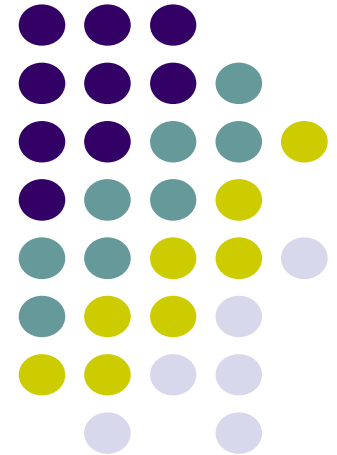


Logistic Regression: Use & Interpretation of Odds Ratio (OR)

Fu-Lin Wang, B.Med.,MPH, PhD
Epidemiologist
Adjunct Assistant Professor

Fu-lin.wang@gov.ab.ca
Tel. (780)422-1825





Background

- **Odds: The ratio of the probability of occurrence of an event to that of nonoccurrence.**
- **Odds ratio (OR, relative odds): The ratio of two odds, the interpretation of the odds ratio may vary according to definition of odds and the situation under discussion.**

- **Consider the 2x2 table:**

	Event	Non-Event	Total
Exposure	a	b	a+b
Non-Exposure	c	d	c+d
Total	a+c	b+d	N

A 2x2 Table for Two Binary Variables



	Lung Ca	No Lung Ca	Total
Smoking	80	20	100
Non-Smoking	20	80	100
Total	100	100	200

Odds for Lung Cancer_{smokers} = $80/20=4.00$

- The probability of having lung cancer among smokers is 4 times of not having lung cancer.

Odds Ratio for Lung Cancer_{smokers} = $(80/20) / (20/80) =16.00$

- The probability of developing lung cancer among smokers is 16 times of that non-smokers.

Why is the odds ratio useful?



If the odds measures exposure-disease relationship

- Determine the strength of association: Strong ($OR > 3$), moderate ($OR = 1.6 - 3.0$), weak ($OR = 1.1 - 1.5$)
- Assess the impact of confounding variables
- Estimate the relative risk for a disease in relation to a given risk factor



Why is the odds ratio useful (cont'd)?

If the odds measures other event to non-event (reference) relationship or spatial/temporal trend

- The likelihood to delivery LBW babies for mothers 35 years or older is 2.5-times of that for mothers 20-34 years
- The rate of MVA in Northern Alberta is 4 times more than that in Calgary
- The rate increased 2-folds, from 3 per 100,000 population in 1990 (reference) to 9 per 100,000 in 2010

Why Do We Need Logistic Regression?



- LBW was reported high in our region. Is it true?
- What are the factors that contribute to a lower rate?
- Tell me what will be the LBW rate in next 20 years in our region.

Logistic Procedure



- Logistic regression models the relationship between a binary or ordinal response variable and one or more explanatory variables.
- **Logit (P_i) = $\log\{P_i/(1-P_i)\} = \alpha + \beta'X_i$**
where P_i = response probabilities to be modeled
 α = intercept parameter
 β = vector of slope parameters
 X_i = vector of explanatory variables

Performing a Logistic Regression



```
Proc logistic data = sample;
```

```
  Class mage_cat;
```

```
  Model      LBW = year mage_cat drug_yes drink_yes  
              smoke_9 smoke_yes /  
            lackfit outroc=roc2;
```

```
  Output out=Probs Predicted=Phat;
```

```
run;
```


Why Re-Coding Data to Binary?



- **While explanatory variables can be continuous and ordinal types, it is useful to recode them into binary sometime.**
- **When we want to use a fixed group as the reference, coding a variable into binary makes it easier to use and interpret.**
- **Teen age mother vs. mother 20-34 years or mother 35+ vs. mother 20-34 years, for instance.**

Re-Coding Data to Binary



data sample;

set &srcData;

Smoke_Yes=0; Smoke_9=0; Drug_Yes=0;
Drink_Yes=0; Mage_Teen=0; Mage_Old=0;

if EverSmoke = 1 then Smoke_Yes= 1;
if EverSmoke in (9, .) then Smoke_9 = 1;
if Street_Drug = 1 then Drug_Yes = 1;
if ALCOHOL_Preg= 1 then Drink_Yes = 1;

if Mage_cat= 2 then Mage_Old = 1;
if Mage_cat= 0 then Mage_Teen = 1; run;

