Evaluation of Gene-Environment Interaction for Ovarian Cancer

Presented by:
Grace Choi, Jake Conway, and Niko Escanilla
Road Map

- Background
- Gap in Knowledge & Hypothesis
- Study Summary
- Summary Statistics
- Analyses & Results
- Conclusion
- Discussion

http://www.gpsmagazine.com/2008/02/garmin_nuvi_750_review.php#.U6ZNwxbRy68
Background: General

- Ovarian Cancer
  - Abnormal cell growth in the ovaries
    - Stromal
    - Germ Cell
    - Epithelial
- Treatment
  - Surgery
  - Chemotherapy
  - Radiation Therapy
- Statistics
- Key Mutations

http://www.sgosonline.org/assets/images/Scottsdale_2014/lab%202.jpg
http://www.ovariancancer.org/about/statistics/
Gap in Knowledge & Hypothesis

• Is there an interplay between the BRCA1/2 genes and known reproductive and gynecological risk factors for ovarian cancer?

• Gene-environment interactions work on an multiplicative scale in relation to ovarian cancer.
Study Summary

- Population: Israeli Women
  - March 1, 1994 - June 30, 1999
- Blood samples
- Test BRCA1/2 mutation
- Two controls per case
  - Selected from central population registry
  - Matched on age within 2 years
  - Area of birth and place
  - Length of residence

Study Summary

- Additional Data
  - Age, Ethnicity, Gynecological Surgery, Personal History of Breast Cancer, Family History of Breast or Ovarian Cancer, Parity, and Oral Contraceptive Use

- Environmental factors of specific interest
  - Parity (Dichotomized)
    - 1 child or less (0) VS. More than 1 child (1)
  - Oral Contraceptive Use (Dichotomized)
    - Use for 6 or less years (0) VS. Use for more than 6 years (1)

- Ethnicity: Ashkenazi
  - Large part of data comes from Ashkenazi population
  - Higher rate of BRCA1/2 mutation
## Summary Statistics

### Table 1: Characteristics of Women by BRCA1/2 Mutation Status

<table>
<thead>
<tr>
<th>Factors</th>
<th>Number of Women with no Mutation (%)</th>
<th>Number of Women with Mutation (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n = 1327</td>
<td>n = 252</td>
</tr>
<tr>
<td>1 Cancer</td>
<td>592(45)</td>
<td>240(95)</td>
</tr>
<tr>
<td>2 Oral Contraceptive (&gt;6yrs)</td>
<td>56(4)</td>
<td>16(6)</td>
</tr>
<tr>
<td>3 No Children</td>
<td>110(8)</td>
<td>21(8)</td>
</tr>
<tr>
<td>4 Age (&gt;50)</td>
<td>994(75)</td>
<td>186(74)</td>
</tr>
<tr>
<td>5 Ashkenazi</td>
<td>883(67)</td>
<td>219(87)</td>
</tr>
<tr>
<td>6 Personal History of Cancer</td>
<td>32(2)</td>
<td>36(14)</td>
</tr>
<tr>
<td>7 Undergone Gynecological Surgery</td>
<td>164(12)</td>
<td>19(8)</td>
</tr>
<tr>
<td>8 No Family History of Cancer</td>
<td>1,199(90)</td>
<td>192(76)</td>
</tr>
</tbody>
</table>

### Table 2: Characteristics of Women by Ovarian Cancer Status

<table>
<thead>
<tr>
<th>Factors</th>
<th>Number of Women with No Cancer (%)</th>
<th>Number of Women with Cancer (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n = 747</td>
<td>n = 832</td>
</tr>
<tr>
<td>1 BRCA1/2 Mutation</td>
<td>12(2)</td>
<td>240(29)</td>
</tr>
<tr>
<td>2 Oral Contraceptive (&gt;6yrs)</td>
<td>41(5)</td>
<td>31(4)</td>
</tr>
<tr>
<td>3 No Children</td>
<td>43(6)</td>
<td>88(11)</td>
</tr>
<tr>
<td>4 Age (&gt;50)</td>
<td>542(73)</td>
<td>638(77)</td>
</tr>
<tr>
<td>5 Ashkenazi</td>
<td>509(68)</td>
<td>593(71)</td>
</tr>
<tr>
<td>6 Personal History of Cancer</td>
<td>14(2)</td>
<td>54(6)</td>
</tr>
<tr>
<td>7 Undergone Gynecological Surgery</td>
<td>108(14)</td>
<td>75(9)</td>
</tr>
<tr>
<td>8 No Family History of Cancer</td>
<td>683(91)</td>
<td>708(85)</td>
</tr>
</tbody>
</table>
Methods for Analyses: Case-Control Design

