SUBGROUP EFFECTS/INTERACTION
(April 20, 2009; rev. Feb. 24, 2012)

[This is a pdf version of the Subgroup Effects sub-page of the Scanlan's Rule page of jpscanlan.com with endnotes converted to footnotes.]

Prefatory notes:

1. This item was originally a sketch created preparatory to the 2009 JSM presentation.[1] It has occasionally been expanded in various ways. Originally the discussion in the item and its references principally concerned the fact that, given reasons to expect that factors similarly affecting two groups with different baseline rates will result in different proportionate changes in those rates, perceptions of differential effects based on observed differing risk reduction ratios for groups with different baseline rates are mistaken. I eventually came to recognize that a more important matter involves recognition that, regardless of what one says about whether observed patterns reflect subgroup effects (differential effects, interaction, heterogeneity, effect modification), the distributional forces that underlie the points made about subgroup effects must be taken into account in making treatment decisions. It has been suggested that, absent sound evidence that an intervention differentially affects groups with different baseline rates (as the concept of differential effect or interaction is generally understood), one should assume that the same rate ratio applies across all baseline rates.[2] Such assumption underlies the guidance for applying an observed risk reduction to varying baseline rates in order to calculate the clinically relevant absolute risk reduction and corresponding number need to treat (NNT) provided by, among other entities the BMJ *publication* Clinical Evidence *and* University of Oxford's Centre for Evidence Based Medicine*. Thus, for example, if an intervention is observed to reduce an outcome rate by 10%, absent sound evidence to the contrary, such guidance assumes that the intervention reduces a rate of 10% by 1 percentage point (NNT = 100) and a rate of 30% by 3 percentage points (NNT = 33.3). But putting aside issues concerning the meaning of an overall reduction and how one derives it (such as discussed in the* Comparing Averages *sub-page and which are by no means unimportant[3]), estimation of the reduction for each baseline rate should instead reflect an underlying assumption that the risk distributions for each subgroup are shifted equivalent amounts. Thus, for an example based on the figures in Table 1 below, if Group A's figures reflected the observed reduction in an adverse outcome effected by an intervention –*

---

[1] Interpreting Differential Effects in Light of Fundamental Statistical Tendencies, presented at 2009 Joint Statistical Meetings of the American Statistical Association, International Biometric Society, Institute for Mathematical Statistics, and Canadian Statistical Society, Washington, DC, Aug. 1-6, 2009.
PowerPointPresentation : http://www.jpscanlan.com/images/Scanlan_JSM_2009.ppt
Oral Presentation: http://www.jpscanlan.com/images/JSM_2009_ORAL.pdf

[2] See Kent DM, Rothwell PM, Ionnadis JPA, et al. Assessing and reporting heterogeneity in treatment effects in clinical trials: a proposal. Trials 2010,11:85 (http://www.trialsjournal.com/content/11/1/85), which is the subject of my May 26, 2011 comment.

[3] Considerations of such issues may suggest that analyses should be based solely on comparisons of reductions of outcome rates of different subgroups rather than comparisons of subgroups with an overall rate, however such rate might be derived.

12.5% baseline rate reduced to 5.0% – the estimate of the most likely absolute reduction in the rate for Group B whose baseline rate is 21.7% would not be the 12.2 percentage point reduction based on the 61% relative reduction of Group A's adverse outcome rate from 12.5% to 5.0% or the 6.9 percentage point reduction based on the observed 8.8% increase in Group As favorable outcome rate from 87.3% to 95.0%. Rather, it would be a reduction of 11.7 percentage points based on the .5 standard deviation shifting of the mean of Group A's risk distribution reflected by the 12.5% and 5.0% figures for the control and treated groups. Tables 3 and 4 and the accompanying text have been added to briefly illustrate the implications of the different approaches.

