# the conundrum of interaction

## an epidemiologic perspective

Charles DiMaggio

Departments of Anesthesiology and Epidemiology
College of Physicians and Surgeons
Columbia University
New York, NY 10032

cjd11@columbia.edu

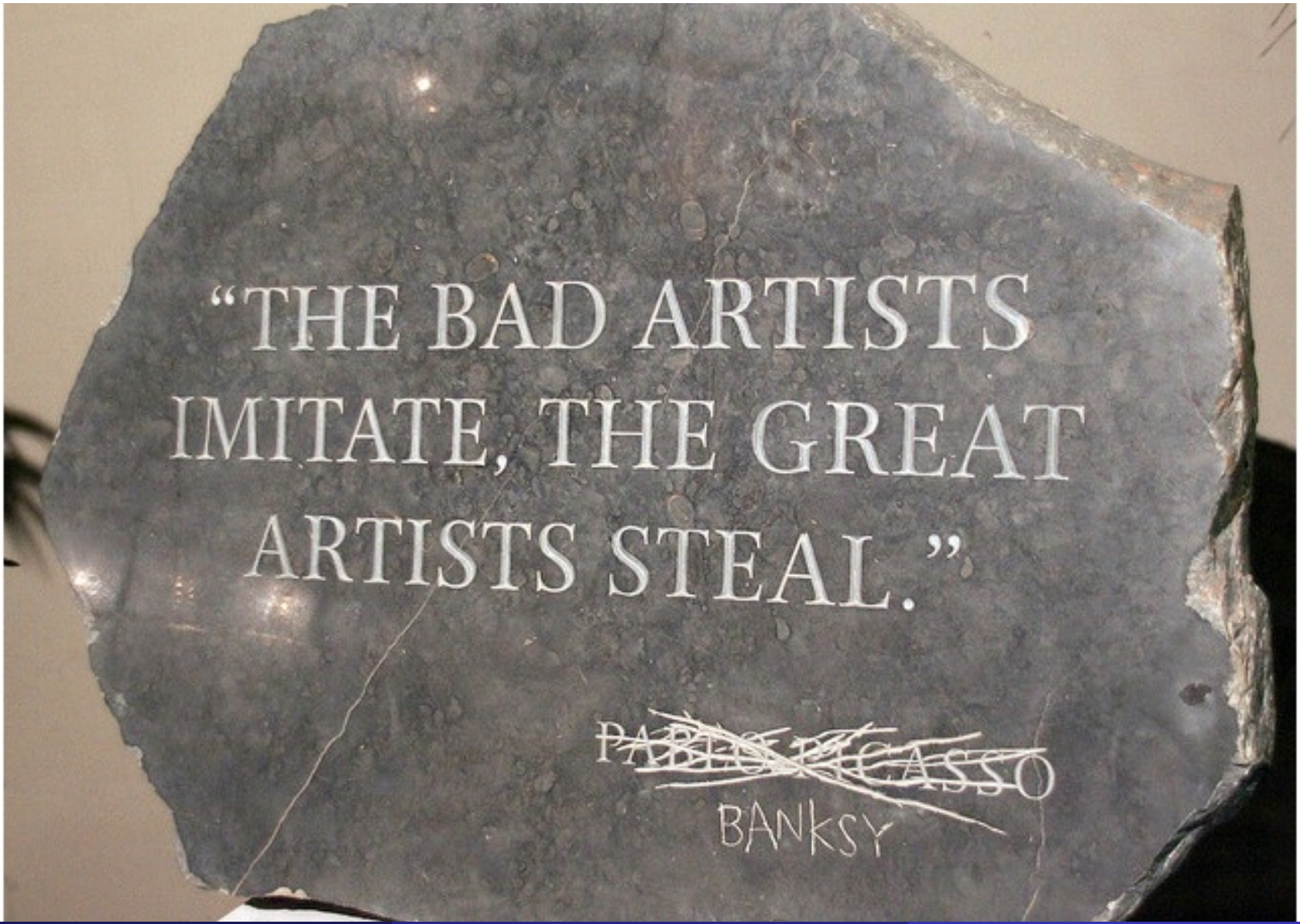
Boston-area
SAS users group
October 2013

# Outline


1. Review Classical Statistical Interaction

2. Interaction: An Epidemiological Perspective
   - The Conundrum of Interaction
   - Components and Causes
   - Biologic Interaction
   - Some Conclusions

