The Mismeasure of Group Differences in the Law and the Social and Medical Sciences

Institute for Quantitative Social Science
at Harvard University
Cambridge, Massachusetts, Oct. 17, 2012

James P. Scanlan
Attorney at Law
Washington, DC
jps@jpscanlan.com

There is currently no posted narrative accompanying this presentation. But an October 9, 2012 Letter to Harvard University written preparatory to the workshop provides a fair guide to the illustrations according to the following scheme: Letter Section (LS) A – Workshop Sections (WS) 1 and 5; LS B – WS 2; LS C – WS 4; LS D – WS 3.

Key Points

One: Standard measures of differences between outcome rates (proportions) are problematic for appraising the comparative situation of groups reflected by a pair of rates because – for reasons inherent in the underlying risk distributions – each measure tends to be systematically affected by the prevalence of an outcome.

- Relative differences
- Absolute differences

Two: Efforts to appraise differences in the circumstances of two groups reflected by a pair of outcome rates in the law and the social and medical sciences have been universally undermined by failure to recognize the way the chosen measures tend to be affected by the prevalence of an outcome.

Key Points (cont'd)

Three: Even when broadly correct, research is misleading by implying that the measures employed effectively quantify a difference in circumstances of two groups

Four: There exists only one answer to the question of whether differences in the circumstances of advantaged and disadvantaged groups reflected by outcome rates have increased or decreased or are larger in one setting than another.

Five: That answer can be divined, albeit imperfectly, by deriving from pairs of outcome rates the difference between means of the underlying risk distributions.

Caveat One

- Do not be distracted by the fact that one commonly finds departures from the patterns described here. Observed patterns are invariably functions of
  - (a) the strength of the forces causing rates to differ and
  - (b) the prevalence-related/distributionally-driven forces described here.
- Society's interest is solely in (a).
- Only with an understanding of (b) can one discover (a).

Caveat Two

- Do not think that presenting relative and absolute differences (or even both of the two relative differences and absolute differences) by any means addresses the issues raised here.
- The fundamental problem is that none of the measures is statistically sound.

Caveat Three

- Do not be distracted by the fact that distributions may not be normal.
- That may complicate efforts to interpret differences by use of a theoretically sound measure.
- But it is no basis for relying on standard measures as if the patterns described here (or patterns like them) did not exist.

References

- Harvard University Measurement Letter (Oct. 9, 2012)
- United States Department of Justice Measurement Letter (Apr. 23, 2012)
- Institutional Correspondence subpage and Consensus Section to Measuring Health Disparities page
- "Race and Mortality" (*Society 2000;* "Can We Actually Measure Health Disparities?" (*Chance* 2005)

Outline

1. The two relative differences

2. Absolute differences and odds ratios

3. Different perspectives/choice of measure fallacy; measure unaffected by prevalence

4. Pay for performance/Massachusetts Medicaid P4P program

5. Illogic of assumption of constant rate ratio; sound approach to subgroup analysis and calculation of NNT

1. The Two Relative Differences

Table 1. Explanation of Terms

(a) AG Fav Rt	(b) DG Fav Rt	(c) AG Adv Rt	(d) DG Adv Rt	(1) RR Fav	(2) RR Adv	(3) Abs Df	(4) Odds Ratio
90%	80%	10%	20%	1.125	2.00	0.10	2.25

RR = “relative risk” aka “rate ratio”; relative difference = RR -1

(1) RR Fav = a/b (1.125; relative difference is 12.5%)

(2) RR Adv = d/c (2.00; relative difference is 100%)

(3) Abs Df = a-b (10 percentage points)

(4) Odd Ratio = (a/c)/(d/b) (2.25)

Abbreviations

- NCHS: National Center for Health Statistics (Health People 2010, 2020 etc.)

- AHRQ: Agency for Healthcare Research and Quality (National Healthcare Disparities Report)

Interpretive Rule 1 (IR1): The Two Relative Differences (Heuristic Rule X (HRX), Scanlan's Rule)

The rarer an outcome

(a) the greater tends to be the relative difference in experiencing it and

(b) the smaller tends to be the relative difference in avoiding it.

Illustrative Data

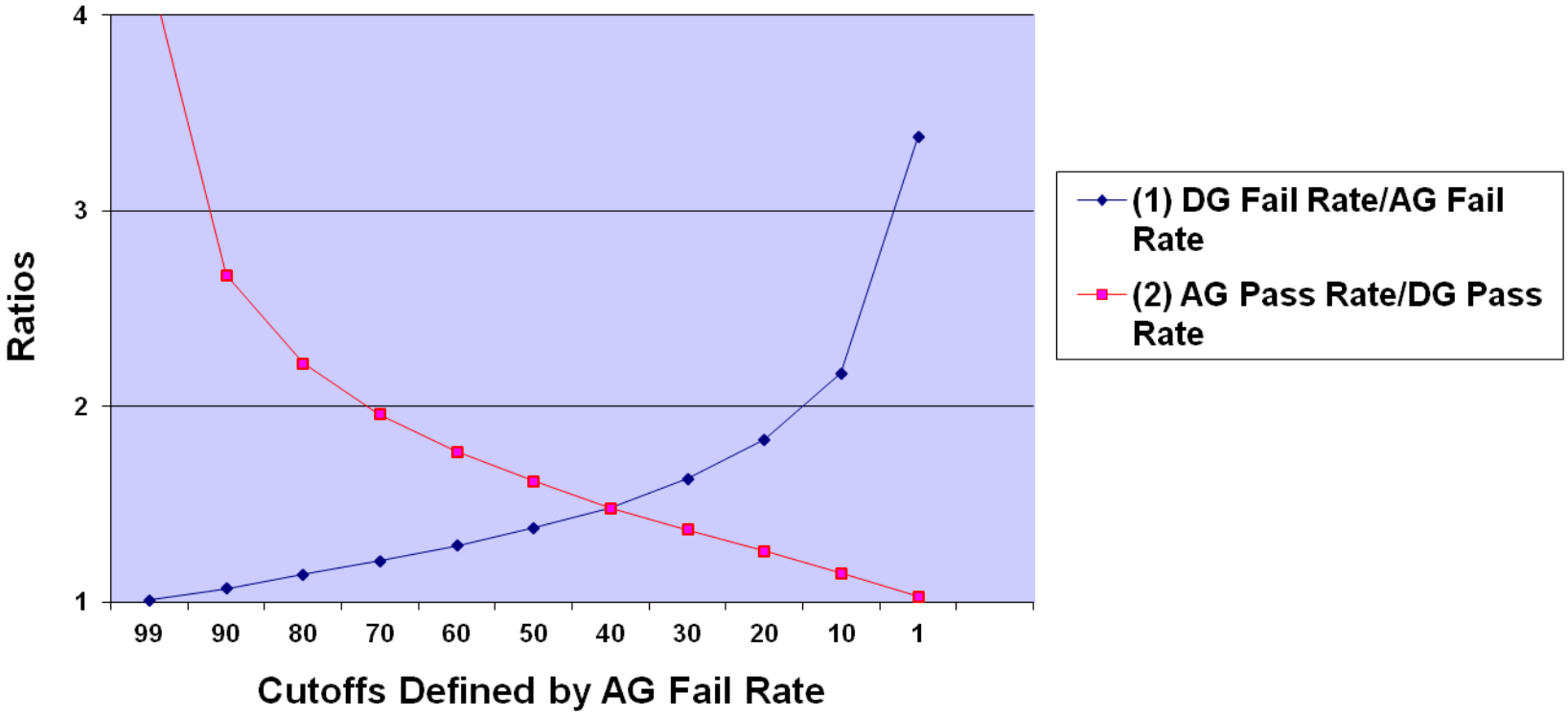
- Income Illustrations
- NHANES Illustrations
- Framingham Illustrations
- Life Table Illustration
- Credit Score Illustrations
- Test Score Data