2. Commencing in the spring of 2011, the [Framingham Illustrations](#), [Life Tables Illustrations](#), [NHANES Illustrations](#), and [Income Illustrations](#) sub-page were added to the Scanlan's Rule page to illustrate the patterns described in the introduction to that page using published data on various outcomes. The illustrations are equally pertinent to the points discussed here.

3. As discussed on the [Meta-Analysis](#) sub-page of the Scanlan's Rule page, the points discussed here apply as well to meta-analyses. Just as the rate ratio is a flawed tool for appraising subgroup effects where subgroups have different baseline rates, the rate ratio (or odds ratio) is a flawed tool for combining the results of a number of studies involving different baseline rates.

4. While this item principally focuses on the clinical setting, the points made about proportionate changes in different baseline rates, including that it is illogical to expect a factor to cause equal proportionate changes in different baseline rates, apply as well in the analyses of whether various groups benefit equally from general improvements in health and healthcare. See, e.g., [Comment on Korda IJE 2007](#) and the [Explanatory Theories](#) sub-page of the Scanlan's Rule page.

5. This item was initially around 700 words. As it grew in length and complexity I added bracketed numbers to facilitate ease of reference in other documents. In the February 21, 2012, revision, I added headings to make the item easier to follow. In order to facilitate cross-referencing from other materials that may have mentioned the bracketed section numbers, those bracketed numbers are retained.

A. Background

 [1] The [Scanlan's Rule](#) page and various other pages of this site and materials they make available explain the pattern – inherent in the shapes of normal risk distributions and observable in virtually any data set showing proportions of groups falling above or below multiple points on a continuum of factors associated with experiencing an outcome – whereby the rarer an outcome the greater tends to be the relative difference in experiencing it and the smaller tends to be the relative difference in avoiding it. A corollary to this pattern is one whereby a factor that increases or decreases an outcome for two groups with different baseline rates will tend to cause a larger proportionate change in the outcome in the group with the lower baseline rate, while causing a larger proportionate change in the opposite outcome for the other group.[4] Moreover,

---

[4] Since observers sometimes consider the statement in the text countered by evidence of situations where the described pattern does not occur, I note here, as I have in many places, observed patterns will be functions of both

since a factor cannot cause equal proportionate decreases in an outcome for each of two groups and at the same time cause equal proportionate increases in the opposite outcome for the two groups, it is unreasonable to expect equal proportionate changes in either outcome (an issue separately treated on the Illogical Premises sub-page of the Scanlan's Rule page).[5]

By way of example, a factor that improves test performance will tend to cause a greater proportionate decrease in the failure rate of the group with the higher average score while causing a larger proportionate increase in the pass rate of the group with lower average score. Similarly, a factor that reduces blood pressure will tend to cause greater proportionate reductions in hypertension in groups with lower average blood pressure while causing greater proportionate increases in rates of avoiding hypertension in other groups.[6] Thus, studies that identify differential effects of factors on subgroups (also termed "interaction") solely on the basis of differing proportionate changes in outcome rates may not be identifying a meaningful effect.

References in the Measuring Health Disparities (MHD) exploring these issues include A5 (*Chance* 1994*)*, A10 (*Society* 2000), B8 (APHA 2006), B19 (JSM 2009), D15 (Comment on Kaplan JECH 2006), D24 (Comment on Gan NEJM 2000), D30 (Comment on Mustard JECH 2003), D32 (Comment on Thurston AJE 2005), D33 (Comment on Martikainen JECH 2007), D36 (Comment on James JECH 2007), D59 (Comment on Kawachi Health Affairs 2005), D89 (Comment on Sun BMJ 2010), D95, D122 (Comments on Altman BMJ 2003), D105 (Comment on Mullins Health Affairs 2010), D107 (Comment on Berrington de Gonzalez MEJM 2010), D110 (Comment on Kent Trials 2010), D111 (Comment on Gabler Trials 2009), D112 (Comment on White BMC Med Res Meth 2005), D118 (Comment on Schwartz BMJ 2007), D124 (Comment on Wang Emerging Themes in Epidemiology 2009). Two recent items (D121 Comment on Chatellier BMJ 1996), D119 Comment on Cook BMJ 1995) address the articles underlying the *Clinical Evidence* guidance on adjusting for baseline risk discussed in prefatory note 1.