# But first...



# Credit where credit is due...

- SAS Institute
  - Statistics and Regression
  - https://support.sas.com/edu/schedules.html?ctry=us&id=1321
- Ezra Susser, Sharon Schwartz, Alfredo Morabia
  - Psychiatric Epidemiology: Searching for the Causes of Mental Disorders
  - http://www.amazon.com/Psychiatric-Epidemiology-Searching-Disorders-Psychiatry/dp/0195101812
- Melanie Wall
  - Columbia University Departments of Psychiatry and Biostatistics
  - Are you looking for the right interactions?
- Kenneth Rothman
  - Epidemiology: An Introduction
  - http://www.amazon.com/Epidemiology-Introduction-Kenneth-J-Rothman/dp/0199754551

# Outline

1. Review Classical Statistical Interaction

2. Interaction: An Epidemiological Perspective
   - The Conundrum of Interaction
   - Components and Causes
   - Biologic Interaction
   - Some Conclusions

# drug dosage, disease and blood pressure

- 4 anti-hypertensive drug dosages in the setting of 3 diseases.
- outcome variable is systolic blood pressure.
- does combination of drug dosage and disease interacts to affect blood pressure in unexpected way?

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk}$$

where

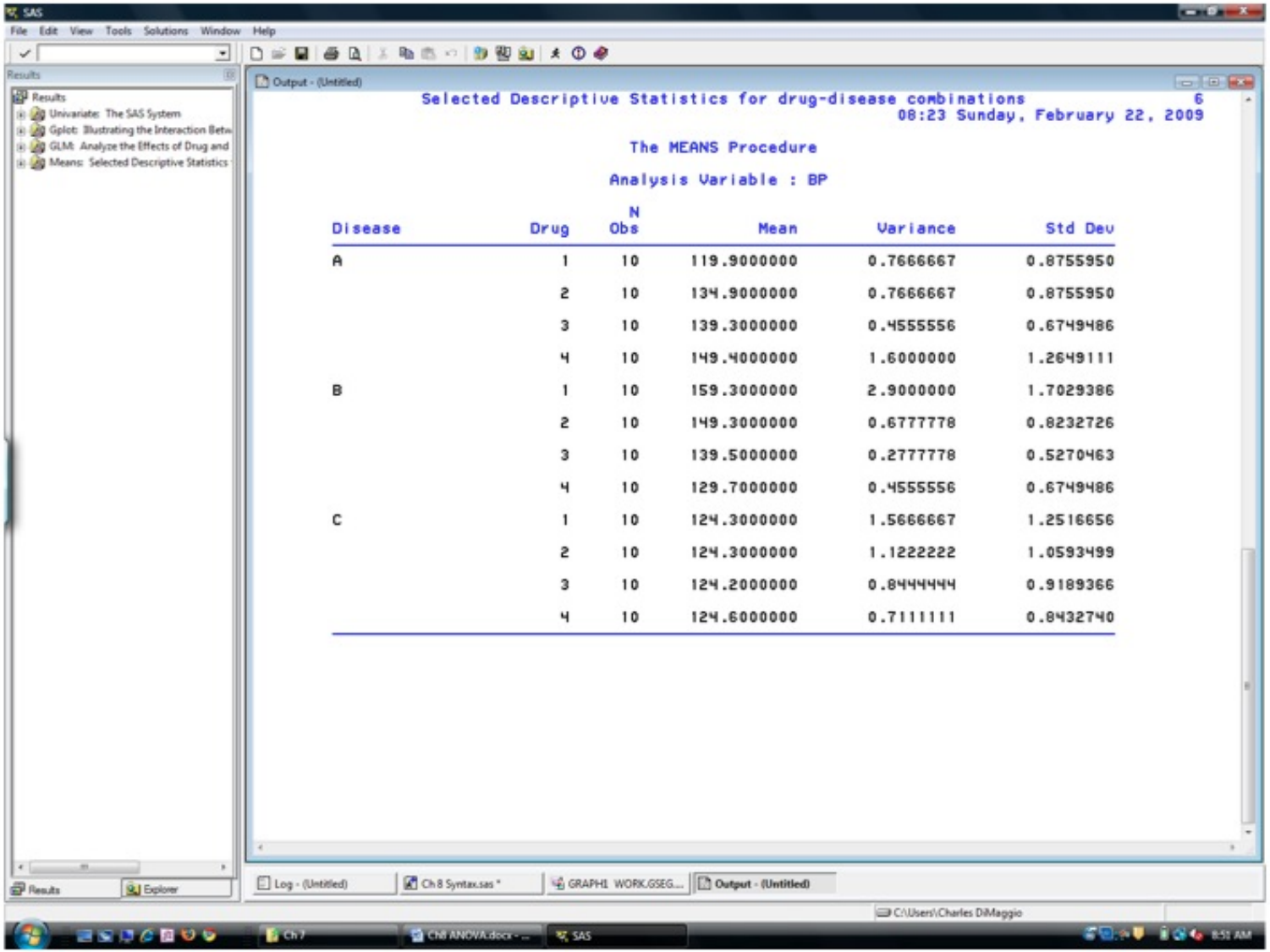
- $y_{ijk}$ is the observed blood pressure for each subject
- $\mu$ is the overall population mean blood pressure
- $\alpha_i$ is the effect of disease $i$
- $\beta_J$ is the effect of drug dosage $j$
- $(\alpha\beta)_{ij}$ is the *interaction* between disease $i$ and drug dose $j$, and
- $\epsilon_{ijk}$ is the residual or error term for each subject

