

10 Contemporary Controversies in Adverse Impact Analyses of Selection Rates

The Center for Corporate Equality Teleconference
Thursday, September 1, 2011

Presenters:

David B. Cohen, M.S.

Mike Aamodt, Ph.D.

Eric M. Dunleavy, Ph.D.

Format of Teleconference

- Participant phone lines are muted.
- Submit questions by sending an email to webinar@cceq.org
- If you have any technical difficulties during the call, please email webinar@cceq.org
- Q & A session will be held at the end of the presentation.

The Center for Corporate Equality

- Nonprofit research and think tank organization – 501(c)(6)
- <http://www.cceq.org>
- Mission: To create workplaces free from bias and unlawful discrimination by:
 - Harnessing the synergies between human resource functions
 - Promoting affirmative action and equal employment regulatory compliance
- DCI Consulting Group provides staff services
- Harold Busch– Executive Director
- Recently published a TAC report on adverse impact analysis

A Note About These Controversies

- Some of these topics are not new...but are still “hot”
- We organized a 70-member TAC on this general topic in 2010, and some of these topics were some of the more contentious issues
- The law is always changing, and we included a couple of controversies related to this change
- Experts will disagree on just about anything
- The goal of our presentation is to identify the controversy and articulate the various perspectives on that issue

Adverse Impact: A simple example

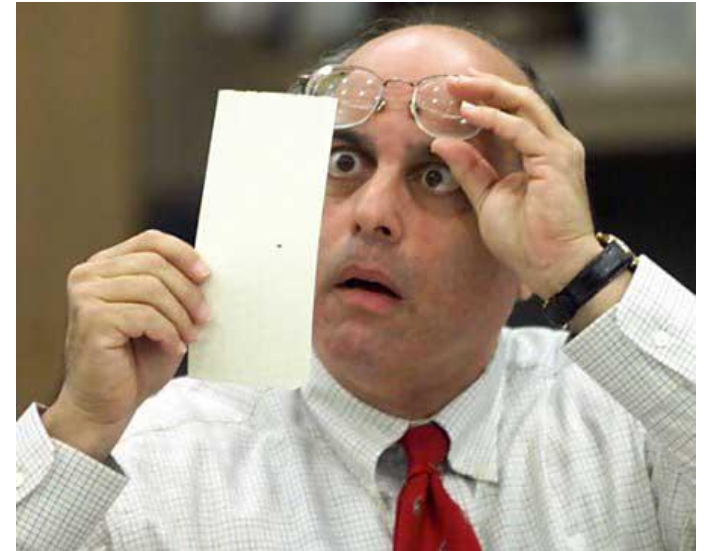
	Selected	Not	Totals
Male	40	60	100
Female	25	75	100
Totals	65	135	200

- Selection rate difference: $.40 - .25 = .15$. Is this 'big'?
- Selection rate ratio: $.25/.40 = .63$ and 4/5ths rule: $.63 < .80$
- Statistical significance (Z) = 2.26 sds, $p < .025$ (1 in 42 chance)
- Shortfall: expect 50/50 in a 'neutral' system, expected = 32.5, observed = 25, shortfall is 7.5.

Controversy 1: What on Earth is a “Selection” for Analysis Purposes?

Who Counts as a *Selection*?

- Hires and Selections Not the Same
- Selection
 - Accepted offer and reported to work
 - Accepted offer but failed to report to work
 - Rejected offer
 - Failed conditional offer
 - Medical or psychological exam
 - Physical ability test?
 - Background check?
 - Drug test?



Areas of Controversy

- Are there disparate rates of turning down an offer?
- What “tests” can legitimately be used post-offer?
 - Medical and psychological must be post-offer
 - Are post-offer tests used to skirt adverse impact?
 - A cognitive ability test given by a clinical psychologist as part of the post-offer psychological
 - A background check that “fails” an applicant with a conviction record or a bad credit history
 - In the control of the employer?
- If these are not “selections”, what are they?
 - Rejections?
 - Withdrawals?

Controversy 2: Which Groups Should be Compared in an Adverse Impact Analysis?

What is the Comparison Group?