- Type of observational study

- Compare patients with disease (case) vs no disease (control)

- Retrospective
  
  - Compare frequency of exposure to a risk factor present in each group

  - Help determine relationship between risk factor and disease
Analysis: Standard Logistic Regression

- Standard Logistic Regression
- Looks at the effects of covariates on outcome
- Binary or dichotomous outcome
- Odds Ratio
  - “How much more likely (or unlikely) it is to be present with y=1 than y=0”
Results:
Standard Logistic Regression

Table 6: Standard Logistic Regression Results Table for Ovarian Cancer

<table>
<thead>
<tr>
<th>Factors</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>p-Value</th>
<th>Odds Ratio</th>
<th>95%CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (Intercept)</td>
<td>-0.608</td>
<td>0.229</td>
<td>7.973e-03</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 BRCA1/2</td>
<td>3.153</td>
<td>0.305</td>
<td>4.934e-25</td>
<td>23.417</td>
<td>(12.876-42.587)</td>
</tr>
<tr>
<td>3 Oral Contraceptive Use</td>
<td>-0.590</td>
<td>0.289</td>
<td>4.162e-02</td>
<td>0.555</td>
<td>(0.314-0.978)</td>
</tr>
<tr>
<td>4 Parity</td>
<td>-0.035</td>
<td>0.030</td>
<td>2.477e-01</td>
<td>0.966</td>
<td>(0.910-1.025)</td>
</tr>
<tr>
<td>5 Age Group</td>
<td>0.114</td>
<td>0.046</td>
<td>1.335e-02</td>
<td>1.121</td>
<td>(1.024-1.228)</td>
</tr>
<tr>
<td>6 Ethnicity</td>
<td>0.085</td>
<td>0.097</td>
<td>3.813e-01</td>
<td>1.089</td>
<td>(0.900-1.317)</td>
</tr>
<tr>
<td>7 Cancer History</td>
<td>0.564</td>
<td>0.348</td>
<td>1.052e-01</td>
<td>1.758</td>
<td>(0.888-3.481)</td>
</tr>
<tr>
<td>8 History of Gynecological Surgery</td>
<td>-0.244</td>
<td>0.087</td>
<td>5.082e-03</td>
<td>0.784</td>
<td>(0.661-0.929)</td>
</tr>
<tr>
<td>9 Family History of Cancer</td>
<td>0.323</td>
<td>0.135</td>
<td>1.100e-02</td>
<td>1.381</td>
<td>(1.059-1.801)</td>
</tr>
</tbody>
</table>
Results:
Standard Logistic Regression

Table 7: Standard Logistic Regression on the Interaction Between BRCA1/2 Gene Mutation Given Environmental Factors and that Subjects have Cancer

<table>
<thead>
<tr>
<th>Factors</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>p-Value</th>
<th>Odds Ratio</th>
<th>95%CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-0.616</td>
<td>0.229</td>
<td>7.230e-03</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oral Contraceptive Use</td>
<td>-0.623</td>
<td>0.305</td>
<td>4.127e-02</td>
<td>0.537</td>
<td>(0.295-0.976)</td>
</tr>
<tr>
<td>BRCA1/2</td>
<td>3.622</td>
<td>0.670</td>
<td>6.290e-08</td>
<td>37.431</td>
<td>(10.076-139.044)</td>
</tr>
<tr>
<td>Parity</td>
<td>-0.032</td>
<td>0.030</td>
<td>2.969e-01</td>
<td>0.969</td>
<td>(0.913-1.028)</td>
</tr>
<tr>
<td>Age Group</td>
<td>0.114</td>
<td>0.046</td>
<td>1.334e-02</td>
<td>1.121</td>
<td>(1.024-1.228)</td>
</tr>
<tr>
<td>Ethnicity</td>
<td>0.085</td>
<td>0.097</td>
<td>3.796e-01</td>
<td>1.089</td>
<td>(0.900-1.317)</td>
</tr>
<tr>
<td>History of Gynecological Surgery</td>
<td>-0.243</td>
<td>0.087</td>
<td>5.140e-03</td>
<td>1.739</td>
<td>(0.661-0.930)</td>
</tr>
<tr>
<td>Cancer History</td>
<td>0.555</td>
<td>0.349</td>
<td>1.131e-01</td>
<td>0.784</td>
<td>(0.877-3.449)</td>
</tr>
<tr>
<td>Family History of Cancer</td>
<td>0.324</td>
<td>0.135</td>
<td>1.671e-02</td>
<td>1.382</td>
<td>(1.060-1.802)</td>
</tr>
<tr>
<td>Oral Contraceptive Use:BRCA1/2</td>
<td>0.472</td>
<td>1.130</td>
<td>6.759e-02</td>
<td>1.604</td>
<td>(0.175-14.681)</td>
</tr>
<tr>
<td>BRCA1/2:Parity</td>
<td>-0.198</td>
<td>0.219</td>
<td>3.655e-01</td>
<td>0.820</td>
<td>(0.534-1.260)</td>
</tr>
</tbody>
</table>
Methods for Analyses: Case-Only Design