# examine assumptions

- observations for each predictor combination
  - independent
  - identically, normally distributed
  - approximately equal variances (homoscedasticity)
- simple PROC MEANS of the 12 combinations of 4 drug doses and 3 diseases to begin

```
proc means data=bp_drug mean var std;
   class disease drug;
   var BP;
   title 'Selected Descriptive Statistics';
run;
```

# SAS output: ? interaction



Selected Descriptive Statistics for drug-disease combinations 6
08:23 Sunday, February 22, 2009

The MEANS Procedure

Analysis Variable : BP

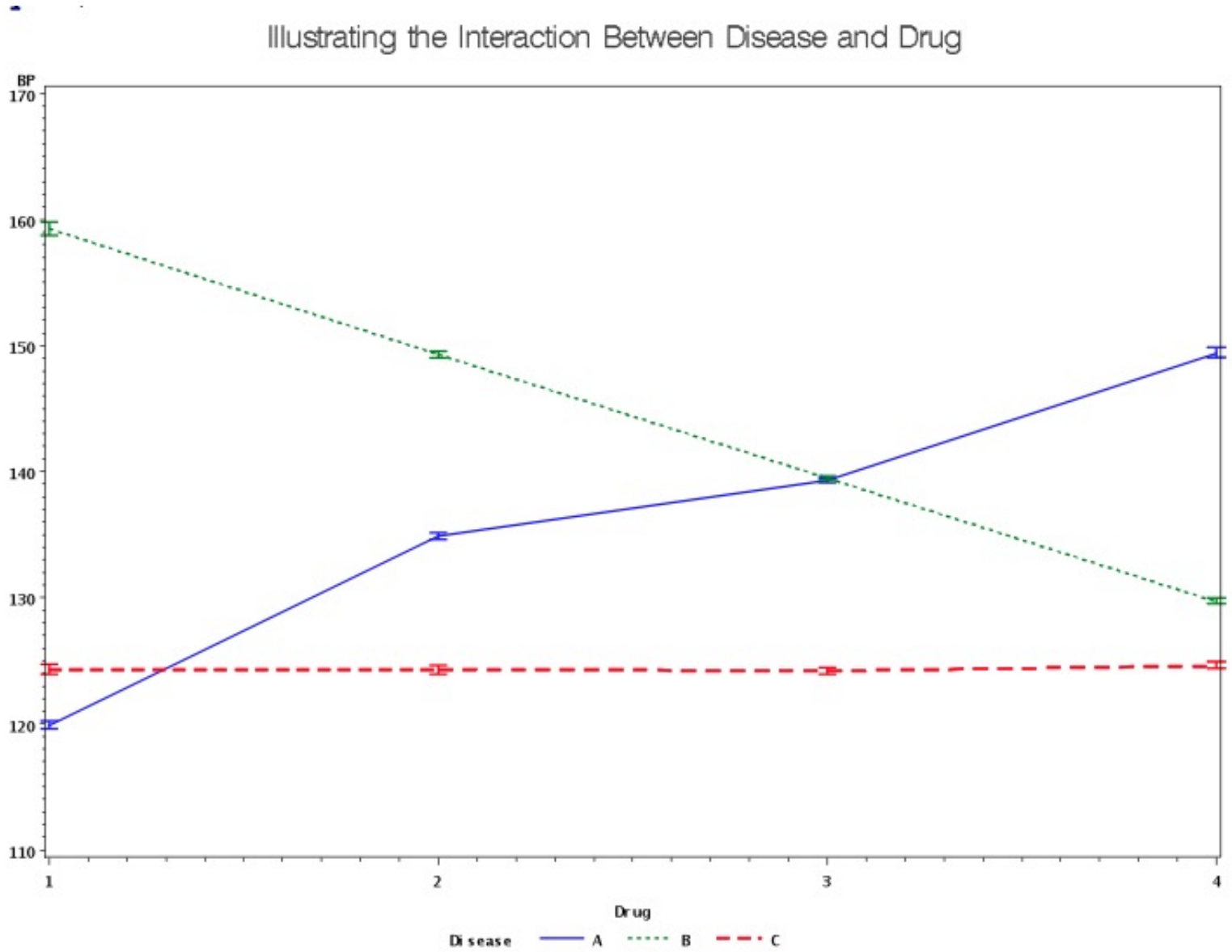
| Disease | Drug | N Obs | Mean | Variance | Std Dev |
|---|---|---|---|---|---|
| A | 1 | 10 | 119.9000000 | 0.7666667 | 0.8755950 |
| | 2 | 10 | 134.9000000 | 0.7666667 | 0.8755950 |
| | 3 | 10 | 139.3000000 | 0.4555556 | 0.6749486 |
| | 4 | 10 | 149.4000000 | 1.6000000 | 1.2649111 |
| B | 1 | 10 | 159.3000000 | 2.9000000 | 1.7029386 |
| | 2 | 10 | 149.3000000 | 0.6777778 | 0.8232726 |
| | 3 | 10 | 139.5000000 | 0.2777778 | 0.5270463 |
| | 4 | 10 | 129.7000000 | 0.4555556 | 0.6749486 |
| C | 1 | 10 | 124.3000000 | 1.5666667 | 1.2516656 |
| | 2 | 10 | 124.3000000 | 1.1222222 | 1.0593499 |
| | 3 | 10 | 124.2000000 | 0.8444444 | 0.9189366 |
| | 4 | 10 | 124.6000000 | 0.7111111 | 0.8432740 |

