Measuring Equity: Calipers, Yardsticks, and Judgement

Terrence Willett
CAIR Conference Anaheim 2018
Flyers on college campuses

https://www.sacbee.com/latest-news/article221031490.html
Why should we care about equity?

- Because it’s the right thing to do
- Expand society’s talent pool
- Reduce societal conflict
- Because we are required to
And it has

BIAS AND SYSTEMIC RACISM.
Disproportionate Impact (DI) Title 5 § 55502(e)

- (e) “Disproportionate impact” in broad terms is a condition where access to key resources and supports or academic success may be hampered by inequitable practices, policies, and approaches to student support or instructional practices affecting a specific group. For the purpose of assessment, disproportionate impact is when the percentage of persons from a particular racial, ethnic, gender, age, or disability group, who are directed to a particular service or course placement based on an assessment test or other measure is significantly different from the representation of that group in the population of persons being assessed, and that discrepancy is not justified by empirical evidence demonstrating that the assessment test or other measure is a valid and reliable predictor of performance in the relevant educational setting.
Equity Definitions

● “What is “equity” and how is it different from “diversity”? Equity refers to achieving parity in student educational outcomes, regardless of race and ethnicity. It moves beyond issues of access and places success outcomes for students of color at center focus.” – Center for Urban Education

● Equity at Cabrillo College is the commitment to cultivate an inclusive teaching and learning environment by providing equitable support, access, and opportunities for success emphasizing historically underrepresented and emerging student populations.

● Or, all student outcomes are equal upon disaggregation.

● How close is close enough to be equal?

● Are there situations that increase equity and inequity at the same time?
Equity Scenarios (after Stoup 2016)

AB-504 Community colleges: Student Success and Support Program funding

- AB 504 “require[s] the Chancellor of the California Community Colleges to establish a **standard methodology**, for measurement of student equity and disproportionate impact for disaggregated subgroups of the student population of the California Community Colleges, for use in the student equity plans of community college districts, as specified.” [emphasis added]

- How should we define subgroups?
- How atomic should dissaggregations be?
Common Equity Metrics

- 80% Rule
- Proportionality Index
- Percentage Point Gap
80% Rule

- “Equal Employment Opportunity Commission (EEOC) 80% Rule, outlined in the 1978 Uniform Guidelines on Employee Selection Procedures, and was used in Title VII enforcement by the U.S. Equal Opportunity Commission, Department of Labor, and the Department of Justice”
  – CCCCO 2015 Equity Guidelines

- Example:
  - Assume reference group has a completion rate of 60%
  - \(60\% \times 80\% = 48\%\)
  - Any group with a rate less than or equal to 48% is defined as disproportionately impacted
Proportionality Index

● “Compares the percentage of a disaggregated subgroup in an initial cohort to its own percentage in the resultant outcome group” – CCCCO 2015 Equity Guidelines

● Group A comprises 40% of the college population and 30% of completions
  ○ \( \frac{.3}{.4} = 0.75 \) therefore underrepresented in completions and disproportionately impacted

● Group B comprises 40% of the college population and 40% of completions
  ○ \( \frac{.4}{.4} = 1 \) therefore equally represented in completions and not disproportionately impacted

● Group C comprises 40% of the college population and 50% of completions
  ○ \( \frac{.5}{.4} = 1.2 \) therefore overrepresented in completions and not disproportionately impacted
Percentage Point Gap

● “Compares the percent of students in a disaggregated subgroup who succeed in an outcome with the percent of all students who succeed in the same outcome”
  – CCCC CO 2015 Equity Guidelines

● Example: Subgroup completion rate is 40% and college average is 50%
  40% - 50% = -10%

  ○ Original proposed metric indicated -3% and less is considered disproportionate impact

  ○ 2017 CCCC CO memo uses normal distribution to determine significance
What are the problems with the traditional metrics we are trying to fix?

- They do not account for the size of the population or group
- Opinions vary on which is the most difficult to explain
- The Legislature only wants one metric
- Institutional performance indicators are population data, not samples therefore there is no sampling error and so no statistics (although data quality is always an issue!)
- Can a single metric do the job without bias for all colleges?
- How precise do we need to be?
Tangent: Why did I take all those statistics?

- Predictive analytics
- Comparing distributions ex: Ethnic distribution of community compared to that enrolled at the college
- Estimating effect sizes
- Cluster analysis
- Fun!
- Other?
Binomial and Hypergeometric Trials

- **Binomial**
  - Fixed number of trials
  - Each trial independent or “with replacement”
  - Example: Probability of flipping at least 2 heads out of 5 attempts.