Understanding Distribution – Proc Freq



```
Proc freq data=sample; table smoke_yes*LBW/nopercent  
nocol chisq cmh1;
```

```
Proc freq data=sample; table smoke_yes*(Mage_Teen  
Mage_Old mage_cat)/nopercent norow chisq cmh1;
```

```
Proc freq data=sample; table smoke_yes*(drug_yes  
drink_yes)/nopercent chisq;
```

```
run;
```



Run the Macros for Data Preparation

- `%inc '\\edm-go-a-file-3\user$\fu-lin.wang\methodology\Logistic Regression\recode_macro.sas';`
- `%recode;`

Distribution of Maternal Smoking and LBW



Low Birth Weight (< 2500 g)

Maternal
Smoking

	1 (Yes) n=68	0 (No) n=932
1 (n=237)	11.0%	89.0%
0 (n=763)	5.5%	94.5%

Odds Ratio (95%CI): 2.11 (1.27-3.53)



Use Class Statement for Odds Ratio



```
Proc logistic data = sample desc outest=betas2;
```

```
  Class mage_cat;
```

```
  Model LBW = year mage_cat drug_yes drink_yes  
             smoke_9 smoke_yes /  
             lackfit outroc=roc2;
```

```
  Output out=Probs_2 Predicted=Phat;
```

```
run;
```




Use Recoded Data for Odds Ratio



```
Proc logistic data = sample desc outest=betas3;
```

```
Model LBW = year mage_Teen Mage_Old  
drug_yes drink_yes  
smoke_9 smoke_yes /  
lackfit outroc=roc3;
```



```
Output out=Probs_3 Predicted=Phat;  
run;
```



Run the Macros for logistic regression

- `%inc '\\edm-go-a-file-3\user$\fu-lin.wang\methodology\Logistic Regression\logistic_macro.sas';`

Logistic Regression - Class Statement



Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
YEAR	0.951	0.850	1.064
mage_cat 0 vs 2	0.313	0.105	0.927
mage_cat 1 vs 2	0.405	0.226	0.724
Drug_Yes	0.494	0.102	2.381
Drink_Yes	2.047	0.797	5.258
Smoke_9	1.506	0.548	4.135
Smoke_Yes	2.384	1.338	4.247

Logistic Regression - Recoded Data



Effect	Odds Ratio Estimates		
	Point Estimate	95% Wald Confidence Limits	
YEAR	0.951	0.850	1.064
Mage_Teen	0.773	0.286	2.089
Mage_Old	2.472	1.382	4.421
Drug_Yes	0.494	0.102	2.381
Drink_Yes	2.047	0.797	5.258
Smoke_9	1.506	0.548	4.135
Smoke_Yes	2.384	1.338	4.247

Logistic Regression - Model Fitness



Model Fit Statistics

Criterion	Intercept Only	Intercept & Covariates
AIC	498.869	492.644
SC	503.777	531.906
-2 Log L	496.869	476.644

Identical for AIC, SC and -2 Log L
and other statistics between two models

Association of Predicted Probabilities and Observed Responses

Percent Concordant	63.4	Somers' D	0.306
Percent Discordant	32.9	Gamma	0.317
Percent Tied	3.7	Tau-a	0.039
Pairs	63376	c	0.653

Impact of Excluding Missing Smoking



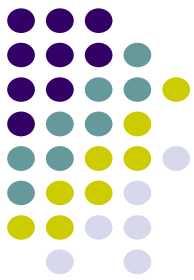
Effect	Odds Ratio Estimates		
	Point Estimate	95% Wald Confidence Limits	
YEAR	0.961	0.862	1.071
Mage_Teen	0.785	0.290	2.124
Mage_Old	2.439	1.365	4.357
Drug_Yes	0.487	0.101	2.349
Drink_Yes	2.047	0.797	5.260
Smoke_Yes	2.288	1.299	4.028