# means plot

- BP by dosage by disease

```
proc gplot data=bp_drug;
   symbol c=blue w=2 interpol=std1mtj line=1;
       /* interpolation method gives s.e. bars */
   symbol2 c=green w=2 interpol=std1mtj line=2;
   symbol3 c=red w=2 interpol=std1mtj line=3;
   plot BP*drug=disease; /* vertical by horizontal */
   title 'Illustrating Interaction Between Disease and Drug';
run;
quit;
```

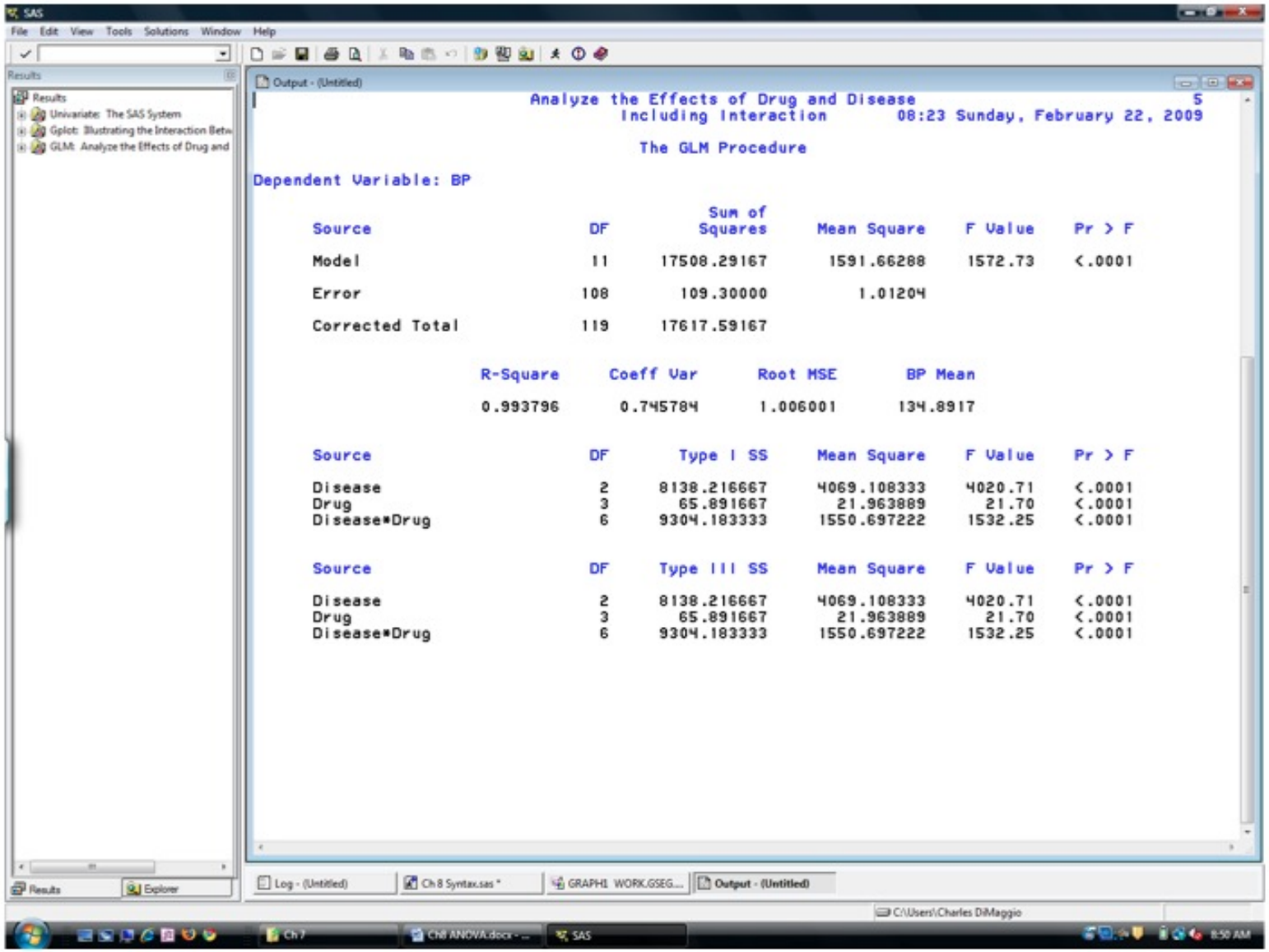
# SAS Output: interaction



# examine the interaction term with GLM

```
proc glm data=bp_drug;
   class disease drug;
   model BP=disease drug disease*drug;/*note intx term*/
   title 'Analyze the Effects of Drug and Disease';
   title2 'Including Interaction';
run;
quit;
```

# SAS output: "good" model, interaction



Analyze the Effects of Drug and Disease 5
