



**National Industry Liaison Group
August 6, 2006 (1:00 - 3:00 p.m.)**

**The Mantel-Haenszel and Breslow-Day
Defined: Combining Adverse Impact
Analyses Across Multiple Events**

Presented by Biddle Consulting Group, Inc.

The following presentation is not to be construed as legal advice. For specific legal advice please consult your corporate counsel or a labor attorney.



Contact Information

Dan A. Biddle, Ph.D.

CEO, Biddle Consulting Group, Inc.

193 Blue Ravine, Ste. 270

Folsom, CA 95630

1-800-999-0438

www.biddle.com

Email: dan@biddle.com



Overview of Biddle Consulting Group, Inc. (BCG)

- Since 1974
- Over 200+ cases in the EEO/AA area (both plaintiff and defense cases)
- Pioneers in the EEO/AA field
- Administrative Skills Testing (OPAC)
- 911 Dispatcher Testing (Criticall)
- AAP Software and Services
- EEO Litigation Assistance (expert consulting and witness services)
- Adverse Impact and Test Validation, 2nd Ed.



Agenda

- Adverse Impact Overview
- Mantel-Haenszel
- Breslow-Day
- Mantel-Haenszel / Breslow-Day Case Studies



Adverse Impact Overview



**Adverse Impact:
*in the beginning . . .***

80%? 70%? 90%?



Adverse Impact — What's Tried and True:

- Statistical Significance
- Practical Significance



Adverse Impact — What's Tried and True:

- Statistical Significance:
 - 5%
 - 0.05
 - 1 chance in 20
 - 2.0 Standard Deviations (*actually 1.96*)



Two Types of Adverse Impact

SELECTION RATE COMPARISON

- 2 X 2 Table Comparison
- Evaluates hires, promotions, terminations
- “Hypergeometric”

Men Pass	Women Pass
Men Fail	Women Fail

AVAILABILITY COMPARISON

- Utilization Analysis
- Single Group Test
- “Binomial”

Availability %
Women
Total



When Does Adverse Impact Result in “Disparate Impact Discrimination”?

SELECTION RATE COMPARISON

- 2 X 2 Table Comparison
- Evaluates hires, promotions, terminations
- “Hypergeometric”

Statistically Significant Result



No Job Relatedness / Validity



Disparate Impact Discrimination

AVAILABILITY COMPARISON

- Utilization Analysis
- Single Group Test
- “Binomial”
- **See p. 58955 of App Regs**

Statistically Significant Result



6 “Possible Ingredients”



“Adverse Inference” or Evidence for Disparate Treatment Cases



Adverse Impact: Selection Rate Comparison (1991 CRA) & UGESP

Amends Section 703 of the 1964 Civil Rights Act (Title VII) (k) (1) (A). An unlawful employment practice based on disparate impact is established under this title only if:

- A(i) a complaining party demonstrates that a respondent uses a particular employment practice that causes a disparate impact on the basis of race, color, religion, sex, or national origin, and the respondent fails to demonstrate that the challenged practice is **job-related for the position in question and consistent with business necessity**; **OR**,
- A(ii) the complaining party makes the demonstration described in subparagraph (C) with respect to an **alternate employment practice**, and the respondent refuses to adopt such alternative employment practice.



Adverse Impact—Availability Comparison

- Statistical Significance + Six Possible Ingredients for “Adverse Inference” or Disparate Treatment
 - #1 **Failure to keep applicant records** (sometimes referred to as an “adverse inference”—see 4D of the Guidelines)
 - #2 **Failure to run/keep adverse impact analyses** on the selection or promotional processes (also an “adverse inference”—see 4D of the Guidelines)
 - #3 **Discriminatory recruiting practice** (e.g., Hazelwood School District v. United States)
 - #4 **Discriminatory reputation “chilled” or “discouraged”** certain group members from applying
 - #5 **Promoting employees through “appointment only”** process (rather than conducting promotional processes)
 - #6 **Invalid “Basic Qualifications”**



Adverse Impact—Availability Comparison Statistical Significance + Six Possible Ingredients for “Adverse Inference” or Disparate Treatment