Table 2: Simplified Illustration of Effects of Lowering Test Cutoff

(*National Law Journal* 2012, *Recorder* 2012)

Cut Point	Outcome	AG	DG	RR Pass	RR Fail
High	Pass	80%	63%	1.27	
High	Fail	20%	37%		1.85
Low	Pass	95%	87%	1.09	
Low	Fail	5%	13%		2.60

Fig. 1. Ratios of (1) DG Fail Rate to AG Fail Rate and (2) AG Pass Rate to DG Pass Rate at Various Cutoff Points Defined by AG Fail Rate



Corollary 1 to IR1

As an outcome changes in overall prevalence,

(a) the group with a lower baseline rate outcome will tend to undergo a larger proportionate change in the rate, while

(b) other group will tend to undergo a larger proportionate change in the rate for the opposite outcome.

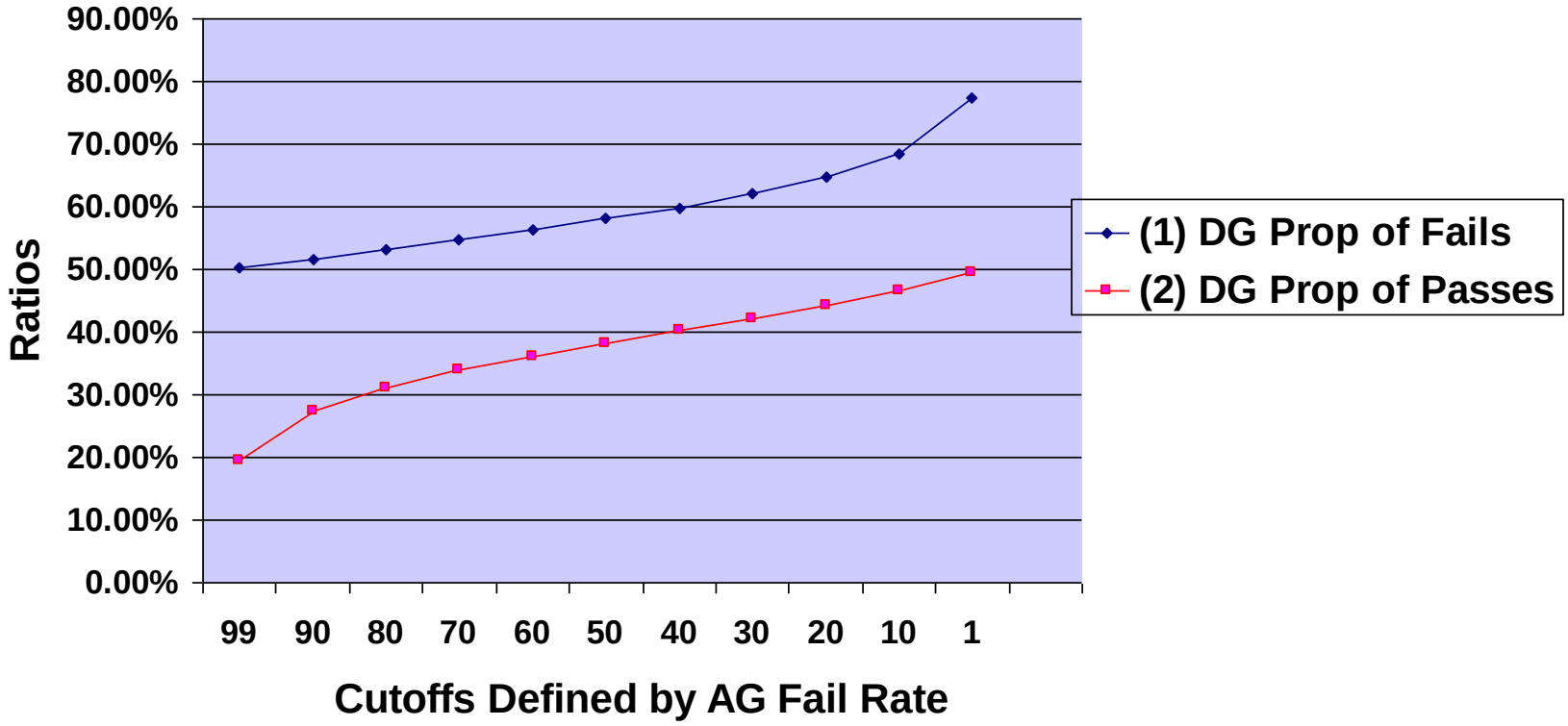
Corollary 2 to IR1

When an outcome declines in overall prevalence, the group most susceptible to the outcome will tend to comprise both

(a) a larger proportion of those continuing to experience the outcome and

(b) a larger (*sic*) proportion of those no longer experiencing the outcome. (Feminization of Poverty, Table 1 of Chance 2006)

Fig. 1a. Proportion DG Comprises of (1) Persons Who Fail and (2) Persons Who Pass at Various Cutoff Points Defined by AG Fail Rate



IR1 Implications

- Test pass/test fail (proficiency/non-proficiency)
- Poverty/non-poverty (Feminization of Poverty)
- Mortality/survival (Mortality and Survival)
- Immunization/no immunization (Immunization Disparities)
- Hypertensive/normal (NHANES Illustrations, ICHPS 2008)
- Low folate/adequate folate (NHANES Illustrations, Comment on Dowd IJE 2008)
- Loan rejection/loan approval (Lending Disparities)
- Expulsion/retention (Discipline Disparities)

Table 3: Changes in Total and Black Rates of Pneumococcal and Influenza Vaccination Rates, 1989-1995
(HHS *Progress Review: Black Americans*, Oct. 26, 1998)

Type	Yr	Total	Blk	RR Fav	RR Adv	AbsDf	EES
Pneumo	1989	15%	6%	2.50	1.11	0.09	0.53
Pneumo	1995	34%	23%	1.48	1.17	0.11	0.33
Influenza	1989	33%	20%	1.65	1.19	0.13	0.42
Influenza	1995	58%	40%	1.45	1.43	0.18	0.47