Including Interaction 08:23 Sunday, February 22, 2009

The GLM Procedure

Dependent Variable: BP

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| Model | 11 | 17508.29167 | 1591.66288 | 1572.73 | <.0001 |
| Error | 108 | 109.30000 | 1.01204 | | |
| Corrected Total | 119 | 17617.59167 | | | |

| R-Square | Coeff Var | Root MSE | BP Mean |
|---|---|---|---|
| 0.993796 | 0.745784 | 1.006001 | 134.8917 |

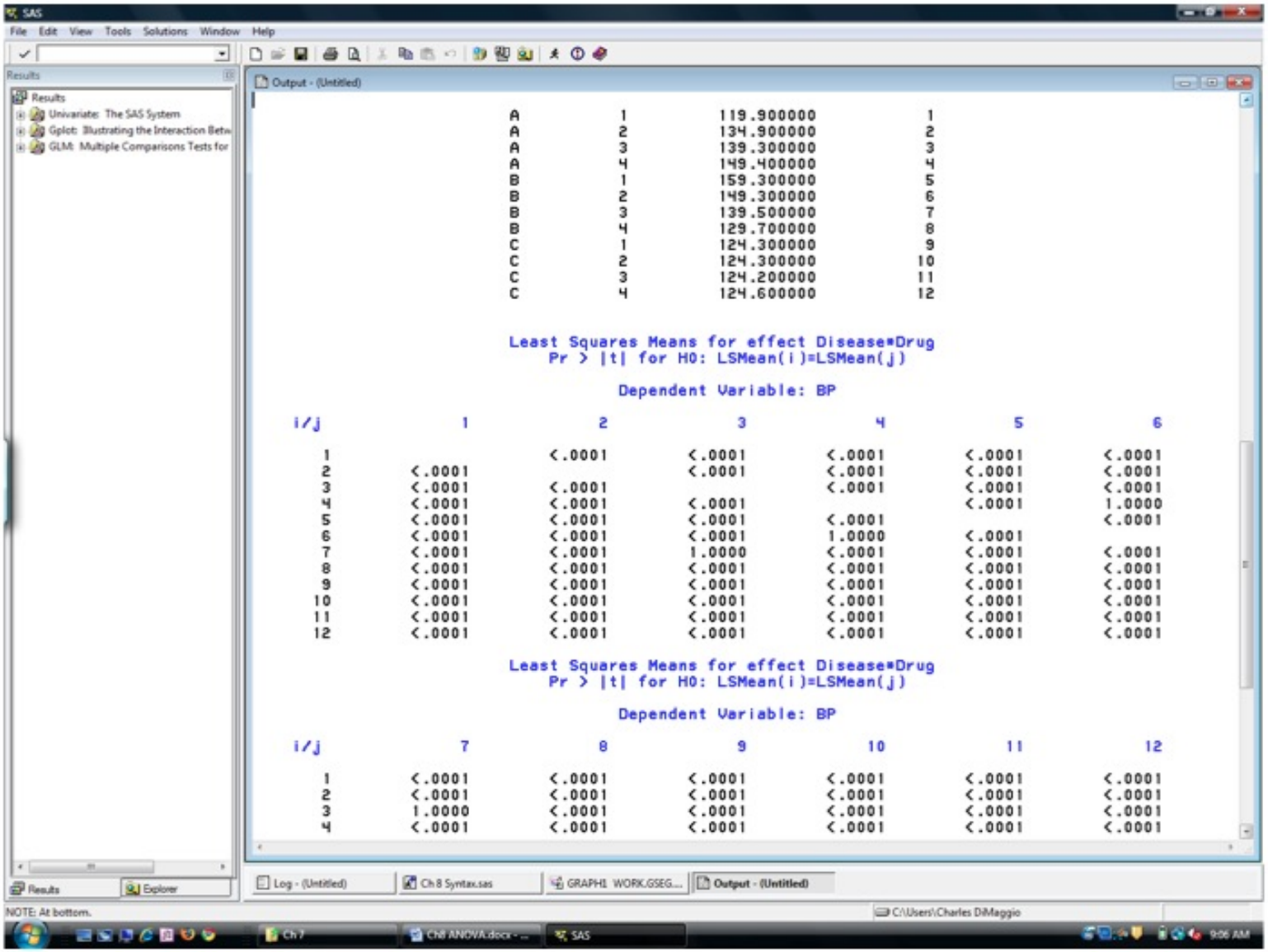
| Source | DF | Type I SS | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| Disease | 2 | 8138.216667 | 4069.108333 | 4020.71 | <.0001 |
| Drug | 3 | 65.891667 | 21.963889 | 21.70 | <.0001 |
| Disease*Drug | 6 | 9304.183333 | 1550.697222 | 1532.25 | <.0001 |