Table 1 below (which is based Table 1 of BSPS 2006) illustrates the issues. It shows a situation where an intervention shifts the risk distributions of two groups with different baseline adverse outcome rates by half a standard deviation (with the advantaged group (AG) rate in BSPS 2006 Table 1 used as the treated rate and the disadvantaged group (DG) rate used as the control rate). In other words, according to the reasoning above and on other pages of this site, the intervention

---

(a) the distributionally-driven forces just described and (b) differences between underlying distributions in the settings being compared (which in the context addressed here correspond to meaningful subgroup effects).

[5] Tables 3 and 4 show, inter alia, the varying absolute risk reductions estimates across a range of baseline rates depending on whether one assumes that an observed proportionate reduction in the adverse outcome rates applies to each baseline rate of experiencing that outcome or that an observed proportionate increase in the favorable outcome applies to each baseline rate of experiencing that outcome. The patterns presented under the Method 2 and Method 2 Alt columns might be deemed another illustration that the assumption of proportionate effects for different baselines is fundamentally illogical (as discussed further in Section D).

[6] While these and other examples are based on the way that a certain factor will affect all persons in a range between particular cut points, the matter would be more scientifically viewed in terms of a factor's causing the same change in the two groups' risk distributions (i.e., moving the means the same distance, measured in standard deviations, in one direction or the other).

affects the two groups in exactly the same way. But the penultimate column shows that the intervention effects a greater proportionate reduction in the adverse outcome rate in the group with the lower adverse outcome baseline rate (Group A). The final column, on the other hand, shows a larger proportionate increase in the favorable outcome rate for the group with the higher adverse outcome baseline rate (Group B).

Table 1: Comparative Effects of .5 Standard Deviation Change in Risk Distributions of Groups with Different Baseline Rates

Group	Point	Control	Treated	Relative Risk	Adv Risk Reduc	Fav Risk
A	M	12.71%	5.00%	0.39	60.67%	8.84%
B	L	21.77%	10.00%	0.46	54.06%	15.04%

Researchers who evaluate differential subgroup effects based on percentage reductions in adverse outcome rates would mistakenly conclude that Group A derived a greater benefit from the intervention. Researchers who evaluate differential subgroup effects based on percentage increases in favorable outcome rates would mistakenly conclude that Group B derived a greater benefit from the intervention. The point here, however, is not that there exist no genuine differential subgroup effects, but that the standard tools for identifying them are flawed.[7]

B. An Approach to Measuring the Effect of an Intervention That is Unaffected by the Baseline Rate

[2] The only statistically sound method for determining whether there exist subgroup effects that are not functions of different baseline rates appears to be that described on the Solutions and Solutions Database sub-pages of the Measuring Health Disparities page (MHD) of jpscanlan.com (with or without the adjustment for irreducible minimums discussed on the Irreducible Minimums sub-page of MHD). The method involves deriving from a pair or rates (such as the rates for an advantaged and a disadvantaged group or a treated group and a control group) the difference, in terms of percentage of a standard deviation, between means of the hypothesized underlying distributions. As discussed on the Solutions sub-page, the method that is mechanically implemented in the Solutions Database yields the same result as that derived formulaically by a probit analysis. The Solutions sub-page also discusses some of the shortcomings of the method. Notwithstanding those shortcomings, the method at least is based on a plausible assumption, whereas the assumption of a constant relative effect based either on

---

[7] The issues about spurious subgroup effects discussed here are different from the statistical significance issues described in such places as Wang R, Lagakos SW, Ware JH, et al. Statistics in medicine – Reporting of Subgroup Analyses in Clinical Trials. *N Engl J Med* 2007;357:2187-2194, and Lagakos SW. The challenge of subgroup analyses – Reporting without distorting. N *Engl J Med* 2006;354:1667-1669. The statistical significance issues would also exist with regard to valid tools for identifying subgroup effects. But in cases where standard measures of sub-group effects such as proportionately different reductions in rates are in the direction to which the statistical forces described here would tend to lead, that would add support to the view that the effects were not a result of chance (though, again, leaving open whether they reflected anything other than the statistical forces).

the treatment and control rates of experiencing the outcome or the treatment and control rates of failing to experience the outcome is an implausible assumption.