Table 4: Changes in Black and White Hepatitis-B Vaccination Rates Before and After School-Entry Vaccination Requirement (see Comment on Morita)

Period	Grade	Year	White Rate	Black Rate	Fav Ratio	Adv Ratio	AbsDf	EES
PreRq	5	1996	8%	3%	2.67	1.05	0.05	0.47
Post 1	5	1997	46%	33%	1.39	1.24	0.13	0.34
Post 2	5	1998	50%	39%	1.28	1.22	0.11	0.29
PreRq	9	1996	46%	32%	1.44	1.26	0.14	0.37
Post 1	9	1997	89%	84%	1.06	1.45	0.05	0.24
Post 2	9	1998	93%	89%	1.04	1.57	0.04	0.26

IR1 Implications – General (2)

- Less discriminatory alternatives (Discipline Disparities (B-D), Disparate Impact, Less Discriminatory Alternative – Substantive))
  - Lending Issues
  - Performance/retention standards
  - Disqualifying criteria (arrest/convictions/bad credit)
  - Mandatory sentencing (three-strikes etc.)
  - Discretion/review

IR1 Implications – Subpopulations

- Racial differences in infant mortality among highly-educated ("Race and Mortality")
- Occupational differences in mortality among British Civil Servants (Whitehall Studies)
- Racial and socioeconomic differences in mortality among younger age groups (Life Tables Illustrations)
- Racial differences in mortgage rejection rates among high income applicants (Disparities – High Income)
- Racial differences in completion/non-completion rates at elite universities
- Suburban discipline disparities (Suburban Disparities)
- Nordic health disparities

Figure 2. B/W Ratios for Bad Health and W/B Ratios for Good Health, by Income Level (from Comm. Paper Figure 8)

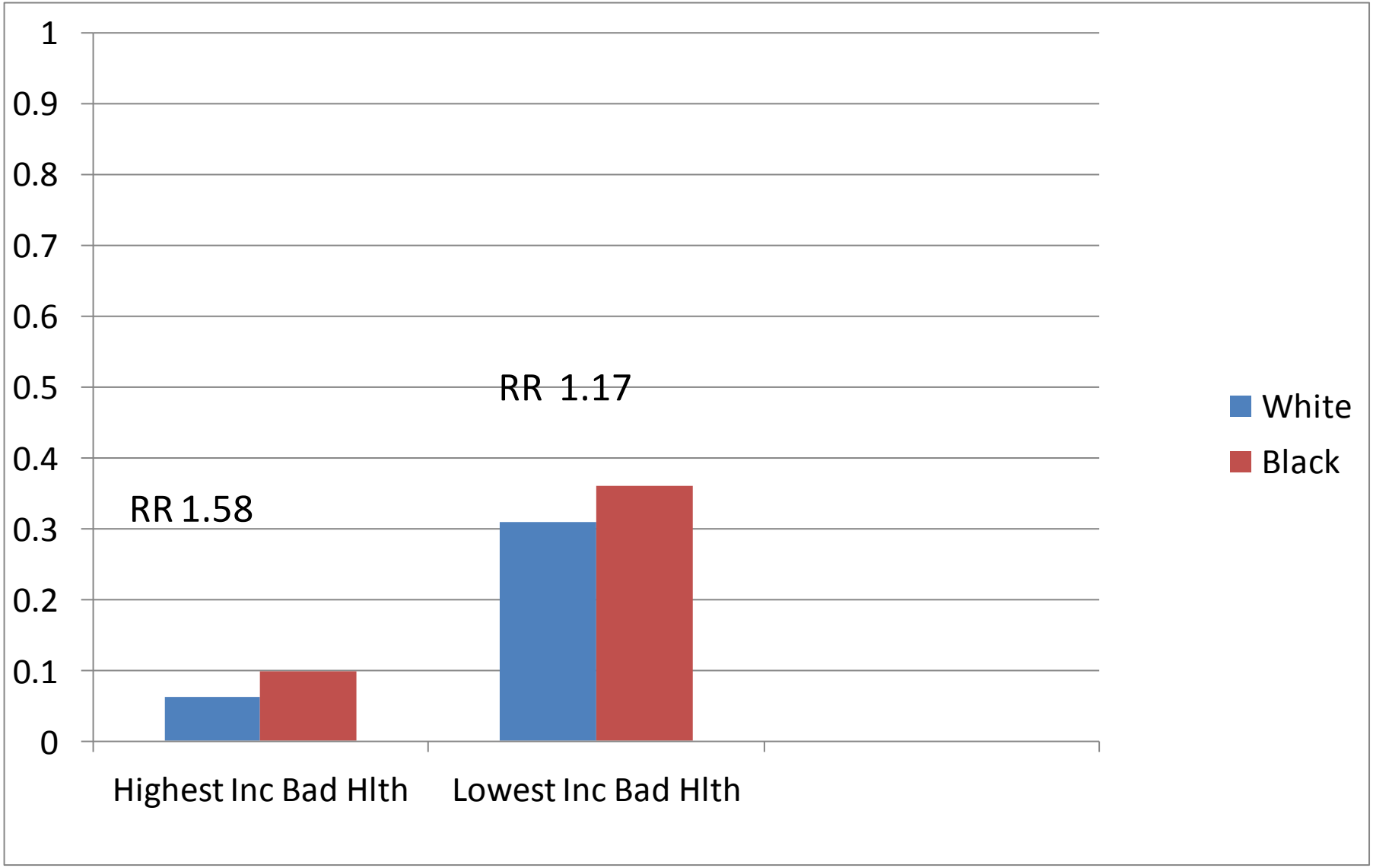


Fig. 2a. B/W Ratios for Bad Health and W/B Ratios for Good Health, by Income Level (from Comm. Paper Figure 8)