- UGESP says “group with the highest rate”
- But to whom should this be compared?
 - Minority v. Non-minority?
 - Highest rate v. all others combined?
 - Highest rate v. all others individually?
 - Every individual group with each other group?
 - Combination of groups with similar rates?

Does it depend on the theory?

- Organization prefers Hispanics but doesn't dislike anyone
 - Highest selected (Hispanics) to all other groups combined?
- Organization dislikes African Americans but doesn't really have a preference among other groups
 - Lowest selected (African Americans) to all other groups combined?
- Organization prefers Hispanics, dislikes African Americans, but otherwise has no preference
 - Highest selected (Hispanics) to lowest selected (African Americans)

Does it Matter?

	White v. Minority	White v. Black	White v. Hispanic
White			
Applicants	60	60	60
Selections	30	30	30
Selection rate	.50	.50	.50
Minority			
Applicants	30	15	15
Selections	8	4	4
Selection rate	.27	.27	.27
Standard Deviation	2.11*	1.62	1.62

Does it Matter?

	White v. Minority	White v. Black	White v. Hispanic
White			
Applicants	60	60	60
Selections	30	30	30
Selection rate	.50	.50	.50
Minority			
Applicants	30	20	10
Selections	10	4	6
Selection rate	.33	.25	.60
Standard Deviation	1.50	2.35*	-0.59

2.20 Compared to Hispanics

Some recent EEO enforcement

- Interesting comparisons in OFCCP audits and conciliation agreements
 - Asians to African-Americans (for one job title out of hundreds)
 - Hispanics to non-Hispanics
 - Total minority to non-minority (White)

Controversy 3: What Do We Do With Applicants That Do Not Self-Identify?

What Do You Do with Applicants Who Don't Self-Identify?

- 8.26% Don't Self-Identify
 - 5.45% race and gender
 - 2.61% race
 - 0.20% gender
- Options
 - Exclude all from the analysis
 - Backfill race and gender for those that are hired (self-identification)
 - Backfill race and gender for those that are interviewed (visual ID)
 - Use applicant name and background to guess their race and gender

Areas of Controversy

- Does it matter?
 - Using duplicate applicants only added race for 16 out of 4,000 duplicate applicants
 - Less than 1% of missing information could be backfilled from hires list
 - It is mostly white applicants who do not self-identify
 - Can change adverse impact results
- Can you accurately identify from
 - Name?
 - Visual?
 - Clubs, schools, address?

Pat

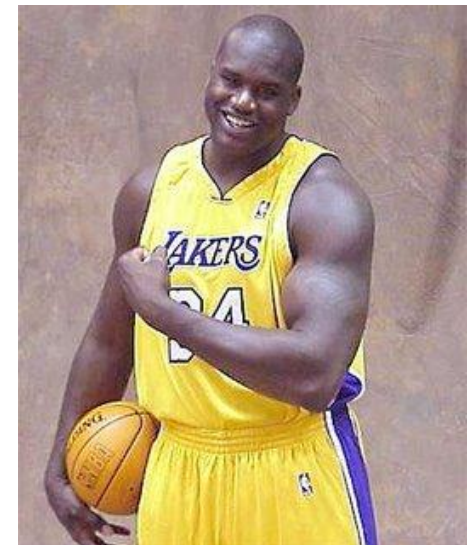


Name that Race/Ethnicity/Nationality!

- Fitzgerald
- Kennedy
- Lynch
- O'Neal

Name that Race/Ethnicity/Nationality!

- Fitzgerald
- Kennedy
- Lynch
- O'Neal



Other areas of controversy

- Accuracy
 - Review of research using surname lists suggests accuracy is between 75% and 90% depending on the race/ethnicity
 - More accurate for males than females
- Problems with visual identification
 - Accuracy and cheating
 - Encourages recruiters to focus on race and gender
 - Could this be used as proof of discrimination?

Controversy 4: How Do You Handle Applicants Who Apply Multiple Times?

An Example

One applicant applied 22 times!

