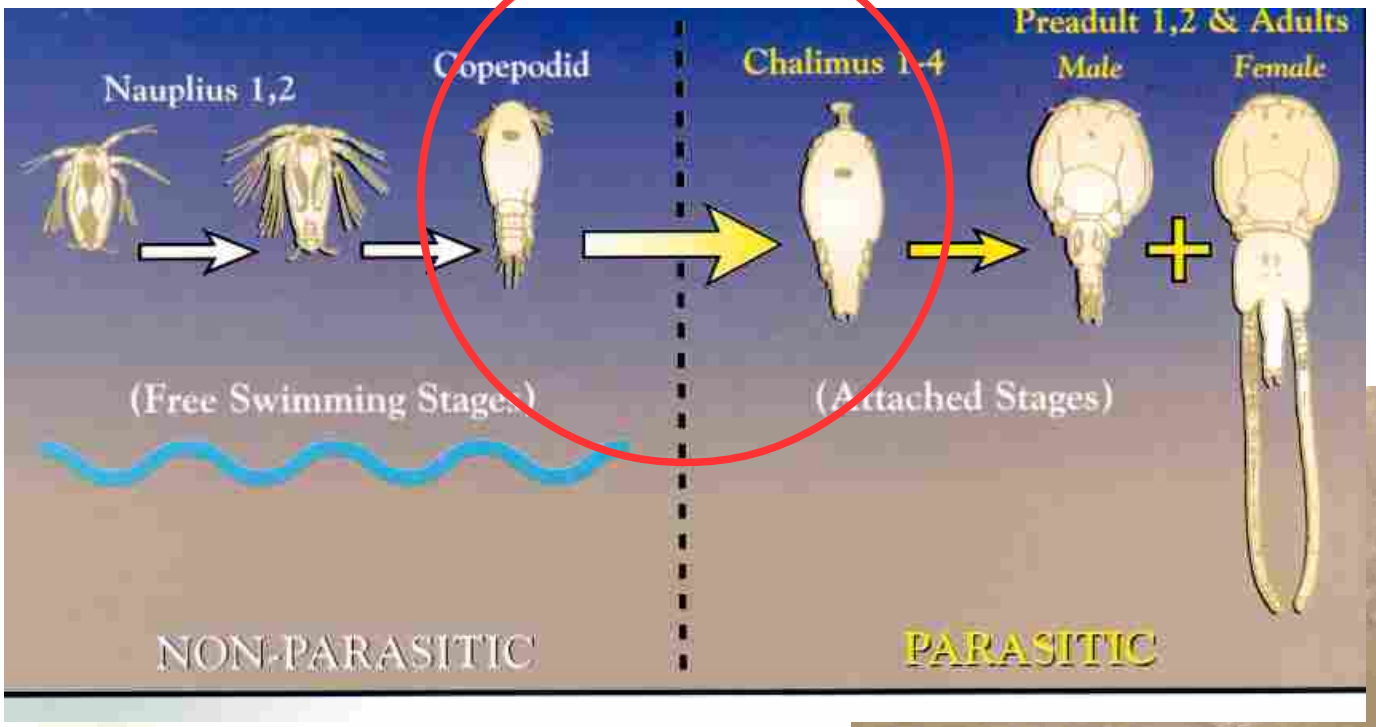
based on:

- lice counts on fish farms
- distance
- water temperature
- model of fecundity

Variables

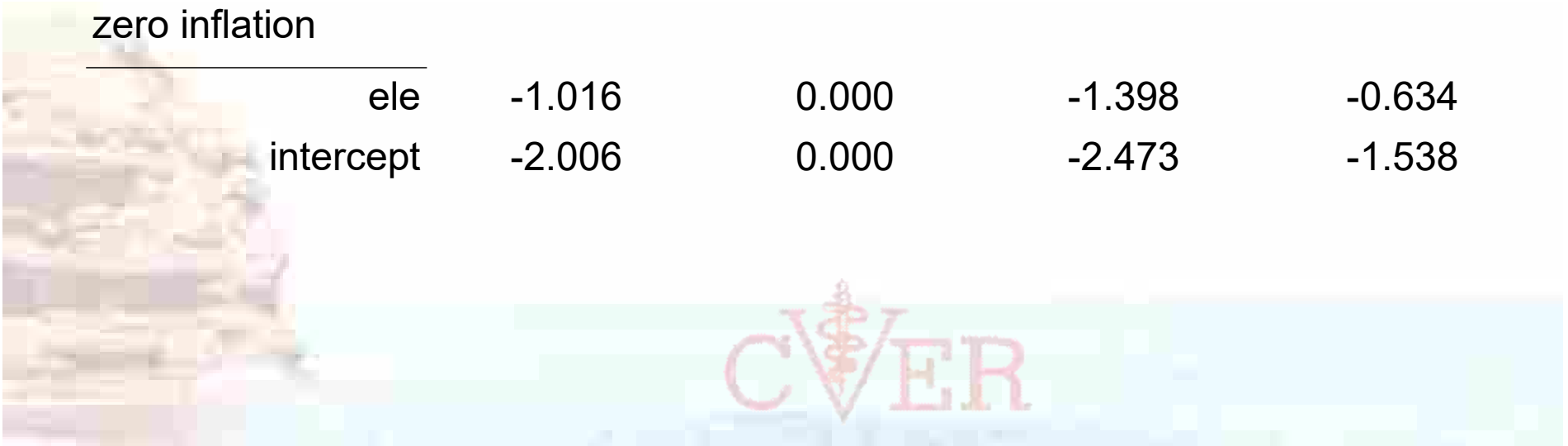
Variable	mean	min	max
temp	9.83	5.96	13.84
ele	14.02	9.22	17.44
N_attch	24.7	0	770
N_mob	11.8	0	105
Standardized Variables			
temp	0	-1.93	2.00
ele	0	-2.28	1.62
N_attch	0	-1.44	2.59
N_mob	0	-1.42	2.06