[3] Most early treatments of the method involved appraising the size of differences between the rates of two groups in order, for example, to determine whether the size of the difference increased or decreased during a period when there occurred an overall increase or decrease in the prevalence of an outcome. But the same principles apply with respect to evaluating the effects of an intervention, as illustrated in Tables 3 through 8 of the JSM 2009 presentation. That is, the same method used for appraising the size of the difference between the rates of two groups and hence for determining whether such difference was larger or smaller at different points in time (or before and after an intervention) can be used to appraise the effect of a factor or intervention on the rates of each group and hence to determine whether the factor or intervention had a greater effect on one group than another.

Table 2 below replicates Table 7 of the presentation. It illustrates the situation where observed patterns of changes in relative differences accord with the distributionally-based patterns. The final column shows in terms of the estimated effect size (EES) discussed on the Solutions sub-page, a measure of the effect of the intervention that is theoretically unaffected by the different baseline rates. The .34 and .32 EES figures show that the group that had the larger proportionate reduction in the adverse outcome (though smaller proportionate increase in the favorable outcome) experienced a larger effect in a meaningful sense, though the difference was quite small.[8]

Table 2 (2009 JSM Table 7): Comparison of the Effects of Beta Blockers on Mortality among Heart Patients at Different Ages

Age	BetaRate	NoBetRate	Adverse Reduction	Favorable Increase	EES
<70	11.30%	18.70%	39.60%	9.10%	0.34
>80	22.60%	33.10%	31.70%	15.70%	0.32

---

[8] As reflected in Tables 3 through 8, one can either evaluate the difference between the rates of two groups at two points in time (or before and after an intervention) or evaluate the size of the change each group's rate underwent during the period at issue (or as a result of the intervention). For reasons that should be evident once one thinks the matter through, the absolute differences between the EES figures reflecting the difference between the two groups' rates at two points in time will be the same as the absolute difference between EES figures reflecting the size of the changes in the two groups' rates. As shown in Table 8 of JSM 2009, EES between the rates of those under 70 and those over 80 was .45 among the NoBeta and .48 among the Beta. The difference between those two figures (.03) differs from the differences between the two EES figures in Table 2 of the text above (.02) only because of rounding.

C. Applying the Risk Reduction an Intervention is Observed to Achieve as to One Baseline Rate in Order to Estimate the Absolute Risk (and NNT) the Intervention is Likely to Achieve as to Other Baseline Rates

[4] As mentioned in prefatory note 1, most of the discussion above is focused on determining what is and is not a differential effect, while the crucial issue involves estimating the absolute risk reduction for differing baseline rates (from which can be derived the number need to treat (NNT)). Table 3, which is explained more fully in the Three Estimates of Absolute Risk Reduction document and which involves an observed reduction of an adverse outcome from 12.7% to 5.0% as in the first row or Table 1 supra (which underlies row M5 in Table 3)), shows the various estimates of absolute risk reductions across a range of baseline rates depending on the method of applying an observed reduction of a particular baseline rate. Since the NNT is simply the reciprocal of the absolute risk reduction when the absolute risk reduction is cast as a percent,[9] I omit that information from the table.

Method 1 is based on the above-described Solutions/probit approach to measuring differences between rates. It estimates the absolute risk reduction for each baseline rate on the basis that the intervention will achieve the same 5.standard deviation shift in the means of the underlying distributions reflected by the 12.7% and 5.0% figures.