| Source | DF | Type III SS | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| Disease | 2 | 8138.216667 | 4069.108333 | 4020.71 | <.0001 |
| Drug | 3 | 65.891667 | 21.963889 | 21.70 | <.0001 |
| Disease*Drug | 6 | 9304.183333 | 1550.697222 | 1532.25 | <.0001 |

# LSMEANS

```
proc glm data=bp_drug;
   class disease drug;
   model BP=drug disease drug*disease;
   lsmeans disease*drug / adjust=tukey pdiff=all;
                /*looking at combinations*/
   title 'Multiple Comparisons Tests for Drug and Disease';
run;
quit;
```

# SAS output: lots of it



```
A          1          119.900000          1
A          2          134.900000          2
A          3          139.300000          3
A          4          149.400000          4
B          1          159.300000          5
B          2          149.300000          6
B          3          139.500000          7
B          4          129.700000          8
C          1          124.300000          9
C          2          124.300000         10
C          3          124.200000         11
C          4          124.600000         12


Least Squares Means for effect Disease*Drug
Pr > |t| for H0: LSMean(i)=LSMean(j)

Dependent Variable: BP

i/j          1          2          3          4          5          6

 1                  <.0001     <.0001     <.0001     <.0001     <.0001
 2      <.0001                 <.0001     <.0001     <.0001     <.0001
 3      <.0001      <.0001                <.0001     <.0001     <.0001
 4      <.0001      <.0001     <.0001                <.0001     1.0000
 5      <.0001      <.0001     <.0001     <.0001                <.0001
 6      <.0001      <.0001     <.0001     1.0000     <.0001
 7      <.0001      <.0001     1.0000     <.0001     <.0001     <.0001
 8      <.0001      <.0001     <.0001     <.0001     <.0001     <.0001
 9      <.0001      <.0001     <.0001     <.0001     <.0001     <.0001
10      <.0001      <.0001     <.0001     <.0001     <.0001     <.0001
11      <.0001      <.0001     <.0001     <.0001     <.0001     <.0001
12      <.0001      <.0001     <.0001     <.0001     <.0001     <.0001

Least Squares Means for effect Disease*Drug
Pr > |t| for H0: LSMean(i)=LSMean(j)

Dependent Variable: BP

i/j          7          8          9         10         11         12

 1      <.0001     <.0001     <.0001     <.0001     <.0001     <.0001
 2      <.0001     <.0001     <.0001     <.0001     <.0001     <.0001
 3      1.0000     <.0001     <.0001     <.0001     <.0001     <.0001
 4      <.0001     <.0001     <.0001     <.0001     <.0001     <.0001
```

# Outline

1 Review Classical Statistical Interaction

2 Interaction: An Epidemiological Perspective
- The Conundrum of Interaction
- Components and Causes
- Biologic Interaction
- Some Conclusions

# heterogeneity of effect

- measure of disease risk (either absolute or relative) behaves differently in the presence or absence of another variable
- but...depends on how we *measure* an exposure-disease association (!)
- can measure disease risk on either an absolute scale like risk differences, or on a relative scale, like risk ratios
- Interaction may be present on the additive measurement scale scale, but absent on the multiplicative scale (!)