Times Applied	N	%
1	13,416	78.9
2	2,181	12.8
3	693	4.1
4	309	1.8
5	146	0.9
6	85	0.5
7	61	0.4
8	36	0.2
9	26	0.1
10	11	0.1
11 +	32	0.2



Issues

- Where were the duplicates?
 - Same position, same requisition
 - Same position, different requisitions
 - Different positions
- Were the data aggregated?
 - Across requisition for the same position
 - Across positions
 - Across years

Does it Matter?

- Dunleavy, Mueller, Buonasera, Kuang, & Dunleavy (2008)
 - Used simulated data
 - Depending on the situation, a single, lesser-qualified applicant can solely cause or disguise adverse impact

Controversy 5: Is the 4/5th Rule Deceased?

4/5th Rule (UGESP, 1978)

- A selection rate for any race, sex, or ethnic group which is less than four-fifths (4/5) (or eighty percent) of the rate for the group with the highest rate will generally be regarded by the Federal enforcement agencies as evidence of adverse impact, while a greater than four-fifths rate will generally not be regarded by Federal enforcement agencies as evidence of adverse impact.

Areas of Controversy

- Was used early and often after UGESP was published in 1978
- These days? We don't see 4/5th rule often
 - OFCCP doesn't use it at all (Cohen a& Dunleavy, 2010)
 - EEOC is no fan
 - What about judges? It depends on the circuit
- Wait! The Supreme Court used it in Ricci didn't they?
 - Well, yes and no?

Other thoughts

- The 4/5th rule and statistical tests will tell you the same thing anyway RIGHT?
- Was the 4/5th rule intended to be accompanied by a significance test?
- The way it was created was ridiculous?
 - SMEs came up with it

Controversy 6: Which Statistical
Significance Test Should be Used to
Analyze 2 by 2 Data?

Statistical Significance Tests

	Selected	Not	Totals
Male	40	60	100
Female	25	75	100
Totals	65	135	200

• $Z = 2.26$ ($p = .024$)

• $FET = 2.12$ ($p = .034$)

•Various continuity corrected Z statistics = slightly smaller SD value (and slightly larger probability value)

•Mid-p corrected FET= similar to the continuity corrections

What did the TAC say?

- *It is important to determine the context in which the data are being analyzed. Substantial effort should be made to understand the reality of the employment decision process under scrutiny and to mirror that process in analyses. The statistical model (e.g., some form of binomial or hypergeometric) that best mirrors the reality of the employment decision process should be used. It is often difficult to understand how employment decisions were made, and thus challenging to identify the most appropriate model.*

A Recent Study

- Biddle and Morris (2011) conducted a large scale small sample simulation assessing power and error rates of various significance tests
- Concluded that:
 - FET was overly conservative (least control over type 2 errors) by a large margin
 - Z and Mid-p should be used instead of FET
 - FET had most control over type I errors and never exceeded an alpha of .05, but it was close.
- A note on the alpha = .05 cut point problem

Controversy 7: Should Statistical Significance Tests be Adjusted to Account for Multiple Comparisons?

Should Significance Level be Adjusted for Multiple Comparisons?

- Public sector analyses are often different than those conducted in OFCCP audits
 - Annual promotion exam v. AAP analyses
- Typical AAP might have hundreds of adverse impact analyses
 - Analysis first by EEO-1 Category (minimum of 10)
 - Then by job title
 - Example
 - 15 jobs groups
 - 3 race comparisons (W-AA, W-H, W-A) and 1 gender comparison per job group
 - 60 comparisons
 - We would expect 3 to be significant by chance alone (.05)

What could we do?

- How Might We Adjust Alpha?
 - Bonferroni adjustment
 - $.05 \div 60 = .00083 = 3.09 \text{ SD}$
 - Number expected by chance
 - We found two that were significant
 - Three would be expected by chance

Controversy 8: Did the *Stagi vs. Amtrak* Ruling Slay Practical Significance?