Attached Lice



Regression - ZINB

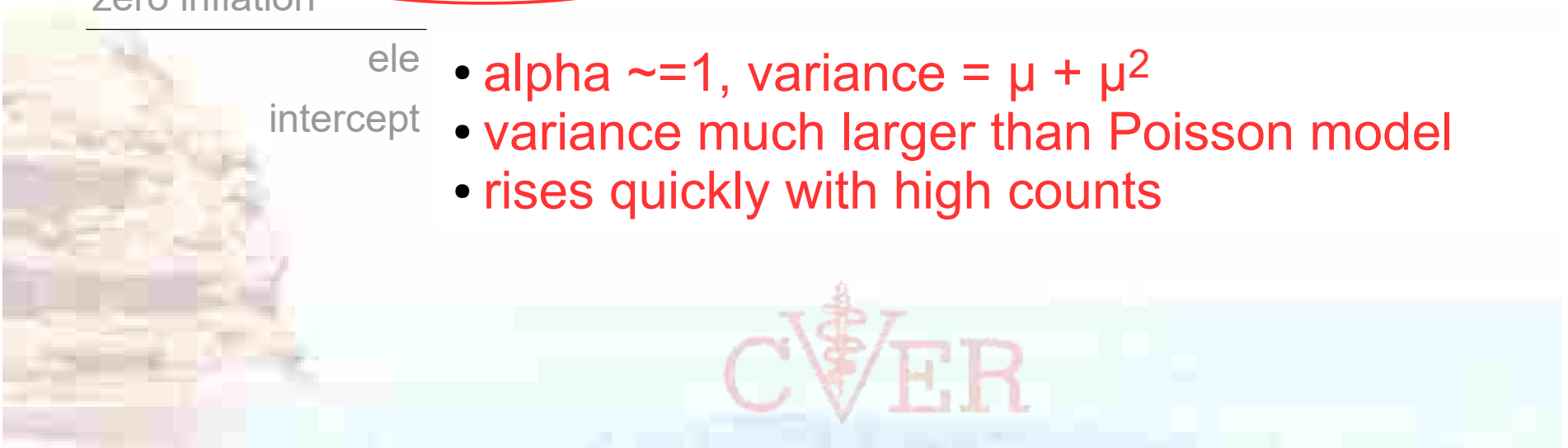
	Coeficient	P	95% Confidence Interval	
negative binomial				
temp	-0.166	0.073	-0.348	0.016
ele	0.909	0.000	0.703	1.114
temp*ele	0.803	0.000	0.614	0.991
ele2	-0.545	0.000	-0.716	-0.374
intercept	2.960	0.000	2.805	3.115
alpha	1.074		0.877	1.315
zero inflation				
ele	-1.016	0.000	-1.398	-0.634
intercept	-2.006	0.000	-2.473	-1.538



Regression - ZINB

	Coeficient	P	95% Confidence Interval	
negative binomial				
temp	-0.166	0.073	-0.348	0.016
ele	0.909	0.000	0.703	1.114
temp*ele	0.803	0.000	0.614	0.991
ele2	-0.545	0.000	-0.716	-0.374
intercept	2.960	0.000	2.805	3.115
alpha	1.074		0.877	1.315
zero inflation				
ele				
intercept				

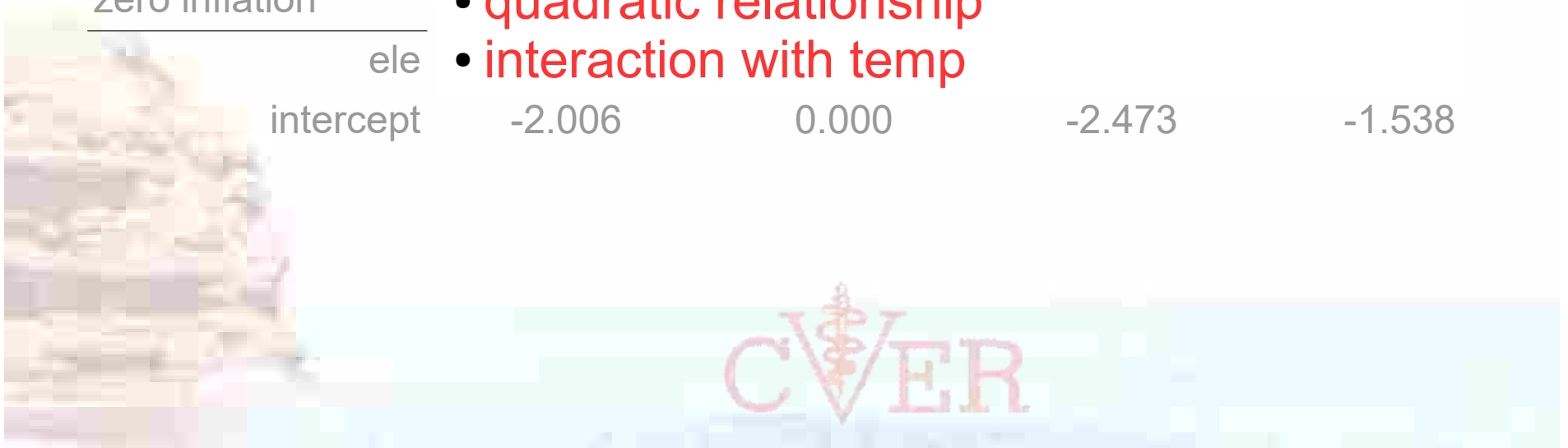
- alpha ~=1, variance = $\mu + \mu^2$
- variance much larger than Poisson model
- rises quickly with high counts