# two kinds of interaction

- *Additive Interaction*
  - $RD_{1,2} \neq RD_1 + RD_2$
  - *absolute* measure differs from *sum* of individual absolute risk measures
- *Multiplicative interaction*
  - $RR_{1,2} \neq RR_1 \cdot RR_2$
  - *relative* measure of joint risk differs from the *product* of the individual ratio measures

# example: stress + genetics = depression interaction

| | | No Stress | Stress |
|---|---|---|---|
| disease rates: | No Genetics | 10 | 17 |
| | Genetics | 10 | 33 |

- additive interaction
  - stress alone $17 - 10 = 7$
  - genetics alone $10 - 10 = 0$
  - stress and genetics $7 + 0 = 7$ vs $33 - 10 = 23$
- multiplicative interaction
  - stress alone $17/10 = 1.7$
  - genetics alone $10/10 = 1$
  - stress and genetics $1x1.7 = 1.7$ vs $33/10 = 3.3$

# example: life events + intimacy ?= depression interaction

| disease rates: | No Life Event | Life Event |
|---|---|---|
| No Intimacy Problems | 1 | 3 |
| Intimacy Problems | 10 | 32 |

- additive interaction
  - intimacy alone $3 - 1 = 2$
  - events alone $10 - 1 = 9$
  - intimacy and events $2 + 9 = 11$ vs $32 - 1 = 31$
- *no* multiplicative interaction
  - intimacy alone $3/1 = 3$
  - events alone $10/1 = 10$
  - intimacy and events $3x10 = 30$ vs $32/1 = 32$

# the conundrum

the *absence* of interaction on one scale, rather than implying the absence of interaction on the other scale, is almost invariably accompanied by the *presence* of interaction on the other scale.

# Outline

1 Review Classical Statistical Interaction

2 Interaction: An Epidemiological Perspective
- The Conundrum of Interaction
- Components and Causes
- Biologic Interaction
- Some Conclusions

# causes and risk factors

- cause - something that makes a difference (subsequent event would not have occurred)
- risk factors - multiple antecedent components necessary for a cause
- *INUS* - *I*nsufficient but *N*ecessary components of *U*nnecessary but *S*ufficient causes
- causal relationships are inherently context dependent
  - strength of any risk factor is relative to and dependent on the presence or absence of its causal partners
  - neural tube defects, genetics and folate

# 5 potential causal relationships

1. Independent Risk Factor - Causes disease through a causal pathway different than that of the exposure of interest (a different causal mechanism)
2. Antecedent - Precedes the exposure
3. Confounder - An alternate risk factor for the disease, but associated with the exposure of interest
4. Mediator - (Also) A risk factor for the disease but (unlike a confounder) does not provide an alternate explanation for disease.
5. *Causal Partners* - Other component members of a causal mechanism that combine with exposure and can result in synergy or *interaction*

# Outline

1. Review Classical Statistical Interaction

2. Interaction: An Epidemiological Perspective
   - The Conundrum of Interaction
   - Components and Causes
   - Biologic Interaction
   - Some Conclusions

# "parallelism"

## Darroch (1997), Rothman and Greenland (1998)

2 risk factors (plus unknown), 4 paths to disease

- $R_{ABU}$ - the risk of disease from the interaction of A and B
- $R_{AU}$ - the risk of disease from A
- $R_{BU}$ - the risk of disease from B
- $R_U$ - the "background" experience where disease occurs in the absence of either A or B

# Biological Interaction is Additive

- does observed $R_{ABU}$ exceeds what we might expect if the two risks did not interact?
- subtract out $R_{AU}$ and $R_{BU}$
  - add back $R_U$ which is subtracted out twice
- if two risk factors biologically independent, then $R_{ABU} = R_{AU} + R_{BU} - R_U$
  - synergy - parallelism $= R_{ABU} - R_{AU} - R_{BU} + R_U$
- Risk Differences: $(RD_{AB} - RD_U) = (RD_A - RD_U) + (RD_B - RD_U)$
- Relative Risks: $(RR_{AB} - 1) = (RR_A - 1) + (RR_B - 1)$
- Any excess risk beyond these equalities is due to interaction

# example: smoking, asbestos and cancer

## risk differences

| | No Asbestos | Asbestos |
|---|---|---|
| No Smoking | 1 | 5 |
| Smoking | 10 | 50 |

- set up biological independence equality: $50 - 1? = (10 - 1) + (5 - 1)$
- $49 \neq 13$
- conclude that the smoking and asbestos interact to cause more cancer than would be expected if either were present alone
- $49 - 13$ or 36 of every 50 cases (72%) of cancer when both smoking and asbestos are present, are due to the interaction between them.