- **Hypergeometric**
  - Fixed number of trials
  - Each trial dependent on previous trial or “without replacement”
  - Example: Probability of poker hands.
hypergeometric distribution

\[ np(1-p) > 9 \]
\[ n/N < 0.1 \]

normal distribution

\[ N > 2000 \]
\[ n/N < 0.1 \]

binomial distribution

\[ np(1-p) > 9 \]

\[ \lambda > 9 \]

Poisson distribution

\[ n > 100 \]
\[ p < 0.05 \]
Comparing 3 Equity Metrics for “Majority-Minority” College

CCCCO Scorecard Completion 11-12 cohort

<table>
<thead>
<tr>
<th>College X Ethnicity</th>
<th>Cohort Count</th>
<th>Cohort Percent</th>
<th>Non-completion Count</th>
<th>Completion Count</th>
<th>% of Completions</th>
<th>Completion Rate</th>
<th>80% Rule</th>
<th>Proportionality Index</th>
<th>PPG</th>
</tr>
</thead>
<tbody>
<tr>
<td>African American</td>
<td>13</td>
<td>1%</td>
<td>9</td>
<td>4</td>
<td>1%</td>
<td>31%</td>
<td>55%</td>
<td>0.69</td>
<td>-14%*</td>
</tr>
<tr>
<td>Native American</td>
<td>5</td>
<td>0%</td>
<td>3</td>
<td>3</td>
<td>0%</td>
<td>50%</td>
<td>89%</td>
<td>1.13</td>
<td>6%</td>
</tr>
<tr>
<td>Asian</td>
<td>5</td>
<td>0%</td>
<td>1</td>
<td>4</td>
<td>1%</td>
<td>80%</td>
<td>143%</td>
<td>1.80</td>
<td>36%</td>
</tr>
<tr>
<td>Filipino</td>
<td>0</td>
<td>0%</td>
<td>0</td>
<td>0</td>
<td>0%</td>
<td>N/A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td>1,483</td>
<td>94%</td>
<td>823</td>
<td>660</td>
<td>94%</td>
<td>45%</td>
<td>80%</td>
<td>1.00</td>
<td>0%</td>
</tr>
<tr>
<td>Pacific Islander</td>
<td>0</td>
<td>0%</td>
<td>0</td>
<td>0</td>
<td>0%</td>
<td>N/A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>59</td>
<td>4%</td>
<td>26</td>
<td>33</td>
<td>5%</td>
<td>56%</td>
<td>100%</td>
<td>1.26</td>
<td>12%</td>
</tr>
<tr>
<td>Total</td>
<td>1,585</td>
<td>100%</td>
<td>882</td>
<td>703</td>
<td>100%</td>
<td>44%</td>
<td>* no DI with CO method</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Pros and Cons of Traditional Metrics

<table>
<thead>
<tr>
<th></th>
<th>80% Rule</th>
<th>Proportional-ity Index</th>
<th>Percentage Point Gap</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pros</strong></td>
<td>• Easiest to calculate</td>
<td>• Fairly easy to</td>
<td>• Easy to calculate</td>
</tr>
<tr>
<td></td>
<td>• Roots in federal regulation</td>
<td>calculate</td>
<td>(unless using CO</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Requires consideration</td>
<td>memo method)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>of participation rates /</td>
<td>Currently mandated</td>
</tr>
<tr>
<td></td>
<td></td>
<td>access metrics</td>
<td>by CO</td>
</tr>
<tr>
<td><strong>Cons</strong></td>
<td>• Doesn’t account for group</td>
<td>• Doesn’t account for</td>
<td>• Arithmetically</td>
</tr>
<tr>
<td></td>
<td>size</td>
<td>group size</td>
<td>problematic</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Will not detect DI</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>in “majority-minority”</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>situations</td>
</tr>
</tbody>
</table>
Hear of any other metrics?