The context

- In *Stagi vs. Amtrak*, a class of women alleged that a policy of requiring union employees to have one year of service in a current position before being promoted to a management position adversely impacted previously laid off women, a policy termed a “one-year blocking rule.”
 - Plaintiffs laid off/bumped down to union job and then applied for management positions they previously held or supervised

District Court Ruling

- Focused on groupings combined into 1 analysis (600 decisions for different jobs)
 - Plaintiff single pool Probit analysis (an adjusted significance test) produced a standard deviation of 3.8 (or a p value of about X).
 - Defense conducted a weighted analysis across SSEG
 - SF of 6 in 600 decisions is meaningful
 - Impact ratio of .96
 - Court rules in favor of defense arguing that sub-pool weighting is more appropriate

Appeals Court ruling

- *Although it was a close case, the district court should not have granted the employer's motion for summary judgment. The employees' expert's decision to aggregate the data, although not obviously correct, was also not obviously incorrect, and so there remained a genuine issue of material fact--whether the one-year rule caused a disparate impact on the employer's female employees.*
- Importantly, the 3rd Circuit also noted that increased numbers make it more likely to exclude chance as a cause of adverse impact. Further, the appeals court ruled that:
 - (1) statistically significant results alone support causation,
 - (2) the 4/5th rule was not persuasive, and
 - (3) there is no additional requirement of practical significance given the inference of causation.
- This combination of ideas almost implies that statistical significance is practical significance because a disparity is probably not due to chance

Disparity analysis specifics

- No rigid mathematical formula
 - 2 pages on statistical significance testing and the “2 SD” threshold
 - Emphasis on causation
 - 4/5th rule and practical significance as a concept are criticized
 - Mention Waisome case (SD = 2.68, IR = .87) but define as an exception
 - Court rulings given more deference than UGESP
 - At no point is the combination of SS and 4/5th rule discussed

At the end of the day...using both practical and statistical measures makes sense.

Here is an example why:

# Applicants		# Selections		Selection Rates			Practical Measures		Statistical Test
Males	Females	Males	Females	Total	Males	Females	Impact Ratio	Diff in rates	SD (Z) test
100	100	99	98	0.985	0.99	0.98	0.99	0.01	0.58
1,000	1,000	990	980	0.985	0.99	0.98	0.99	0.01	1.84
1,200	1,200	1,188	1,176	0.985	0.99	0.98	0.99	0.01	<u>2.01</u>
10,000	10,000	9,900	9,800	0.985	0.99	0.98	0.99	0.01	5.82
100,000	100,000	99,000	98,000	0.985	0.99	0.98	0.99	0.01	18.40
1,000,000	1,000,000	990,000	980,000	0.985	0.99	0.98	0.99	0.01	58.17

Controversy 9: Will the Uniform Guidelines be Revised?

Context

- They should be revised but instead they will “outlive all of us”
- Mike McDaniel keynote presentation on the topic last year and IOP article “*UGESP are a detriment to personnel selection*”
- Two camps:
 - They are old and they stink!
 - They are what they are and reasonable judges will apply them in reasonable ways

What about a regulations review?

- On January 18, 2011, President Obama issued Executive Order 13563, which directs federal agencies to develop a plan to review existing regulations within each agency's purview
- EEOC, DOL, and DOJ all asked for public comment on what regulations should be reviewed and why
- SIOP, SHRM and CCE submitted comments recommending that UGESP be reviewed
 - SIOP focused on validity research
 - CCE focused on adverse impact measurement
 - SHRM mentioned both

And the verdict is.....

- None of the 3 agencies have UGESP on their 'to review' list
- EEOC and DOJ mentioned the SIOP/SHRM/CCE comments in their introduction
- Reality: if UGESP were easy to revise they would have been

CCE response article: 4 options

- Option 1: Maintain the status quo and accept with the disconnects between contemporary personnel selection and the Uniform Guidelines;
- Option 2: Revise the Uniform Guidelines and give that revision the deference of law;
- Option 3: Keep the Uniform Guidelines but educate EEO agencies and the courts to consistently use other useful authorities (e.g., case law, NCME/APA/AERA Standards, SIOP Principles);
- Option 4: Abolish the Uniform Guidelines and adverse impact as a theory.

Controversy 10: Audience Choice

Questions and Comments

- Submit questions by sending an email to webinar@cceq.org