Regression - ZINB

	Coeficient	P	95% Confidence Interval	
negative binomial				
temp	-0.166	0.073	-0.348	0.016
ele	0.909	0.000	0.703	1.114
temp*ele	0.803	0.000	0.614	0.991
ele2	-0.545	0.000	-0.716	-0.374
intercept	2.960	0.000	2.805	3.115
alpha				
zero inflation				
ele				
intercept	-2.006	0.000	-2.473	-1.538

- ele:
- quadratic relationship
- interaction with temp



Regression - ZINB

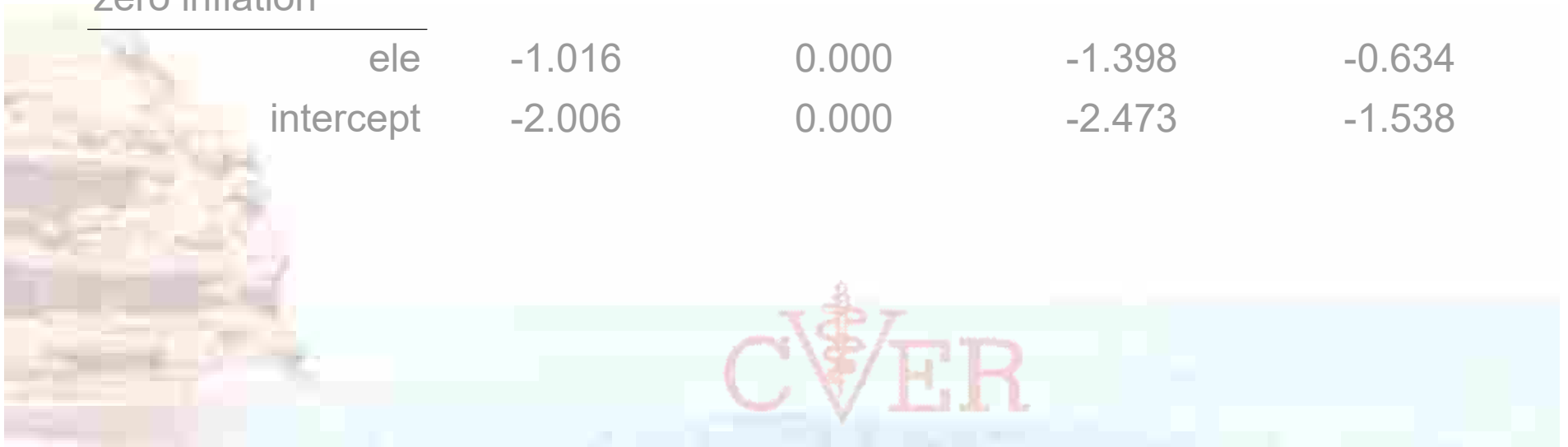
	Coeficient	P	95% Confidence Interval	
negative binomial				
temp	-0.166	0.073	-0.348	0.016
ele	0.909	0.000	0.703	1.114
temp*ele	0.803	0.000	0.614	0.991
ele2	-0.545	0.000	-0.716	-0.374
intercept	2.960	0.000	2.805	3.115
alpha	1.074		0.877	1.315
zero inflation				
ele	-1.016	0.000	-1.398	-0.634
intercept	-2.006	0.000	-2.473	-1.538

- LR test comparing NB and ZINB highly sig.
- ele only factor affecting zero inflation
  - p(zero) goes down linearly with ele