- Chi-square
- Shannon-Wiener Index
- Gini-Simpson Index
- Other information metrics?
- Arbitrary rules of thumb that have face validity
Chi-square

- Contingency table analysis
- Assumptions include mutually exclusive categories, independence of observations, no more than 20% of cells have expected frequency less than five and no cells have zero counts

\[ \chi^2 = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}} \]
<table>
<thead>
<tr>
<th>College X Ethnicity</th>
<th>Completion</th>
<th>Non-Completion</th>
</tr>
</thead>
<tbody>
<tr>
<td>African American</td>
<td>0.57</td>
<td>0.46</td>
</tr>
<tr>
<td>Hispanic</td>
<td>0.03</td>
<td>0.03</td>
</tr>
<tr>
<td>White</td>
<td>1.62</td>
<td>1.31</td>
</tr>
</tbody>
</table>

\[ \chi^2 = 4.02 \]
\[ df = 2 \]
\[ p = 0.13 \]

Note that ethnicities with low counts that violate the assumptions of Chi-square analysis were omitted. Alternatively they could be grouped together although that would challenge interpretation.
Shannon-Wiener Information Index

- Measures diversity of a collection of items. Higher values indicate greater diversity.
- In informatics the index measures “entropy” or the uncertainty of predicting the next bit of information.
- Compare diversity in disaggregated metrics between cohort and completers.

\[ H' = - \sum_{i=1}^{n} (p_i \times \ln(p_i)) \]
<table>
<thead>
<tr>
<th>College X Ethnicity</th>
<th>Cohort ( p \times \ln(p) )</th>
<th>Completion Rate ( p \times \ln(p) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>African American</td>
<td>-0.04</td>
<td>-0.03</td>
</tr>
<tr>
<td>Native American</td>
<td>na</td>
<td>na</td>
</tr>
<tr>
<td>Asian</td>
<td>na</td>
<td>na</td>
</tr>
<tr>
<td>Filipino</td>
<td>na</td>
<td>na</td>
</tr>
<tr>
<td>Hispanic</td>
<td>-0.06</td>
<td>-0.06</td>
</tr>
<tr>
<td>Pacific Islander</td>
<td>na</td>
<td>na</td>
</tr>
<tr>
<td>White</td>
<td>-0.12</td>
<td>-0.14</td>
</tr>
<tr>
<td><strong>H’</strong></td>
<td><strong>0.224</strong></td>
<td><strong>0.233</strong></td>
</tr>
</tbody>
</table>

There is “more diversity” in the completion rate than the cohort. Is the difference enough to indicate DI? For which group is DI occurring?
Gini-Simpson Index

- If selecting two individual items randomly from a collection, what is the probability they are in different categories.
- The Gini coefficient is a measure of the inequality of a distribution, a value of 0 expressing total equality and a value of 1 maximal inequality.
- Compare diversity in disaggregated metrics between cohort and completers.

\[ D = 1 - \sum_{i=1}^{n} p_i^2 \]
There is “more diversity” in the completion rate than the cohort. Is the difference enough to indicate DI? For which group is DI occurring?
Use of term “diversity” brings in value judgements that may be counterintuitive to the statistical interpretation. In our case, if the optimal outcome is for the distributions to be the same, the indices alone do not contain enough information to make a determination.

<table>
<thead>
<tr>
<th>Category</th>
<th>Example Cohort</th>
<th>High Diversity with DI</th>
<th>Equal Diversity with DI</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>5%</td>
<td>33%</td>
<td>90%</td>
</tr>
<tr>
<td>B</td>
<td>5%</td>
<td>33%</td>
<td>5%</td>
</tr>
<tr>
<td>C</td>
<td>90%</td>
<td>33%</td>
<td>5%</td>
</tr>
<tr>
<td>H'</td>
<td>0.39</td>
<td>1.10</td>
<td>0.39</td>
</tr>
<tr>
<td>D</td>
<td>0.19</td>
<td>0.67</td>
<td>0.19</td>
</tr>
</tbody>
</table>
Bonus Method: Disaggregation of Survival Curves

Survival Curves of Ethnicity

- **Ethnicity**
  - White
  - Latino
  - Others

- **y-axis**: Probability (Completion)
- **x-axis**: Time (Month)

The graph shows the survival curves for White, Latino, and Others over time, indicating how long each group maintains a certain probability of completion.
Arbitrary Rules of Thumb With Face Validity

- Subgroup must have at least X students
- Difference between moving from DI to no DI must involve at least X students
- Set minimum levels of achievement for all groups
- Others?
Could we use a school of metrics?

- AB 504 requires “standard methodology” not a single metric
- No single metric is free from bias
- Could calculate equity by three different metrics and then:
  - If any one metric shows DI then there is DI (sensitive)
  - If and only if all three metrics show DI then DI (insensitive)
Thank you!