Method 2 (the approach recommended by Kent et al. referenced in note ii and the three entities mentioned in prefatory note 1) applies the observed 61% relative risk reduction to all baseline rates. The approach will underestimate the absolute risk reduction (and overestimate the corresponding NNT) in situations where the baseline rate is lower than that in the underlying study and overestimate the absolute risk reduction (and underestimate the corresponding NNT) where the baseline rate is greater than that in the underlying study.

Method 3 applies the observed 8.8% relative increase in the favorable outcome to all baseline favorable outcome rates. It yields a pattern of overestimation and underestimation of the absolute risk reduction that is the opposite of the pattern yielded by Method 2.

Method 4 applies the observed odds ratio to all baseline rates, as recommended by Wang et al.[10] It yields an absolute risk reduction estimate that is the closest to that described in Method 1.

---

[9] Because of the confusion than can arise from the presentation of absolute risk differences or changes as percents (as discussed on the Percentage Points sub-page of the Vignettes page), in the tables below I present the estimated absolute risk differences in terms of percentage points. As so presented, the NNT would be 100 divided by the absolute risk reduction.

[10] Wang H, Boissel JP, Nony P. Revisiting the relationship between baseline risk and risk under treatment. *Emerging Themes in Epidemiology* 2009;6:1 (http://www.ete-online.com/content/6/1/1), which is the subject of my January 12, 2012 comment.

Table 3: Four Approaches to Estimating Absolute Risk Reductions (Percentage Point Reductions) (based on BSPS 2006 Table 1 where row M5 reflects the results of trial)

Cut Point	Baseline	Method 1	Method 2	Method 3	Method 4
O 1	3.44%	2.44	2.09	8.53	2.15
N 3	8.38%	5.38	5.08	8.10	5.15
M 5 (study)	12.71%	7.71	7.71	7.71	7.66
L 10	21.77%	11.77	13.21	6.91	12.55
K 20	36.69%	16.69	22.26	5.59	19.23
J 30	49.20%	19.20	29.85	4.49	23.08
I 40	59.48%	19.48	36.09	3.58	24.58
H 50	69.15%	19.15	41.95	2.73	24.14
G 60	77.34%	17.34	46.92	2.00	21.86
F 70	84.61%	14.61	51.34	1.36	17.86
E 80	90.99%	10.99	55.21	0.80	12.33
D 90	96.25%	6.25	58.40	0.33	5.90
C 95	98.38%	3.38	59.69	0.14	2.69
B 97	99.13%	2.13	60.15	0.08	1.47
A 99	99.76%	0.76	60.53	0.02	0.41

Table 4 presents a similar illustration based on the example in the Centre for Evidence Based Medicine's guidance on calculating the number needed to treat. The example is based on a situation where a trial resulted in reduction in event rates from 9.6% for diabetics treated with the usual insulin regimen and 2.8% for diabetics treated with an intensive insulin regimen. The relative risk reduction was 71% and the absolute risk reduction was 6.8 percentage points (with corresponding NNT of 14.7). Based on the 71% observed relative risk reduction, the example then presented absolute risk reduction where the baseline rate was 96% (high hypothetical) and 0.96% (low hypothetical). Those results are shown in the Method 2 column of Table 4. The remaining columns show the estimates for absolute risk reductions calculated in accordance with the other approaches described with regard to Table 3.