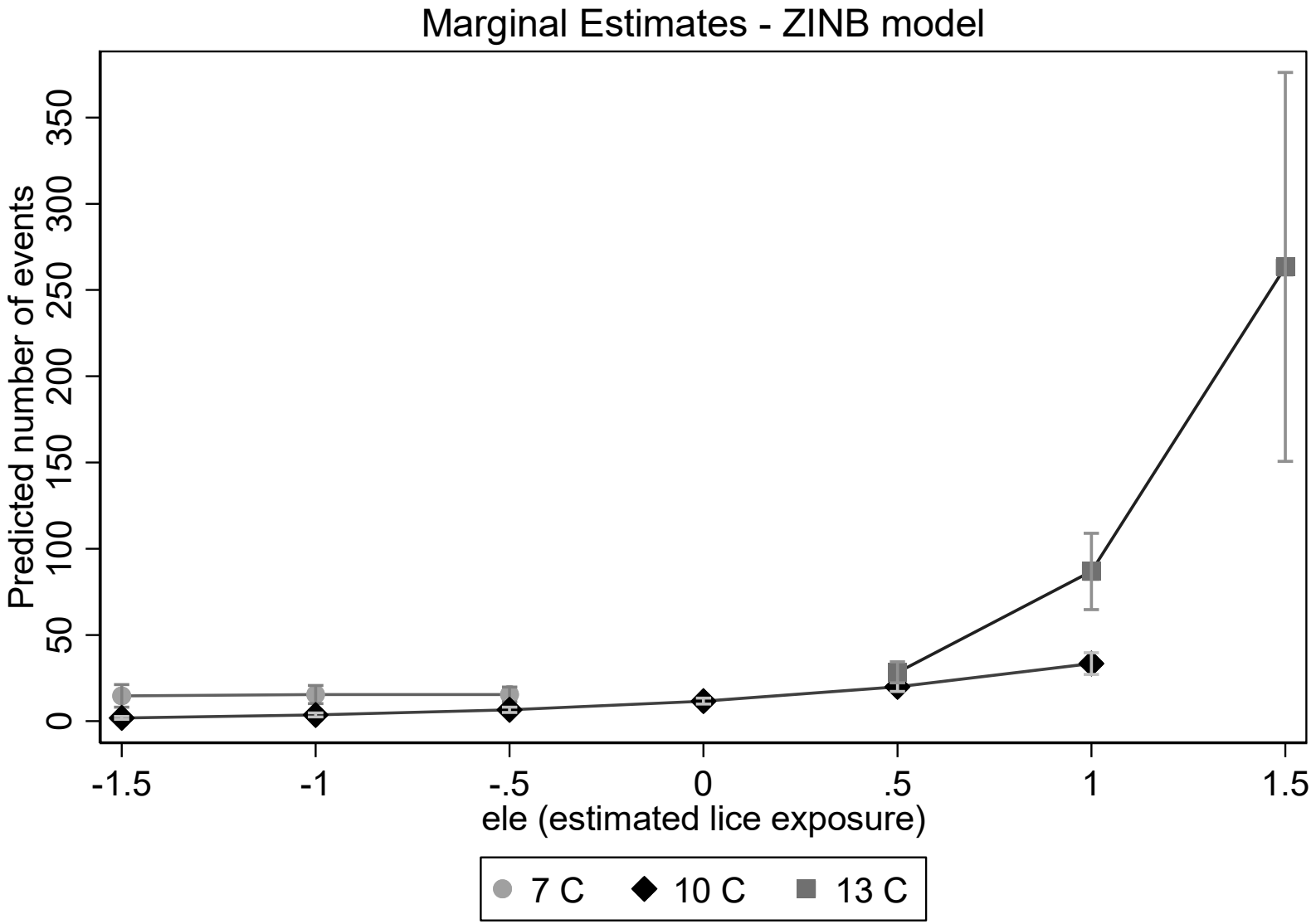
Regression - ZINB

	Coeficient	P	95% Confidence Interval	
negative binomial				
temp	-0.166	0.073	-0.348	0.016
ele	0.909	0.000	0.703	1.114
temp*e				
el				
interce				
alpha				
zero inflation				
ele	-1.016	0.000	-1.398	-0.634
intercept	-2.006	0.000	-2.473	-1.538

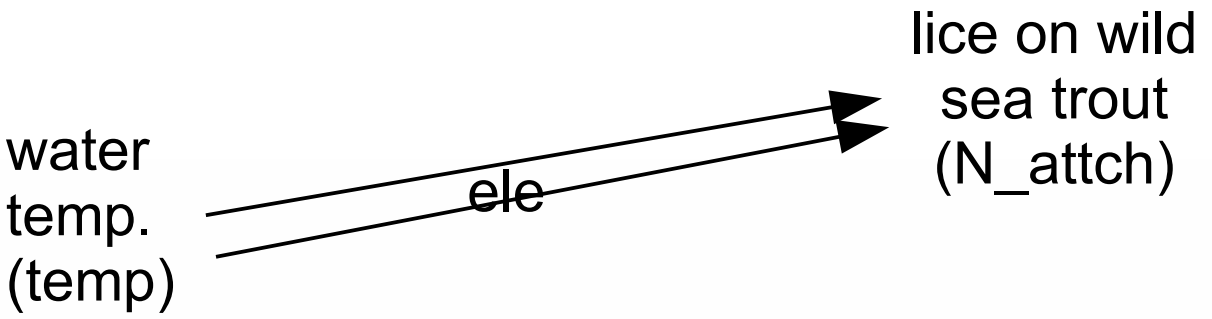
- at average levels of ele, temp only borderline sig.
  - and negative
- evidence of limited direct effect ?



Marginal Estimates

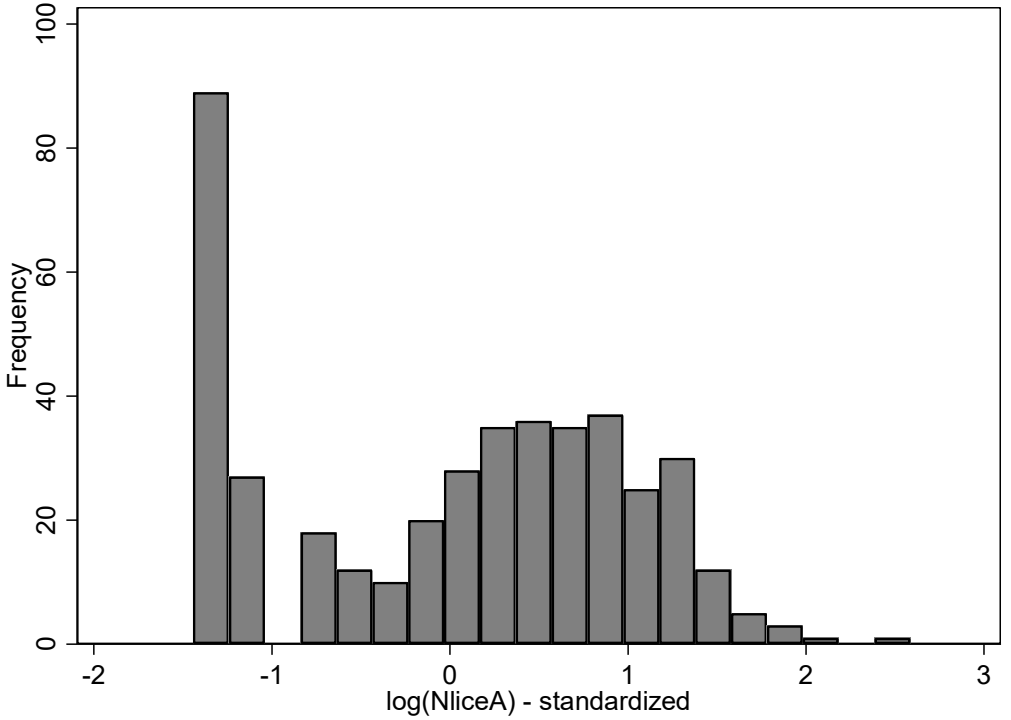


Motivating Example



ZINB vs linear regression

- no software for mediation analysis based on NB regression
- NB regression models log(count)
- would a linear model of log(count) suffice ?