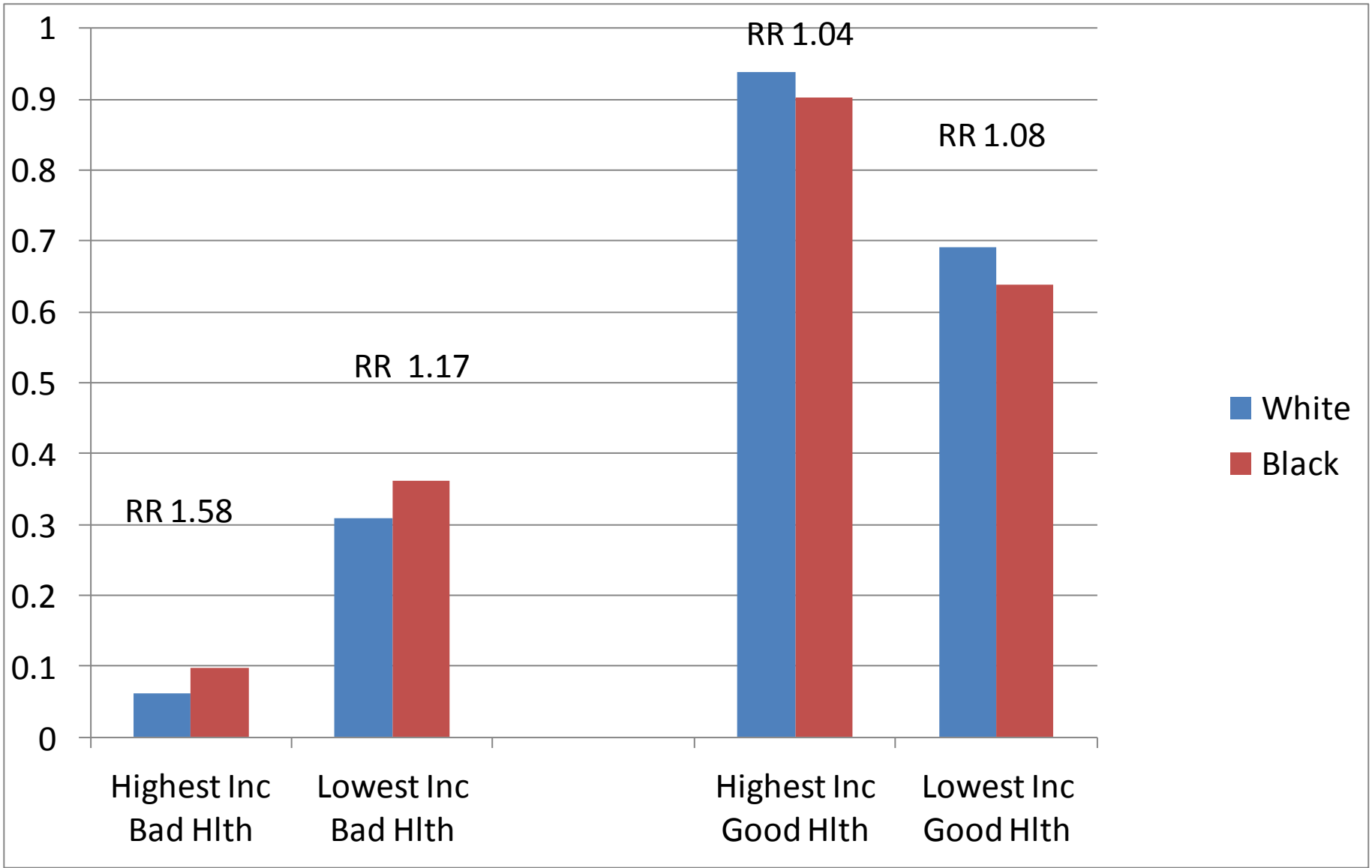


Fig. 3. M/F Ratios of Catheterization, by Whether Previous Infarction (from Steingart NEJM 1991, see Harvard Letter 40-41)

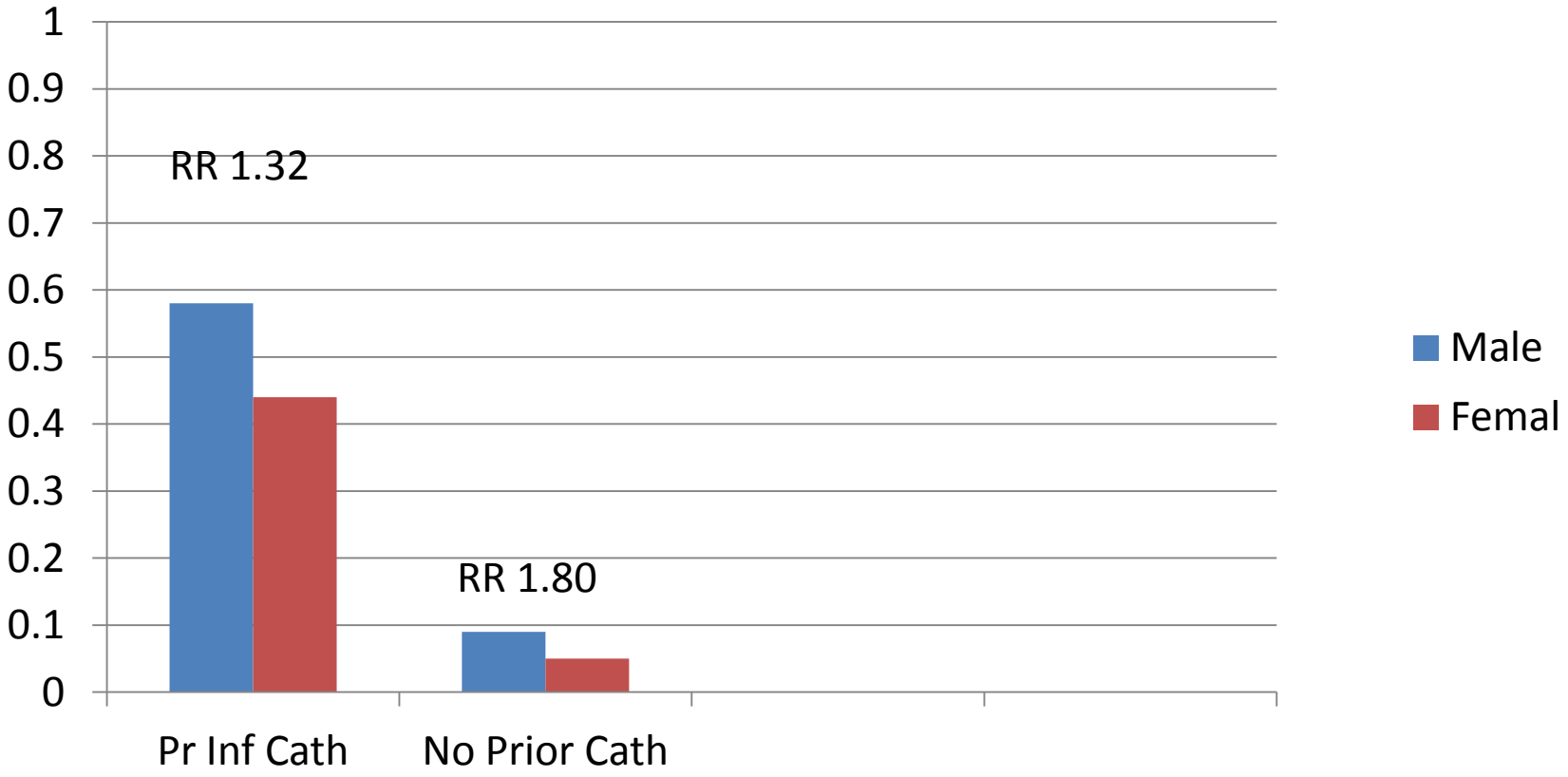
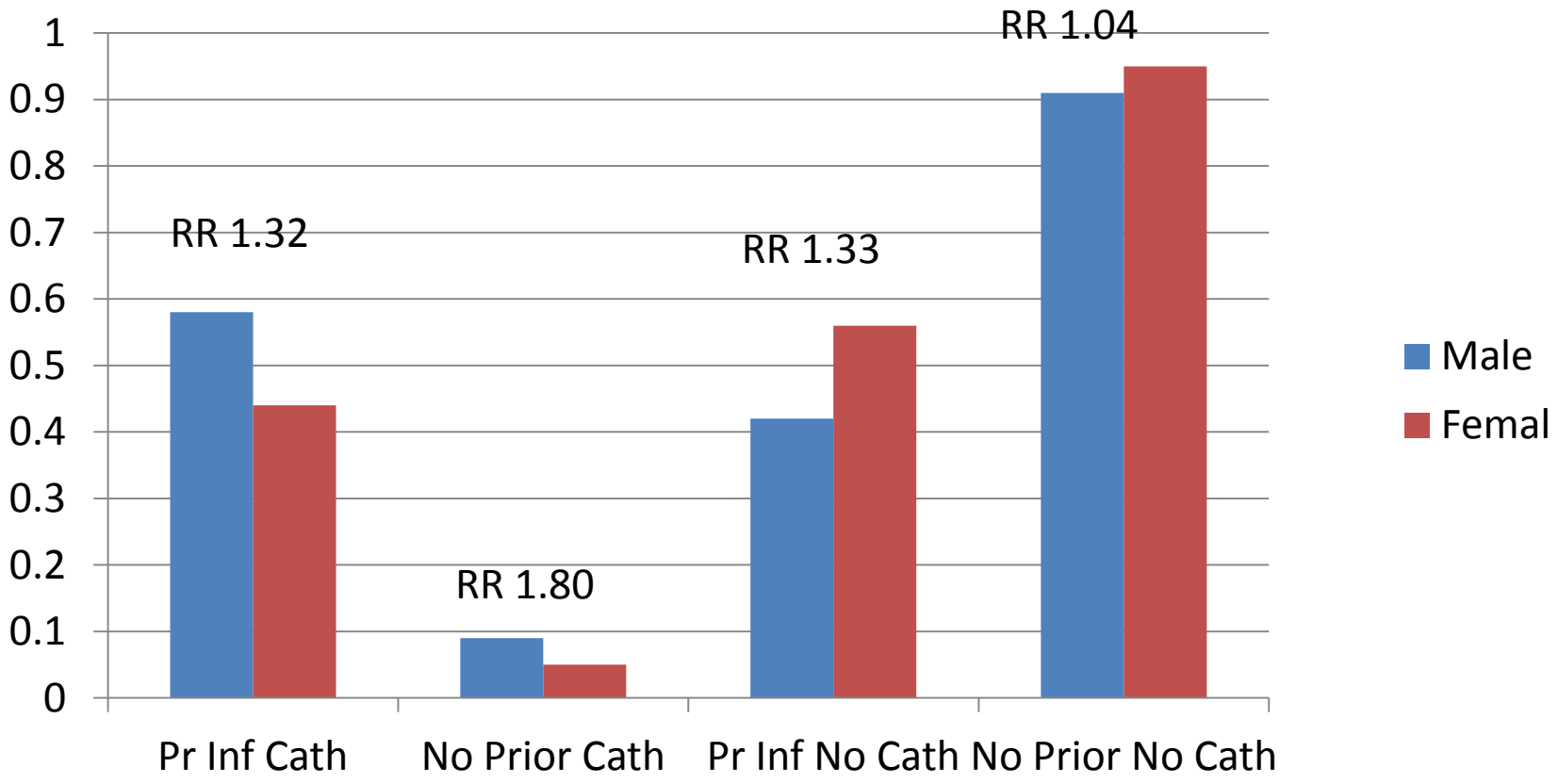