# example: smoking, asbestos and cancer

## relative risks

| | No Asbestos | Asbestos |
|---|---|---|
| No Smoking | 1 | 3.1 |
| Smoking | 6.9 | 13.6 |

- stest the equality $13.6 - 1? = (6.9 - 1) + (3.1 - 1)$
- $12.6 \neq 8$
- conclude (again) that there is interaction
- $(12.6 - 8)/13.6 = 4.6/8 = 34\%$ of the cases when both risk factors are present, is due to interaction
- on multiplicative scale, $3.1x6.9 = 21.4$ and since $13.6 < 21.4$ presence both risk factors results *less* risk than expected ...

# what about SAS?

## Brown Harris (1978) Vulnerability and the effect of stress on depression

```
proc logistic data = brownharris descending;
model depression (event= LAST)  = stress   vulnerability
    stress*vulnerability;
oddsratio stress / at(vulnerability = 0 1);
run;
```

| Parameter | Estimate | ... | Pr > ChiSq |
|---|---|---|---|
| | | ... | |
| Intercept | 2.1722 | ... | <.0001 |
| stress | 2.3869 | ... | 0.0026 |
| vulnerability | 1.3990 | ... | 0.0011 |
| stress*vulnerability | 0.2411 | ... | 0.8262 |

...

| Label | Estimate | 95% Confidence Limits | |
|---|---|---|---|
| stress at vulnerability=0 | 10.880 | 2.299 | 51.486 |
| stress at vulnerability=1 | 13.846 | 3.122 | 61.408 |

$exp(0.2411) = 1.27$
$13.9/10.9 = 1.27$
no (multiplicative) interaction

# coding interaction contrast in proc logistic

## from Melanie Wall

```
PROC NLMIXED DATA=brownharris;
odds = exp(b0 +b1*stress + b2*vulnerability
   + b3*stress*vulnerability); pi = odds/(1+odds);
MODEL depressn~BINARY(pi);
estimate 'p00' exp(b0)/(1+exp(b0));
estimate 'p01' exp(b0+b1)/(1+exp(b0+b1));
estimate 'p10' exp(b0+b2)/(1+exp(b0+b2));
estimate 'p11' exp(b0+b1+b2+b3)/(1+exp(b0+b1+b2+b3));
estimate'p11-p10'exp(b0+b1+b2+b3)/(1+exp(b0+b1+b2+b3))
   - exp(b0+b2)/(1+exp(b0+b2));
estimate 'p01-p00' exp(b0+b1)/(1+exp(b0+b1))
    - exp(b0)/(1+exp(b0));
estimate 'IC= interaction contrast = p11-p10 - p01 + p00'
exp(b0+b1+b2+b3)/(1+exp(b0+b1+ b2+b3))
    - exp(b0+b2)/(1+exp(b0+b2)) - exp(b0+b1)/(1+exp(b0+b1))
    + exp(b0)/(1+exp(b0));
```

# interaction contrast in brown harris data

| Label | Estimate | Error | .... |
|---|---|---|---|
| p00 | 0.1023 | 0.03230 | .... |
| p01 | 0.01036 | 0.007289 | .... |
| p10 | 0.3158 | 0.05332 | .... |
| p11 | 0.03226 | 0.02244 | .... |
| p11-p10 | -0.2835 | 0.05785 | .... |
| p01-p00 | -0.09191 | 0.03311 | .... |
| IC= interaction contrast | -0.1916 | 0.06666 | .... |

conclude additive interaction...

# Outline

1. Review Classical Statistical Interaction

2. Interaction: An Epidemiological Perspective
   - The Conundrum of Interaction
   - Components and Causes
   - Biologic Interaction
   - Some Conclusions

# conceptual implications

- *biological interaction* occurs at the individual level when the effect of one variable depends on the presence of another
  - *synergy or interaction* the underlying process for how *any* cause results in disease at the individual level
- *interaction and effect modification* are are population-level phenomena, ambiguity in how defined...
  - reserve terms *statistical interaction* or *effect modification* for how we try to *capture* this idea of synergy, which we often do through statistical measures of interaction
  - but caution: if looking for interaction, you may well find artifactual and perhaps even misleading results
- *epidemiological* approach to interaction is a priori , conceptual and informed by subject matter expertise
  - think about it during data collection and consider scientifically plausible interactions

# practical implications

- graphical assessment remains an informative initial approach
- when modeling, address biological interaction as an additive phenomenon
- categorize two potential interaction variables into factorial design
  - 11 represents presence of both, 10 and 01 presence of one or the other, and 00 the absence of both
- statistical significance of interaction terms remains useful analytically (though perhaps not biologically)
  - if interaction present, tests for individual factor effects might be misleading due to masking of effects by the interaction