ZINB vs linear regression

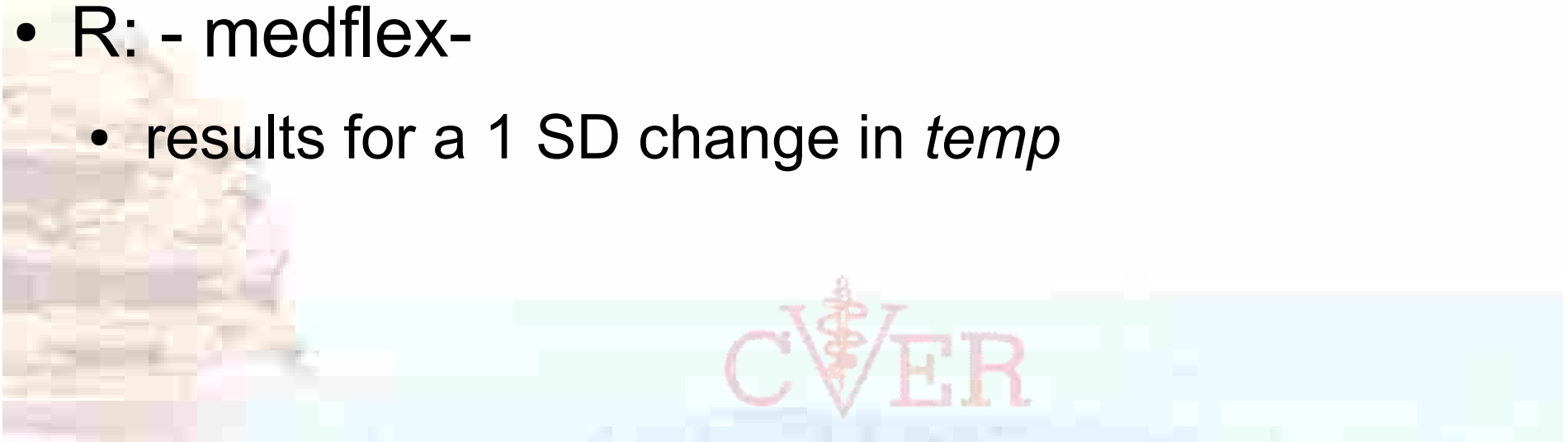
Variable	ZINB	Linear
temp	-0.206*	-0.103
ele	1.0866***	1.084***
ele^2	-0.662***	-0.430***
temp*ele	0.815***	0.688***
Intercept	2.871***	1.921***

* P<0.05 ** P<0.01 *** P<0.001

correlation of predicted values: 0.993

Mediation Analysis

- effect of temp
  - is it mediated through *ele* ?
  - allow for interaction between *temp* and *ele*
- Stata: - medeff-
  - user specifies change in *temp* to consider
- R: - medflex-
  - results for a 1 SD change in *temp*



Direct Effects

effect of temp that are NOT mediated through ele

Effect	Estimate	SE (signif)	Interpretation
Pure direct effect	0.007	0.060	the increase in *N_attch* brought about by a 1 SD increase in *temp* with *ele* held constant at an average temp
Total direct effect	0.156	0.051**	the increase in *N_attch* brought about by a 1 SD increase in *temp* with *ele* held constant at the level resulting from that 1 SD increase in *temp*

* P<0.05 ** P<0.01 *** P<0.001

Indirect Effects

effect of temp that ARE mediated through ele

Effect	Estimate	SE (signif)	Interpretation
Pure indirect effect	0.418	0.047***	the increase in *N_attch* brought about by increasing *ele* by the amount that would result from a 1 SD increase in *temp,* while holding *temp* constant at an average *temp*
Total indirect effect	0.567	0.073***	the increase in *N_attch* brought about by increasing *ele* by an mount that would result from 1 SD increase in *temp,* while holding *temp* constant at that elevated level