- Alternative to case-control design
  - Controls considered to be a sample of the general population
- Used to estimate interaction effect
- Works under two assumptions
  - Rare disease
  - Independence between gene and environmental factor
Methods for Analyses: Case-Only Design

- Back to the two assumptions...

  - Rare disease
    - Case-only estimator is well known to be efficient even when data on unaffected individuals is available
  
  - G-E independence
    - Condition on additional covariates
      - Also condition on covariates that confound association between disease and the gene and/or environmental factor
Analysis:
Logistic Regression-Test for Independence

- Logistic Regression
  - Used to test possible independence
  - Controls only
Results:
Logistic Regression-Test for Independence

- Are the environmental factors independent of the BRCA1/2?
Analysis: Case-Only Estimator

- Standard Logistic Regression
- Will only take into account ONLY the cases in our study
- Why can we do this?
Efficiency of the Case-Only Estimator

Goal:
To show that the case-only estimator is a more efficient method to determine interaction effect than a case-control estimator.

Variables:
- OR = Odds Ratio
- Y = Disease outcome [Controls (Y = 0) or Cases (Y = 1)]
- G = BRCA1/2 Gene Mutation
- E = Environmental Factors (Parity or Oral Contraceptive Use)
- GE = Gene-Environment Interaction
- C = Confounders

Consider the use of a logistic regression for analysis in a case-control design, we say that:
\[
\ln(\text{OR}_Y \mid \text{C,E,G,GE}) = \ln(\text{OR}_{Y=1} \mid \text{C,E,G,GE}) - \ln(\text{OR}_{Y=0} \mid \text{C,E,G,GE})
\]

Under the rare disease assumption,
\[
\ln(\text{OR}_{Y=0} \mid \text{C,E,G,GE}) = \ln(\text{OR}_{\text{population}})
\]

Taking into account the independence assumption,
\[
\ln(\text{OR}_{\text{population}}) = 0
\]

Thus,
\[
\ln(\text{OR}_Y \mid \text{C,E,G,GE}) = \ln(\text{OR}_{Y=1} \mid \text{C,E,G,GE}) - \ln(\text{OR}_{Y=0} \mid \text{C,E,G,GE}) \\
= \ln(\text{OR}_{Y=1} \mid \text{C,E,G,GE})
\]

Case-Control Model:
\[
Y = \beta_0 + \beta_1 G + \beta_2 E + \beta_3 GE + \beta_p C_p
\]

Case-Only Model:
\[
G = \alpha_0 + \alpha_1 E_p + \alpha_2 E_o + \alpha_p C_p
\]
Efficiency of the Case-Only Estimator

To put it in words,

- The case-only estimator is less variable than the case-control estimator
  - Why?
    - Two key assumptions allow us to NOT take into account extra variability from the log odds ratio of the controls
      \[ \ln(\text{OR}_{Y=0 | C,E,G,GE}) \]

(Note: A detailed proof can be discussed during the lunch break!)
Results:

Case-Only Logistic Regression

Table 9: Case-Only Logistic Regression Assessment on the Interaction Between BRCA1/2 Gene Mutation Given Environmental Factors (Parity Dichotomized) and that Subjects have Cancer