- Unless one or more of the 6 ingredients exist, statistically significant underutilization should not be directly equated with discrimination
- Several other factors can sometimes explain underutilization:
 - Job interest
 - Occupational qualifications
 - Labor trends
 - Traditional roles (e.g., engineering vs. clerical)
- Unless one of the “6 ingredients” exist, a specific practice, procedure, or test will need to be identified that caused the adverse impact (using the 80% test, statistical significance tests, and practical significance tests). The only exception is if the agency’s practices cannot be “separated for analysis purposes” (see 1991 Civil Rights Act)



Adverse Impact and Statistical Power

- Statistical **significance** is the point at which differences become large enough that one can claim a “trend” or a “meaningful difference” exists.
- Statistical **power** refers to the ability to see those trends if, in fact, they do exist.
- Statistical power is directly related to **effect size** and **sample size**:
 - **Effect size**: the size of the difference (e.g., in selection rates) between two groups . . . if it’s a large difference then only a small sample size is needed to detect it
 - **Sample size**: a large sample size can detect small differences . . . small sample sizes can only detect large differences



Adverse Impact and Statistical Power

	Men		Women	
	(#)	(%)	(#)	(%)
Pass	100	50.0	90	45.0
Fail	100	50.0	110	55.0
80% Test ⁽¹⁾	Company A		0.90	
Stat. Test-EXACT ⁽²⁾			0.368	
Pract. Test (to Exact Test) ⁽³⁾			N/A	
Stat. Test-ESTIMATED ⁽²⁾			0.317	

	Men		Women	
	(#)	(%)	(#)	(%)
Pass	200	50.0	180	45.0
Fail	200	50.0	220	55.0
80% Test ⁽¹⁾	Company B		0.90	
Stat. Test-EXACT ⁽²⁾			0.179	
Pract. Test (to Exact Test) ⁽³⁾			N/A	
Stat. Test-ESTIMATED ⁽²⁾			0.157	

	Men		Women	
	(#)	(%)	(#)	(%)
Pass	300	50.0	270	45.0
Fail	300	50.0	330	55.0
80% Test ⁽¹⁾	Company C		0.90	
Stat. Test-EXACT ⁽²⁾			0.094	
Pract. Test (to Exact Test) ⁽³⁾			N/A	
Stat. Test-ESTIMATED ⁽²⁾			0.083	

	Men		Women	
	(#)	(%)	(#)	(%)
Pass	400	50.0	360	45.0
Fail	400	50.0	440	55.0
80% Test ⁽¹⁾	Company D		0.90	
Stat. Test-EXACT ⁽²⁾			0.051	
Pract. Test (to Exact Test) ⁽³⁾			N/A	
Stat. Test-ESTIMATED ⁽²⁾			0.045	



Adverse Impact and Statistical Power

- Plaintiff groups realize that statistical significance requires sufficient statistical power, which is why they will typically request data from across multiple years, requisitions, ranks, hiring processes, etc.
- However, simply aggregating all applicants and all hires across strata (as is typically done), *sometimes can result in misleading findings.*



Adverse Impact Over Several Years: Simpson's Paradox

SIMPSON'S PARADOX EXAMPLE				
Testing Year	Group	# Applicants	# Selected	Selection Rate %
2004 Test	Men	400	200	50.0%
	Women	100	50	50.0%
2005 Test	Men	100	20	20.0%
	Women	100	20	20.0%
2004 + 2005 Tests Combined	Men	500	220	44.0%
	Women	200	70	35.0%

- **Standard Deviations:**
 - **Mantel-Haenszel: Z = .024 (NOT significant)**
 - **Fisher Exact Test: Z = 2.12 (Significant)**



Biddle Consulting Group, Inc.

Over 30 Years of Human Resource Consulting Experience

Mantel-Haenszel



Mantel-Haenszel (MH) Defined

- In the context of Adverse Impact Analyses for selection rate comparison analyses (UGESP 4D), the MH is:
 - a statistical tool that allows a researcher to evaluate the adverse impact of a practice, procedure, or test **when used over time**
- The MH is a useful tool for evaluating whether the employer has a “pattern and practice” that is possibly discriminatory



Mantel-Haenszel (MH) Defined

- Statistically speaking:
 - The MH functions like a sample-size weighted chi-square test, which apportions more weight to the events (e.g., years) with larger samples
 - This is because smaller samples have larger “standard errors” and larger samples are more reliable



Mantel-Haenszel (MH) Defined

- The MH outputs a probability value which can be interpreted just like a chi-square test, standard deviation test, Fisher Exact, or many other similar tests
- The probability value (“p-value”) can be converted to a “Standard Deviation” or “Z” value:
 - P-value .10 = 1.65 Z
 - P-value .05 = 1.96 Z
 - P-value .01 = 2.58 Z
 - P-value .001 = 3.29 Z



Biddle Consulting Group, Inc.