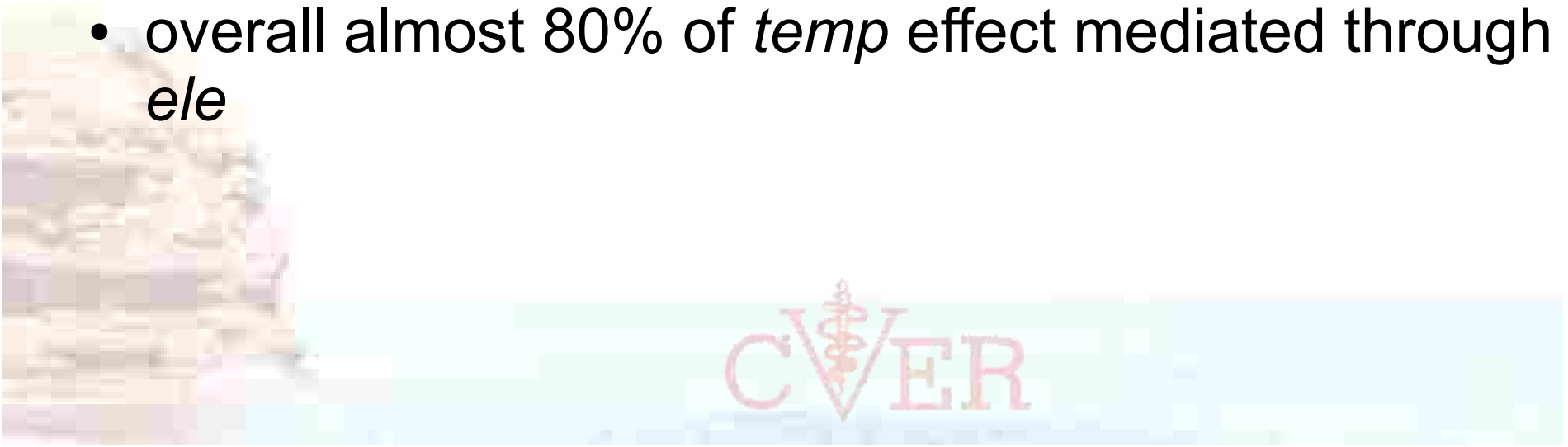
* P<0.05 ** P<0.01 *** P<0.001

% mediated

- total direct effect = 0.156
- total indirect effect = 0.567
- total effect= 0.723
  - a 1 SD increase in *temp* brings about an increase of 0.723 in *N_attch*
- % of total effect mediated through *ele* = 78.4%
  - almost 80% of the effect of rising water temperature results from the impact of water temperature on infestation pressure (*ele*)

Mediation Analysis

- for attached lice
- rising water temperatures
  - interaction with ele
  - small positive direct effect
  - large positive indirect effect
  - overall almost 80% of *temp* effect mediated through *ele*



R: -medflex- vs -mediation-

- after completing this project ... discovered -mediation-
  - much easier to use and understand
  - handles much wider range of models
    - including NB
  - doesn't give direct access to counterfactuals
- comparison of:
  - linear model on log scale
  - NB regression

R: -medflex- vs -mediation-

Proportion mediated		
Model	linear (log) -mdeflex-	NB -mediation-
mediator - ele	63%	65%
mediator – ele, ele^2	61%	59%
interaction between temp and ele	78%	84%

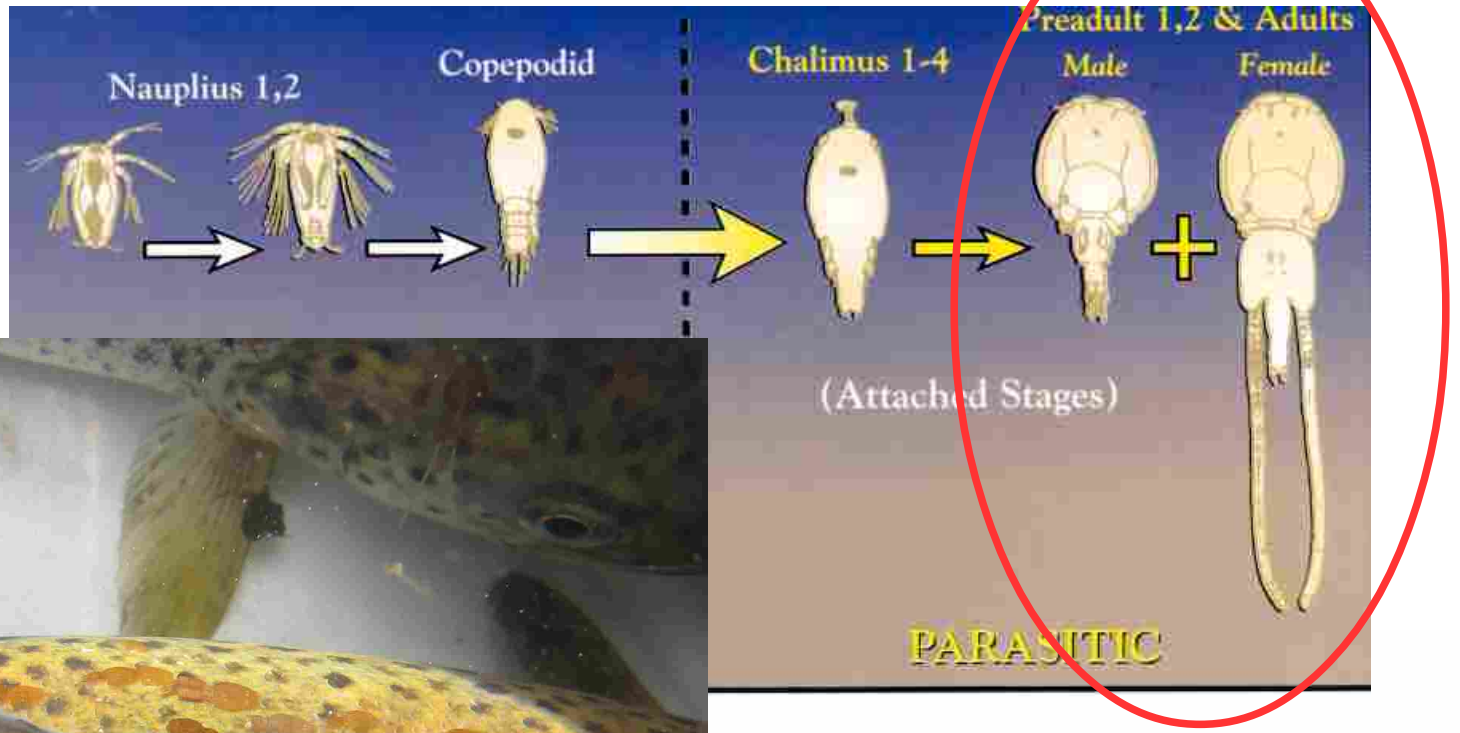
Terminology – no interaction

Stata	**-medflex-**	**-medflex-decomposition**	**-mediation-**
ACME	varnmame1	natural indirect effect	ACME
direct effect	varname0	natural direct effect	ADE
total effect	-	total effect	total effect
% of total effect mediated	-	-	propn. mediated