Fig. 3a. M/F Ratios of Catheterization and F/M Ratios of No Catheterization, by Whether Previous Infarction (from Steingart NEJM 1991, see Harvard Letter 40-41)



Implications of Corollary 1 to IR 1

- Effects of reductions/increases in poverty
- Effects of lowering/raising cutoffs (improving performance)
- Effects of improving health outcomes
- Explanatory theories: “diffusion of innovation,” “inverse equity hypothesis” (Explanatory Theories)
- Effects of chronic conditions on self-rated health (Reporting Heterogeneity, Comment on Delpierre BMC Pub Hlth 2012)
- Subgroup effects (Subgroup Effects, Illogical Premises)

Implications of Corollary 2 to IR1

- Feminization of Poverty
- Racial impact of Proposition 48
- Any discussion of the proportion a group comprise of persons experiencing some adverse outcome (addressed *infra*)

2. Absolute Differences (and Odd Ratios)

Absolute Differences/Odds Ratios

- Absolute differences and differences measured by odds ratios are unaffected by whether one examines the favorable or the adverse outcome.
- But an effective indicator must remain constant when there occurs a change in overall prevalence akin to that effected by lowering a test cutoff.
- Absolute differences and odds ratios tend also to be affected by the prevalence of an outcome but in a more complicate way than the two relative differences.

Interpretive Rule 2(IR 2): Absolute Differences/Odds Ratios

- As an outcome goes from being rare to being universal, absolute differences between rates tend to:

  (a) increase to the point where the first group's rate reaches 50%;

  (b) behave inconsistently until the second group's rate reaches 50%;

  (c) then decline.

- As the prevalence of an outcome changes, differences measured by odds ratios tend to change in the opposite direction of absolute differences.