Over 30 Years of Human Resource Consulting Experience

Breslow-Day



Breslow-Day (BD) Defined

- In the context of Adverse Impact Analyses for selection rate comparison analyses (UGESP 4D), the BD is:
 - a statistical tool that allows a researcher to determine whether multiple events (e.g., years) can be combined into a single MH analysis (or whether they need to be broken into separate analyses)



Breslow-Day (BD) Defined

- Practically speaking, the BD test...
 - gives a “green light” or “red light” to combining several events (e.g., years) together into a combined MH analysis
 - shows whether there are significant “flip-flops” in the trend of events



Breslow-Day (BD) Defined

- Statistically speaking, the BD test...
 - Can be interpreted just like the MH, Fisher Exact, Chi-Square, etc.:
 - P-values less than .05 (or a Z of more than 1.96) indicate that the events **CANNOT** be combined together in an MH analysis
 - Shows which events can be included in an overall MH analysis



Single Event v. Multiple Event Analyses

Selection Rate Analysis

Men Pass	Women Pass
Men Fail	Women Fail

+

Men Pass	Women Pass
Men Fail	Women Fail

+

Men Pass	Women Pass
Men Fail	Women Fail

= MH

Availability Analysis

Availability %
Women
Total

+

Availability %
Women
Total

+

Availability %
Women
Total

= Multiple-Event Binomial Analysis



Mantel-Haenszel / Breslow-Day Case Studies



Conducting Adverse Impact Analyses

Comparison	Single Event	Multiple Events
	For analyzing <u>single data sets</u> from practices, procedures, or tests	For analyzing <u>combined data sets</u> (i.e., multiple tests, years, etc.)
<u>Two Groups'</u> Selection Rates on a Test	Selection Rate Comparison	Combined Data-Sel. Rates
<u>One Group's</u> Representation to Availability	Availability Comparison	Combined Data-Availability
	Load Training Data	
	Load Sample Court Cases	



Selection Rate Comparison: Single Event (Fisher Exact Test or "Chi-Square" Style Test)

Selection Rate Comparison

ENTER DATA ONLY IN YELLOW CELLS

Calculate Exact Test

Clear Data

	Men		Women		White		Black		Hispanic		Asian		Native Amer.		Total Min.	
	(#)	(%)	(#)	(%)	(#)	(%)	(#)	(%)	(#)	(%)	(#)	(%)	(#)	(%)	(#)	(%)
Pass					455	89.6	50	78.1							50	78.1
Fail					53	10.4	14	21.9							14	21.9
80% Test ⁽¹⁾									0.87						0.87	
Stat. Test-EXACT ⁽²⁾																
Pract. Test (to Exact Test) ⁽³⁾																
Stat. Test-ESTIMATED ⁽²⁾									0.007						0.007	

Interpretation of EXACT Statistical Test ⁽⁴⁾:

Likelihood (One Chance In):

Probability as Std. Deviations:

Interpretation of ESTIMATED Statistical Test ⁽⁴⁾:

Likelihood (One Chance In):

Probability as Std. Deviations:

137					137
2.682					2.682



Selection Rate Comparison: Multiple Events (Mantel-Haenszel & Breslow Day)

MULTIPLE EVENTS

Selection Rate Comparison

ENTER DATA ONLY IN YELLOW CELLS

Clear Data

Important Note:

This program uses estimation techniques to analyze the data entered. While the results are likely to be similar to the results of an exact calculation process (which requires advanced statistical software), the user should consider the results only as estimations--especially for small data sets or when the results are close to statistical significance ($p < .05$).