Terminology – interaction

Stata	-medflex-	-medflex-decomposition	-mediation-
ACME1	presented as model with:	total indirect	ACME (treatment)
ACME0	varnmame1	pure indirect	ACME (control)
direct effect 1	varname0	total direct	ADE (treatment)
direct effect 0	interaction term	pure direct	ADE (control)
total effect		total effect	total effect
avg. mediated		-	ACME (average)
avg. direct effect		-	ADE (average)
% of total effect mediated		-	Prop. mediated (average)

Mobile Lice


Nauplius 1,2
Copepodid
Chalimus 1-4
Preadult 1,2 & Adults
Male
Female
(Attached Stages)
PARASITIC

Regression - NB

	Coeficient	P	95% Confidence Interval	
temp	-0.458	0.000	-0.666	-0.250
$temp^2$	-0.206	0.010	-0.364	-0.049
ele	0.626	0.000	0.446	0.807
ele^2	0.189	0.007	0.051	0.328
intercept	2.295	0.000	2.119	2.471
alpha	1.610		1.387	1.868

- no evidence of variance inflation

Regression - NB

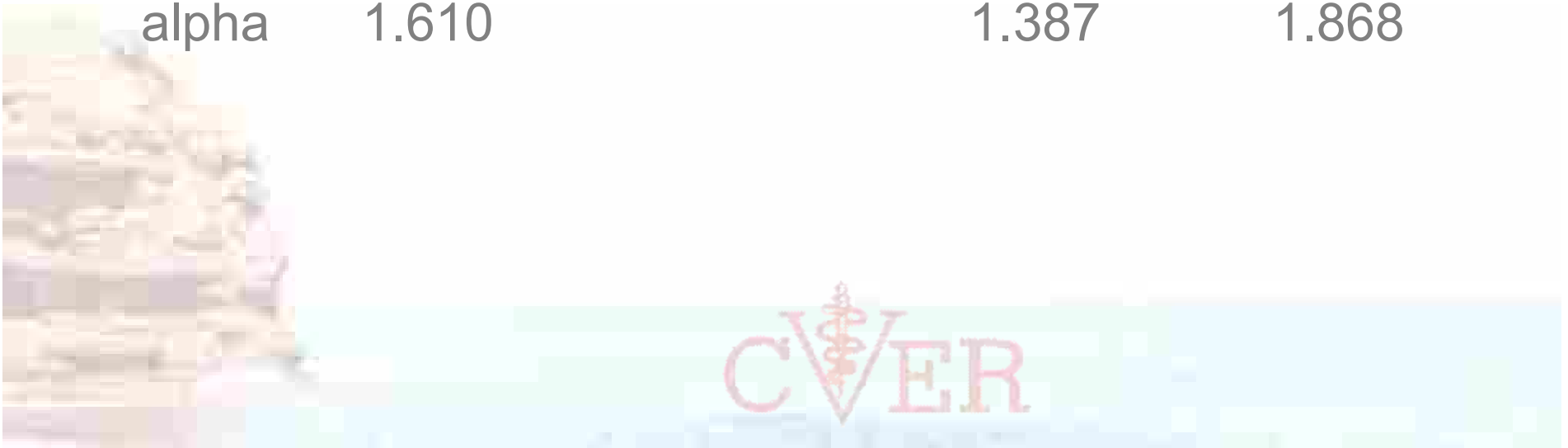
	Coeficient	P	95% Confidence Interval	
temp	-0.458	0.000	-0.666	-0.250
$temp^2$	-0.206	0.010	-0.364	-0.049
ele	0.626	0.000	0.446	0.807
ele^2	0.189	0.007	0.051	0.328
intercept	2.295	0.000	2.119	2.471
alpha	1.610		1.387	1.868

- alpha ~=1.6, variance = $\mu + 2.6(\mu^2)$
- variance much larger than Poisson model
- rises quickly with high counts

Regression - NB

	Coeficient	P	95% Confidence Interval	
temp	-0.458	0.000	-0.666	-0.250
$temp^2$	-0.206	0.010	-0.364	-0.049
ele	0.626	0.000	0.446	0.807
ele^2				
intercept				
alpha	1.610		1.387	1.868

- temp effect is quadratic
- lower mobile lice counts with increasing temperature