Figure 4: Two Normal Distributions

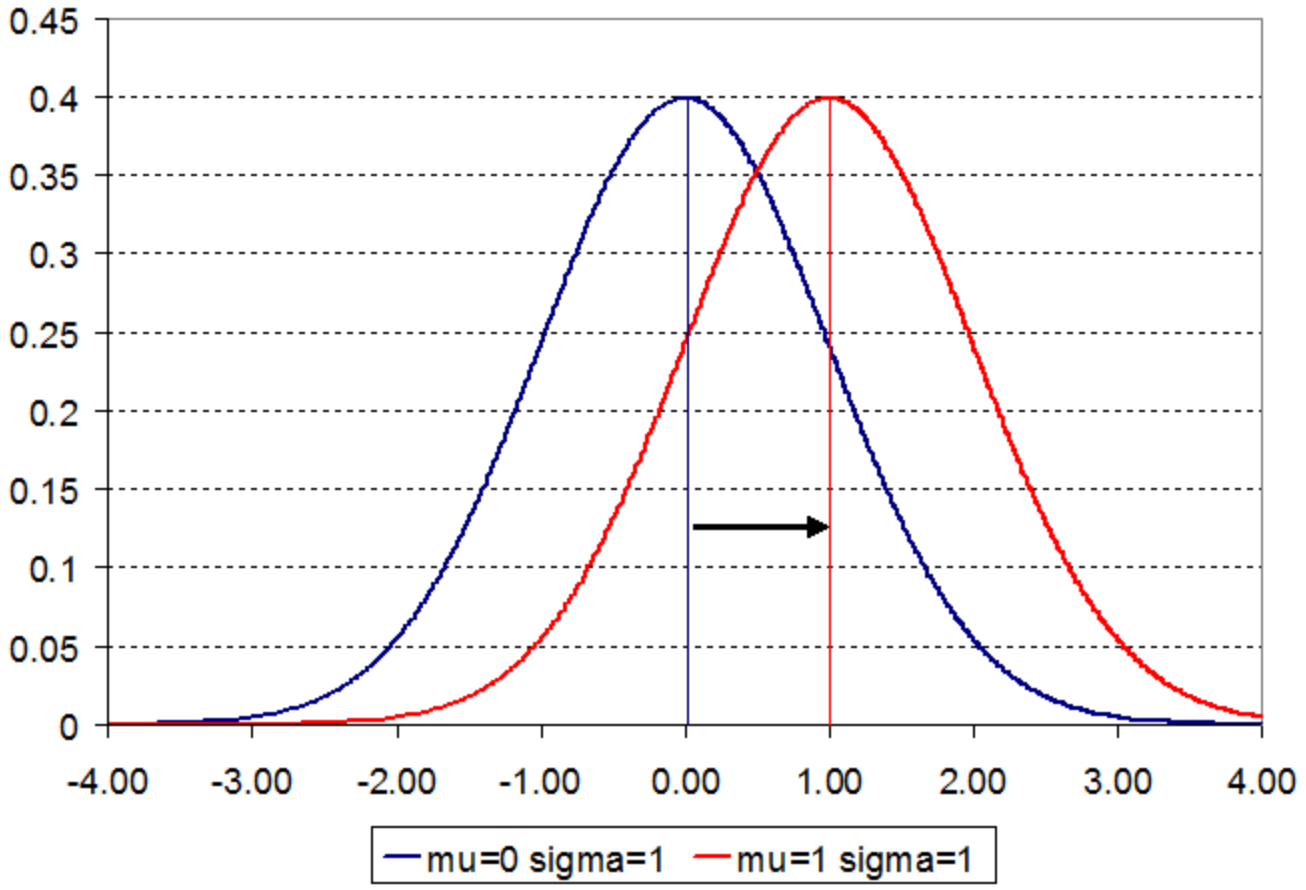


Fig. 5: Ratios of (1) DG Fail Rate to AG Fail Rate, (2) AG Pass Rate to DG Pass Rate, (3) DG Failure Odds to AG Failure Odds; and (4) Absolute Difference Between Rates

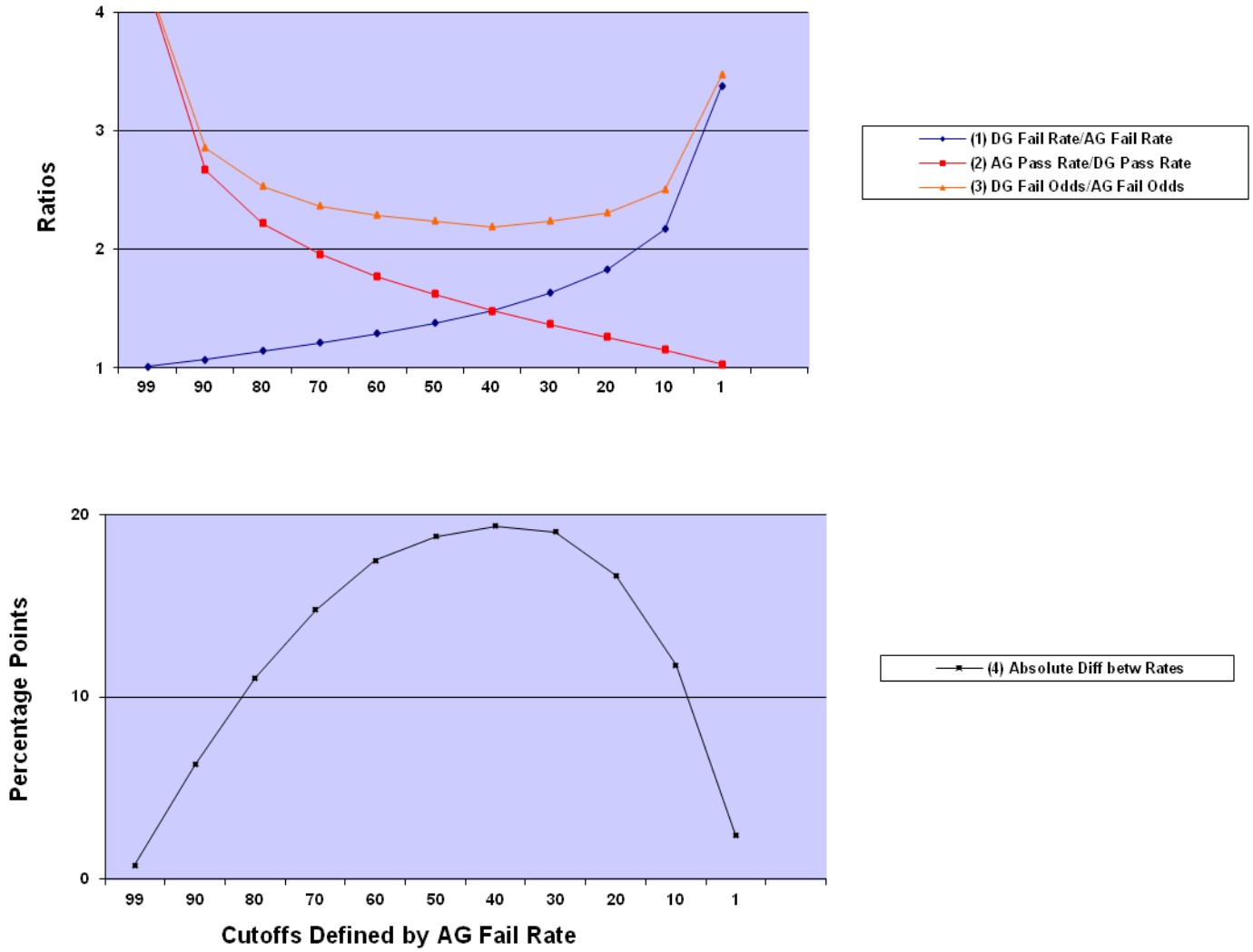
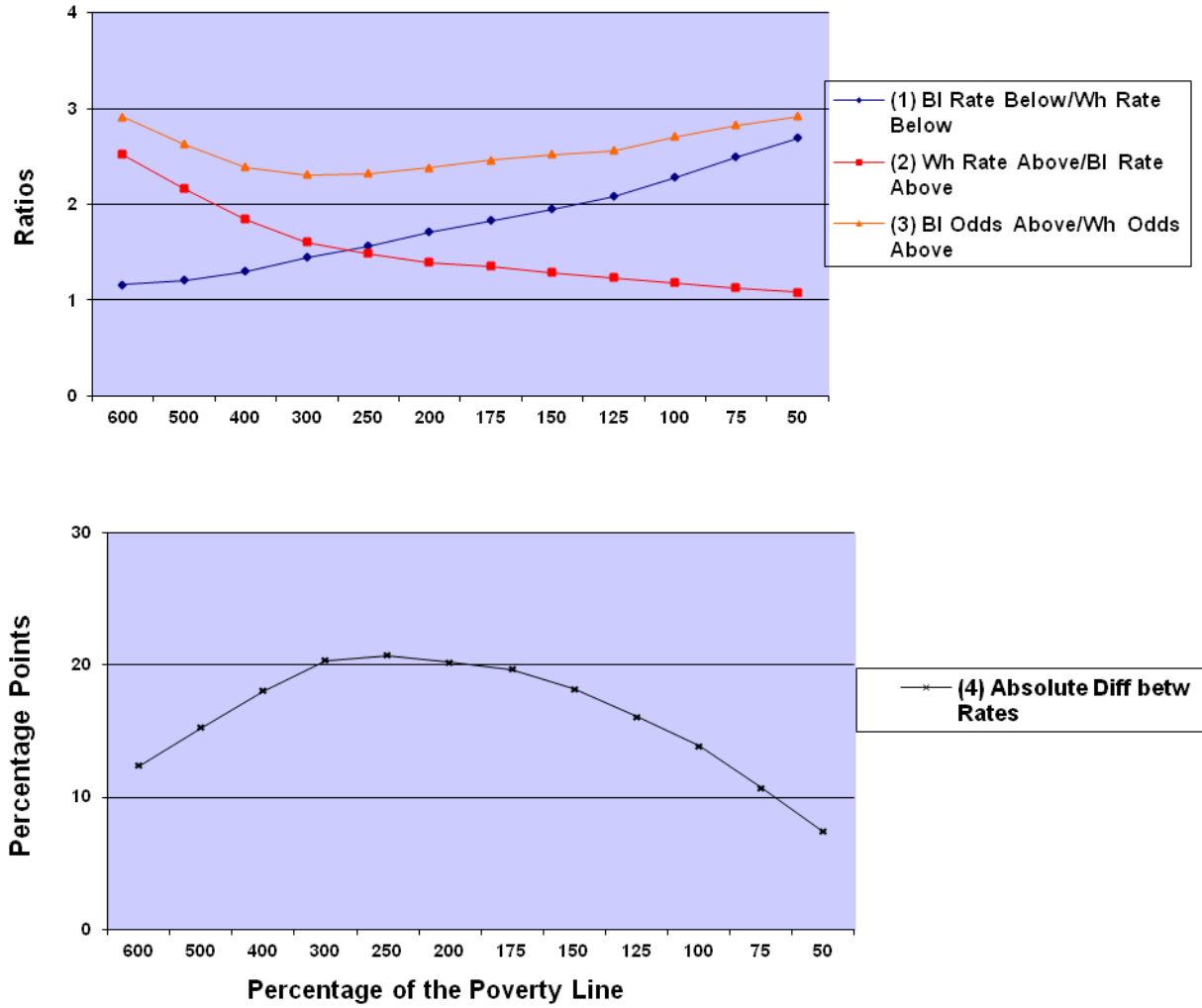