	Reference Group (e.g., Men)		Focal Group (e.g., Women)		Selection Rate Difference ⁽¹⁾
	Applied	Successful	Applied	Successful	
Event 1	238	35	19	3	1.1%
Event 2	147	45	39	7	-12.7%
Event 3	235	54	87	17	-3.4%
Event 4	242	77	143	34	-8.0%
Event 5					
Event 6					
Event 7					
Event 8					
Event 9					
Event 10					
Event 11					
Event 12					
Event 13					
Event 14					
Event 15					
Total	862	211	288	61	

Step 1 - Pattern Consistency: Can the Events Be Combined Into An Overall Analysis?

White = Yes Orange = Warning Red = No

0.726	Pattern Consistency Test for Mantel-Haenszel (BD) ⁽²⁾
0.607	Pattern Consistency Test for Minimum Risk (T X S) ⁽²⁾

Step 2 - Statistical Test Results: Do the Combined Events Result in Statistical Significance?

White = No Orange = Warning Red = Yes

0.035	VALID	Mantel-Haenszel (Version 1) ⁽³⁾
0.040	VALID	Mantel-Haenszel (Version 2) ⁽³⁾
0.063	VALID	Minimum Risk Method ⁽⁴⁾

Step 3: Interpret Degree of Statistical Test Results ⁽⁵⁾

White = NA Orange = Warning Red = Significant

	Likelihood	# Std. Deviations
Degree of MH (Version 1):	29	2.111
Degree of MH (Version 2):	25	2.050
Degree of Min. Risk Test 3:	16	1.863



Availability Comparison: Single Event

Availability Comparison

ENTER DATA ONLY IN YELLOW CELLS

Calculate Exact Test

Clear Data

	Men	Women	White	Black	Hispanic	Asian	Native Amer.	Total Min.
# Selected/Represented	141	141	390	15				15
% Available*	50.0%	50.0%	84.6%	15.4%				15.4%
% Represented ⁽¹⁾	50.0%	50.0%	96.3%	3.7%				3.7%
% Over/Under Represented ⁽²⁾	0.0%	0.0%	11.7%	-11.7%				-11.7%
Stat. Test Results-EXACT ⁽³⁾								
Stat. Test Results-ESTIMATED ⁽⁴⁾	1.000	1.000	1.000	0.000				0.000

Interpretation of EXACT Statistical Test ⁽⁵⁾:

Likelihood (One Chance In):

Probability as Std. Deviations:

Interpretation of ESTIMATED Statistical Test ⁽⁵⁾:

Likelihood (One Chance In):

Probability as Std. Deviations:

1	1	1	#####				#####
-0.060	-0.060	-7.659	7.510				7.510



Availability Comparison: Multiple Events

MULTIPLE EVENTS Availability Comparison

ENTER DATA ONLY IN YELLOW CELLS

Clear Data

	Focal Group		Total in Overall Group	# Times Larger Outside Available Pool Than "Total in Overall Group" (Default = 10) ⁽¹⁾	% Over/Under Represented
	Availability (%)	# Selected/ Represented			
Event 1	38.3%	14	48	10	-9.1%
Event 2	38.5%	12	32	10	-1.0%
Event 3	4.3%	0	26	10	-4.3%
Event 4	13.8%	1	38	10	-11.2%
Event 5	20.9%	6	34	10	-3.3%
Event 6				10	
Event 7				10	
Event 8				10	
Event 9				10	
Event 10				10	
Event 11				10	
Event 12				10	
Event 13				10	
Event 14				10	
Event 15				10	

Step 1 - Pattern Consistency: Can the Events Be Combined Into An Overall Analysis?

White = Yes

Orange = Warning

Red = No

0.347 Pattern Consistency for Statistical Test ⁽²⁾

Step 2 - Statistical Test Results: Do the Combined Events Result in Statistical Significance?

White = No

Orange = Warning

Red = Yes

0.036 VALID Statistical Test (Fisher/Lancaster) ⁽³⁾

0.052 VALID Statistical Test (Generalized Binomial) ⁽⁴⁾

Step 3: Interpret Degree of Statistical Test Results ⁽⁵⁾

White = NA

Orange = Warning

Red = Significant

	Likelihood	# Std. Deviations
Degree of Statistical Test 1:	28	2.095
Degree of Statistical Test 2:	19	1.944



Biddle Consulting Group, Inc.

Over 30 Years of Human Resource Consulting Experience

Thank you!



**Biddle
Consulting
Group, Inc.**