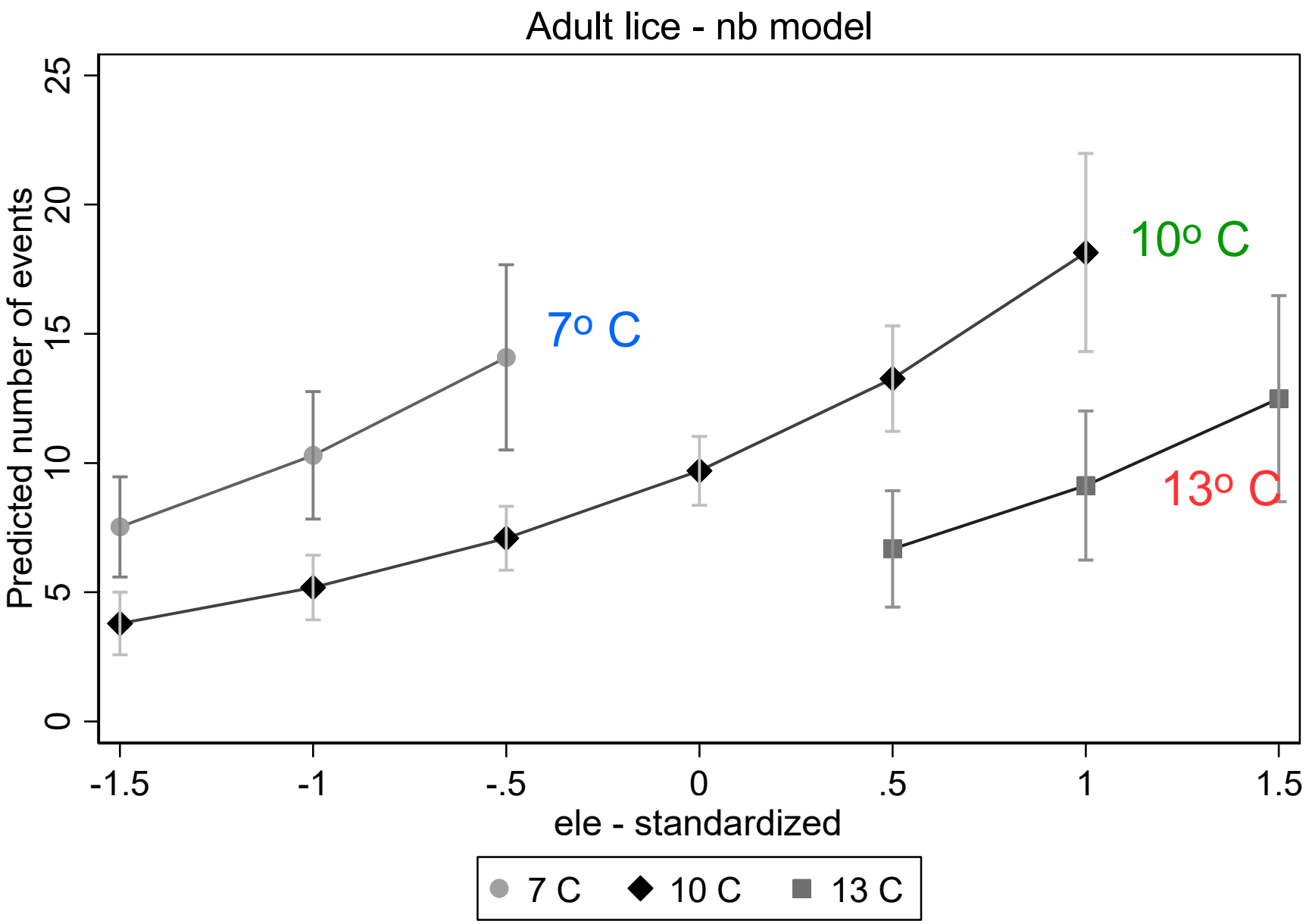


Regression - NB

	Coeficient	P	95% Confidence Interval	
temp	-0.458	0.000	-0.666	-0.250
$temp^2$	-0.206	0.010	-0.364	-0.049
ele	0.626	0.000	0.446	0.807
ele^2	0.189	0.007	0.051	0.328
intercept	2.295	0.000	2.119	2.471
alpha				

- ele effect is quadratic
- higher mobile lice counts with increasing infestation pressure

Marginal Effects



Mediation Analysis

direct and indirect effects of *temp*

Effect	Estimate	SE (signif)	Interpretation
Natural direct effect	-0.43	0.054***	the increase in *N_mob* brought about by a 1 SD increase in *temp* with *ele* held constant at value seen at average *temp*
Natural indirect effect	0.34	0.038***	the increase in *N_mob* brought about by increasing *ele* by the amount that would result from a 1 SD increase in *temp* while holding *temp* constant at an average *temp*
Total effect	-0.1	0.046*	the overall decrease in *N_mob* brought about by increasing *temp* by 1 SD

* P<0.05 ** P<0.01 *** P<0.001

Mediation Analysis

- for mobile lice
- rising water temperatures
  - direct effect = lower lice counts
  - indirect effect = higher lice counts
  - overall = borderline significant reduction in counts with rising temperatures
- counts rose with increasing *ele* at all temperatures

ISVEE 16

HALIFAX 2021

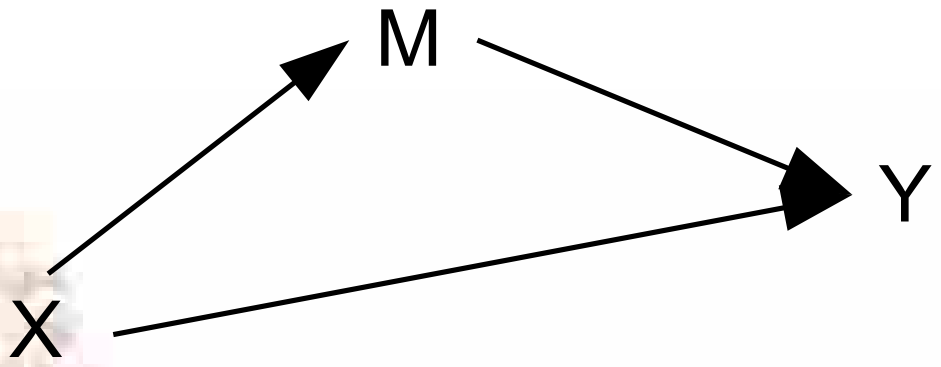
Halifax

... more technical stuff

- Stata /R terminology
- a “peek under the hood”
  - looking at the simulated counterfactual data

A peek under the hood

- assume simple model with:
  - Y = continuous outcome
  - X = dichotomous factor of interest
  - M = dichotomous mediator



Ordinary regression model

$$Y = \beta_c + \beta_0 X + \beta_1 M$$

- β_c = intercept (not relevant)
- β_0 = (direct) effect of X holding M constant
- β_1 = (direct) effect of M holding X constant

Mediation model

$$Y^c = \beta_c + \beta_0 X_0 + \beta_1 X_1$$

- X_0 = a synthesized X value – taken from range of X values
- X_1 = original X value
- Y^c = counterfactual estimate of Y – based on values of X_0 and M
  - β_c = intercept (not relevant)
  - β_0 = (direct) effect of X
  - β_1 = (indirect) effect of X mediated through M


ISVEE 16
HALIFAX 2021