Fig. 6. Ratios of (1) Black to White Rates of Falling Below Percentages of Poverty Line, (2) White to Black Rates of Falling Above the Percentage, (3) Black to White Odds of Falling Below the Percentage: and (4)Absolute Differences Between Rates



Implications of IR2 (1)

- As uncommon procedures (e.g., coronary artery bypass grafting, knee replacement) increase, absolute differences tend to increase; as common procedures (e.g., mammography) increase, absolute differences tend to decrease. (APHA 2007, Comments on Vaccarino etc. NJEM 2005, Schneider JAMA 2001, Trivedi JAMA 2006 (2007), Sequist Arch Int Med 2006, McWilliams Ann Int Med 2009)

- As procedures go from being uncommon to being very common absolute differences tend to increase then decrease.
- Increased proficiency in more difficult subjects will tend to increase absolute differences, while increased proficiency in easier subjects will tend to reduce absolute differences. (Educational Disparities)

Implications of IR2 (2)

- For outcomes or settings with generally low rates, higher rates tend to be associated with larger absolute differences; for outcomes or settings with generally high rates, higher rates tend to be associated with lower absolute difference. (Between Group Variance, Comment on Baicker Hlth Aff 2004)

- Pay for Performance Issues (addressed infra).

3. Fallacy of Validity of Contrasting Measires/Value Judgment; Sound Measure of Disparity

Table 5: Illustration of Appraisals of the Comparative Degree of Employer Bias Using Different Measures of Disparities in Selection/Rejection

(as an illustration that choice of measure does not involve a value judgment and that all standard measures are unsound)

Employer/ Setting	AG Sel Rate	DG Sel Rate	RR Selection	RR Rejection	AbsDf	OR
A	20.0%	9.0%	2.22 (1)	1.14 (4)	0.11 (4)	2.53 (1)
B	40.1%	22.7%	1.77 (2)	1.29 (3)	0.17(2)	2.29 (3)
C	59.9%	40.5%	1.48 (3)	1.48 (2)	0.19 (1)	2.19 (4)
D	90.0%	78.2%	1.15 (4)	2.18 (1)	0.12 (3)	2.50 (2)

•

Approach 1 (relative favorable): A,B,C,D
Approach 2 (relative adverse): D,C,B,A
Approach 3 (absolute difference: C,B,D,A
Approach 4 (odds ratio): A,D,B,C

Table 5a: Illustration of Appraisals of the Comparative Degree of Employer Bias Using Different Measures of Disparities in Selection/Rejection: **Answer to which is most biased.**

Employer/ Setting	AG Sel Rate	DG Sel Rate	RR Selection	RR Rejection	AbsDf	OR
A	20.0%	9.0%	2.22 (1)	1.14 (4)	0.11 (4)	2.53 (1)
B	40.1%	22.7%	1.77 (2)	1.29 (3)	0.17(2)	2.29 (3)
C	59.9%	40.5%	1.48 (3)	1.48 (2)	0.19 (1)	2.19 (4)
D	90.0%	78.2%	1.15 (4)	2.18 (1)	0.12 (3)	2.50 (2)

• Which employer is in fact most biased? They are all the same. Each row reflects the half standard deviation between means underlying Tables 1 and 2 and Figures 1 through 5.

•Moreover, there is no rational argument that they are different.

Additional Factors Supporting Point of Table 5

- Exploring reason for changes in disparities or for why one disparity is larger than another.

- Drawing inferences about other things on the basis of appraisals of the comparative size of disparities or effects.