<table>
<thead>
<tr>
<th>Factors</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>p-Value</th>
<th>Odds Ratio</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>0.581</td>
<td>0.374</td>
<td>1.206e-01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oral Contraceptive Use</td>
<td>1.047</td>
<td>0.403</td>
<td>9.434e-03</td>
<td>2.850</td>
<td>(1.292-6.284)</td>
</tr>
<tr>
<td>Parity*</td>
<td>0.465</td>
<td>0.200</td>
<td>1.989e-02</td>
<td>1.592</td>
<td>(1.076-2.354)</td>
</tr>
<tr>
<td>Age Group</td>
<td>-0.248</td>
<td>0.073</td>
<td>6.360e-04</td>
<td>0.781</td>
<td>(0.677-0.899)</td>
</tr>
<tr>
<td>Ethnicity</td>
<td>-1.000</td>
<td>0.176</td>
<td>1.310e-08</td>
<td>0.368</td>
<td>(0.261-0.519)</td>
</tr>
<tr>
<td>Cancer History</td>
<td>1.673</td>
<td>0.316</td>
<td>1.210e-07</td>
<td>5.327</td>
<td>(2.867-9.897)</td>
</tr>
<tr>
<td>History of Gynecological Surgery</td>
<td>-0.196</td>
<td>0.152</td>
<td>1.972e-01</td>
<td>0.822</td>
<td>(0.609-1.107)</td>
</tr>
<tr>
<td>Family History of Cancer</td>
<td>0.645</td>
<td>0.141</td>
<td>5.050e-06</td>
<td>1.907</td>
<td>(1.445-2.516)</td>
</tr>
</tbody>
</table>

* = Parity Dichotomized
Conclusion

• Objective?
  • Test for an interaction
    • BRCA1/2 vs. Oral Contraceptive Use
    • BRCA1/2 vs. Parity
  • Examine the effect that these factors have on ovarian cancer
Conclusion

Case-Control Estimator

Standard Logistic Regression (No Interactions)

- Oral Contraceptive Use Significant at $\alpha = .05$ level?

Standard Logistic Regression (Interactions)

- Parity Significant at $\alpha = .05$ level?
- Oral Contraceptive Use Significant at $\alpha = .05$ level?
- BRCA1/2 & Oral Contraceptive Use Interaction Significant at $\alpha = .05$ level?
- BRCA1/2 & Parity Interaction Significant at $\alpha = .05$ level?
Conclusion

• Back to the drawing board...

• How do we test for an interaction?

  • We need to check off two important assumptions

    • Assumption 1: Disease under investigation is rare

    • Assumption 2: Independence between gene and each environmental factor

• Assumption 1

  • Ovarian cancer is, in fact, known to be a rare disease
Conclusion

How do we check for an independence assumption?

Test for Conditional Probabilities with a Logistic Regression (Controls Only)

Is $\alpha > 0.05$?

YES

Gene and Environmental Factors are INDEPENDENT!

Case-Only Estimator

Case-Only Logistic Regression

Oral Contraceptive Use Significant at $\alpha = 0.05$ level?

Parity Significant at $\alpha = 0.05$ level?
Conclusion

- Efficiency of case-only design vs. case-control design
- Use of case-only design
  - Determine that there is indeed an interaction between gene and environment for ovarian cancer
- Oral Contraceptive Use & Parity
  - Not as effective in preventing ovarian cancer with those that have mutation as opposed to those that do not have the mutation
Discussion

- Shortcomings
  - Collection of data
    - Interview bias
  - Recall Bias
  - Generalizability
  - Statistical Power
    - Decreases because of binary/dichotomous variables
Discussion

- Future Studies
  - Run analysis on specific parts(s) of BRCA1/2 gene that are mutated
  - Use continuous data for analysis
  - Increase interests of environmental factors in relation to BRCA1/2 gene
  - Test a different population to increase external validity
Acknowledgements

- We would like to recognize the following people for the support and encouragement throughout the SPQS:
  - Dr. Eric Tchetgen Tchetgen
  - Mr. Caleb Miles
  - Dr. Rebecca Betensky
  - Ms. Tonia Smith
  - Ms. Heather Mattie
  - Ms. Ellie Murray
  - Our fellow SPQS participants
Thank You

• Questions?