Table 4: Four Approaches to Estimating Absolute Risk Reductions (Percentage Point Reductions) (based on CEBM example of calculation of absolute differences and NNT)

Scenario	Baseline	Method 1	Method 2	Method 3	Method 4
Actual Trial	9.60%	6.80	6.80	6.80	6.80
High Hypo	96.00%	8.50	68.00	0.30	9.32
Low Hypo	0.96%	0.81	0.68	7.45	0.70

D. The Inevitability of Interaction as the Concept is Currently Understood

[5] Implicit in the above discussion is that the standard tests of interaction involving binary measures, being aimed at identifying departures from a common rate ratio, are fundamentally unsound. A useful online calculator for identifying such interaction is available [here](#).[11] An element of the calculation involves the confidence interval for each relative risk. But for illustrative purposes one can make the confidence intervals (which, in the main, are simply functions of sample size) sufficiently small that the procedure will show a statistically significant interaction in a comparison between any of the Method 1 results in the rows of Table 3 of between any of the Method 1 results in the rows of Table 4. The same would hold for tests of interaction conducted on the relative risk of experiencing the opposite outcome. That is, when observed patterns conform to patterns that would be expected when the factor affects each baseline rate exactly that same way, the standard approach to identifying interaction would find a subgroup effect in every row and do so both as to the adverse outcome and the favorable outcome.

On the other hand, if the Method 2 results were in fact observed for each adverse outcome baseline rate, one would find no interaction regarding the relative risk reduction for the adverse outcome regardless of the confidence intervals. And if the Method 2 Alt results were in fact observed for each baseline favorable outcome rate, one would find no interaction regarding the relative risk increase for the favorable outcome regardless of the confidence intervals. But if one did observe the Method 2 results for any row, one would necessarily find interaction between the results of that row and the reference row in a test involving the favorable outcome rates; and if one observed the Method 2 Alt results for any row, one would necessarily find interaction between the results of that row and the reference row in a test involving the adverse outcome rates.

But the two paragraphs above merely belabor the illogical nature of an assumption that, absent interaction, one will observe a constant relative risk across different baseline rates (the subject of the [Illogical Premises](#) sub-page). The point is nevertheless further developed on the [Inevitability of Interaction](#) sub-page.

Addendum
(Nov. 24, 2011)

Prefatory note to Addendum: Versions of this sub-page prior to November 24, 2011 included as discussion of a body of evidence where interventions have caused larger proportionate reductions of groups with higher baseline rates than lower baseline rates. But I came to believe that the discussion broke up the more important discussion. Further, subsequent to creating that material I came to a better understanding of how widespread is the view that an observed relative risk reduction will apply across a range of baseline risk (as reflected in the guidance

---

[11] The formula underlying the calculate is based on Altman DG, Bland JM. Interaction revisited: the difference between two estimates. BMJ 2003;326:219): Altman DG, Bland JM. Interaction revisited: the difference between two estimates. BMJ 2003;326:219, which is the subject of my September 21, 2010 [comment](#).

discussed in prefatory note 1 and many of the subjects of the comments that are listed in the third paragraph of Section A, inclining me to believe that the issue was less important than I had originally believed. But because it had been intended to be part of the 2009 presentation, I have kept the discussion in this addendum.

There exists a fair body of evidence suggesting that, contrary to the patterns described in Section A, interventions tend to cause a larger proportionate decline of rates among groups with higher baseline rates (such as in the sources summarized in Sackett DL. Why randomized controlled trials fail but needn't: Failure to employ physiological statistics, or the only formula a clinician-trialist is ever likely to need (or understand!). *JAMC* 2001:165(9): 1226-1237). The 2009 JSM paper was intended to seek to reconcile the theory with the contrary evidence, but ultimately did not address the matter. Thus, that issue continues to warrant attention.

While I do not expect many readers to fully understand the following points, which reflect some tentative thinking on my part, it seems possible that three issues/phenomena may be involved in the seeming departures from expectation. The first involves irreducible minimums, a subject discussed on the Irreducible Minimums sub-page of MHD. The second involves a regression toward the mean issue discussed in Sharp SJ, Thompson SG, and Altman DG. The relation between treatment benefit and underlying risk in meta-analysis. *BMJ* 1996;313:735-738. The third involves the tendency for the same effect size (properly measured, as discussed below) to be more often found to be statistically significant higher risk groups, as illustrated in the Statistical Significance SR sub-page of the Scanlan's Rule page, resulting in a form of publication bias.