A Sound Measure of Disparity

- Implied in Table 3
- Derive from a pair of rates the difference between the means of the underlying, hypothesized normal distributions (in terms of percentage of a standard deviation).
- EES for estimated effect size
- Solutions sub-page of MHD
- Probit (Chester Ittner Bliss 1934)

Table 6. Illustration of Meaning of Various Ratios at Different Prevalence Levels

Ratio	DGFailRate	AGFailRate	EES
1.2	60.0%	50.0%	0.26
1.2	18.4%	15.4%	0.12
1.5	75.0%	50.0%	0.68
1.5	45.0%	30.0%	0.39
2.0	40.0%	20.0%	0.59
2.0	20.0%	10.0%	0.44
2.0	1.0%	0.5%	0.24
2.5	24.2%	9.7%	0.60
2.5	7.4%	2.9%	0.44
3.0	44.0%	14.7%	0.90
3.0	14.4%	4.8%	0.60
3.0	2.7%	0.9%	0.44

Table 7. Illustration of Problematic Nature of Representational Comparisons

DG Proportion Pool	DG Proportion Selection	AG/DG Selection Ratio
20%	10%	2.25
30%	20%	1.71
50%	30%	2.33
10%	5%	2.11
50%	25%	3.00

Explanation of Table 6

- Employment discrimination cases and various other matters (e.g., racial profiling analyses) are commonly based on comparisons of the proportion a group comprises of a pool and the proportion it comprises of persons experiencing an outcome.
- We can derive the rate ratios from the two proportions, as reflected in the final column.
- But we need the actual rates in order to derive the EES and determine which setting reflects the greater difference in the forces underlying the observed patterns.

Problems with the Solution

- Always practical issues (we do not really know the shape of the underlying distributions)

- Sometimes fundamental issues (e.g., where we know distributions are not normal because they are truncated portions of larger distributions). (Cohort Considerations, Life Tables Illustrations , Credit Score Illustrations, Comment on Boström and Rosen Scan J Pub Health 2003)

- Irreducible minimum issues (Irreducible Minimums)

- Notwithstanding, the problems the approach remains vastly superior to reliance on any of the standard measures.

- And how else, for example, would we be able to divine that the degrees of bias reflected by the actions of the employers in Table 3 are basically the same?

4. Pay for Performance

- Failure to recognize IR2 has led to perceptions in the US that P4P will increase healthcare disparities and in the UK that P4P will decrease healthcare disparities. Former perception has caused Massachusetts to include disparities measures in its Medicaid P4P program
- Failure to recognize IR2 has also caused Massachusetts to adopt a disparities measurement approach that will tend to lead to increases in healthcare disparities.
- References Pay for Performance and Between Group Variance subpages of MHD and ICHPS 2011

Table 8: Illustration from Werner (*Circulation* 2005) Data on White and Black CABG Rates Before and After Implementation of CABG Report Card (see Comment on Werner)

(1) Period	2 Wh Rt	(3) Bl Rt	(4) Fav Ratio	(5) Adv Ratio	(6) Abs Df (PP)	(7) OR	(8) EES
1	3.60%	0.90%	4.00	1.03	**2.70**	4.11	0.58
2	8%	3%	2.67	1.05	**5.00**	2.81	0.48

Table 10. Illustration of Absolute Differences as to Outcomes of Different Prevalence

Outcome	AG Fav Rt	DG Fav RT	Abs Df
A	20.05%	9.01%	0.11
A	30.15%	15.39%	0.15
B	79.96%	63.31%	0.17
B	89.97%	78.23%	0.12

Massachusetts Medicaid Pay for Performance Healthcare Disparities Criterion

- Between Group Variance (BGV) used for outcomes with rates generally above 80%. Higher BGV means greater disparity.
- BGV is a function of absolute differences and in rate ranges at issue, higher rates tend to be associated with lower absolute differences.
- Approach favors higher performing hospitals (HPH); HPH tend to serve comparatively fewer members of disadvantaged groups; resources diverted to hospitals with fewer minorities for reasons unrelated to health equity.
- Additional issue (unrelated to patterns described in this presentation): BGV increases as minority representation moves toward 50%; decreases at in moves from 50% toward 100%
- References: Between Group Variance sub-page of MHD; Comment on Blustein Hlth Aff 2011.

Misidentification of Subgroup Effects, Illogic of RR as Measure of Association and Miscalculation of NNT

- References: Subgroup Effects, Illogical Premises, Illogical Premises II, and Inevitability of Interaction subpages of Scanlan's Rule page.

- Joint Statistical Meetings 2009, Oral

5. Subgroup Effect

Illogical Premise Regarding Subgroup Effects

- Standard assumption is that a rate ratio will be constant across different baseline rates and departures from such pattern reflect subgroup effects (interaction, effect modification, etc).
- E.g., factor that reduces baseline adverse outcome rate (BAOR) from 20% to 12% (40% relative reduction) is expected to reduce BAOR of 10% to 6%.
- But if factor reduces different BAORs by equal proportionate amounts it necessarily increases baseline favorable outcome rates (BFORs) by different proportionate amounts (80% to 88% = 10.0%; 90% to 94% = 4.44%)
- Since there is no more reason to expect equal proportionate changes to different BAORs than to BFORs, it is illogical to expect equal proportionate changes to either.
- Interaction as to one outcome or as to the other is inevitable.

Table 10. Illustration of Illogic of Assumption of Constant Rate Ratio

Assumption	Control Adv	Treated Adv	Rel Adv Reduction	Control Fav	Treated Fav	Rel Fav Increase
Observed	20%	12%	40%	80%	88%	10.00%
Assume Equal Prop Adv Decrease	10%	6%	40%	90%	94%	4.44%
Assume Equal Prop Fav Increase	10%	1%	90%	90%	99%	10.00%

Rational Basis for Identifying Subgroup Effect

- A factor that reduces a rate of 20% to 12% shifts the underlying distributions by .33 standard deviations.
- A .44 standard deviation would reduce a rate of 10% to 5.32% (i.e., 46.8% relative risk reduction).
- Departure from .33 standard deviation should be benchmark for subgroup effect.

Rational Basis for Calculating Number Needed to Treat

- Based on an observed risk reduction in clinical trial of a 20% BAOR to 12%, NNT would be properly calculated as 12.5 (based on 40% relative risk reduction (RRR) = 8 percentage point (PP) absolute risk reduction (ARR))
- Where BAOR is 10%, standard approach would incorrectly yield NNT of 25 (based on 40% RRR = 4 PP absolute risk reduction.
- Where BAOR is 10%, rational approach would yield NNT of 21.4 (based on 46.8% RRR = 4.68 PP)
- See Tables 3 and 4 of Subgroup Effects for multiple comparison (including based on relative increase in favorable outcome and odds ratios)

Table 11. Illustration of Alternative Methods to Calculate NNT

Outcome	Control	Treated	Rel Risk Reduction	PP Reduction	NNT
Clinically Observed	20%	12%	40.00%	8.00	12.50
Assume Constant RR Adv	10%	6%	40.00%	4.00	25.00
Assume Constant RR Fav	10%	1%	90.00%	9.00	11.11
Assume Constant EES (.333)	10%	1%	46.80%	4.68	21.37