OR reduced from 2.38 to 2.29

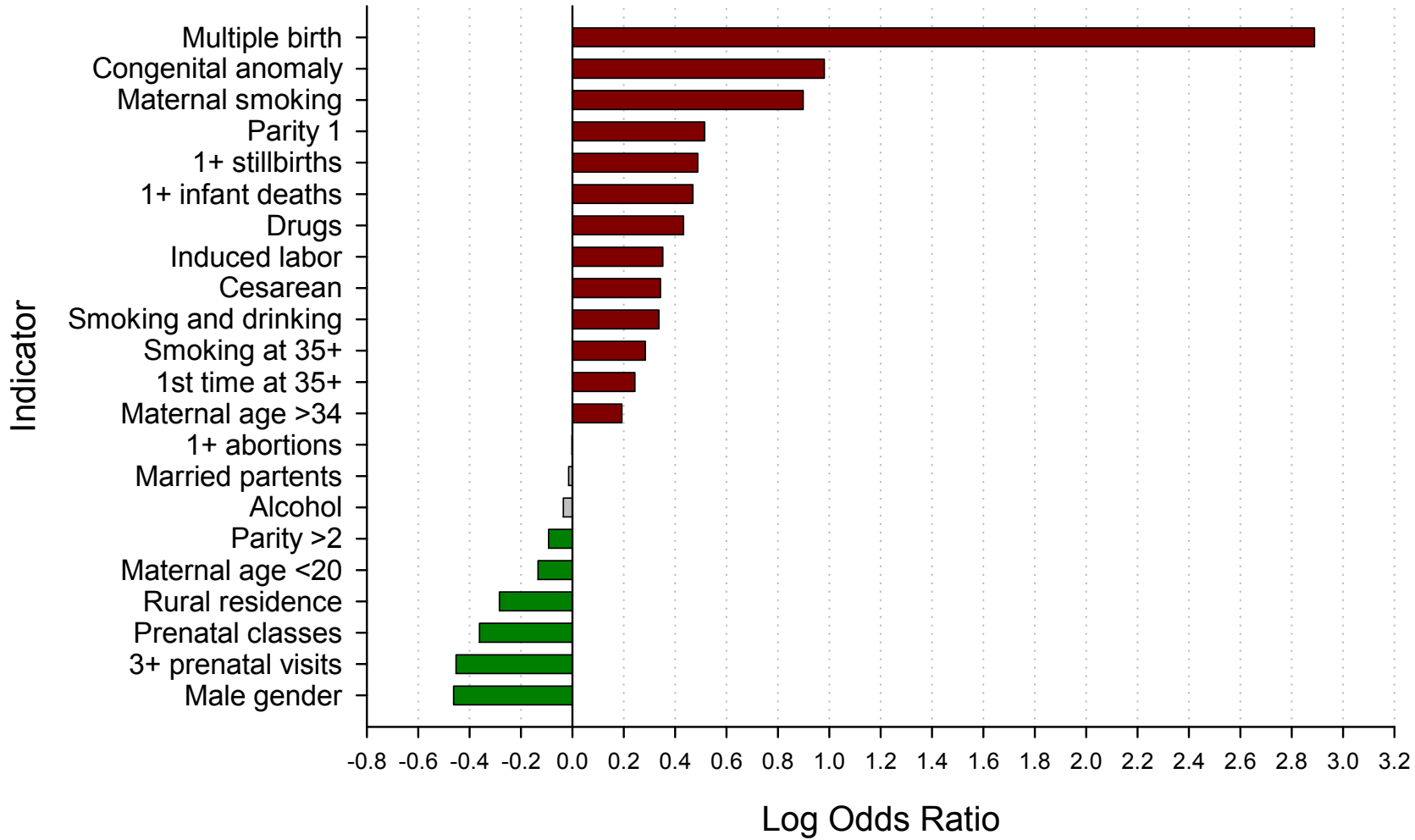


Interpretation of OR in Logistic Regression

- There is a moderate association between maternal smoking and LBW.
- Maternal age is associated with both LBW and maternal smoking.
- After controlling the confounding effect of maternal age (and other variables in the model), the risk for LBW among pregnant women who smoke is about 2.4 times of that non-smoking pregnant women.



Predictors of Low Birth Weight in Term Livebirths, Alberta, 1997 to 2004





Questions?

Pease contact:

Fu-lin.Wang@gov.ab.ca



Thank you!!