

Besting the Odds: Optimal Reporting of Logistic Regression

Morgen M. Miller^a, Edward C. Norton^{a,b}, Kasey Coyne^c, Jason Wang^d, and Lawrence C. Kleinman^{e,f}

a. Departments of Health Management & Policy and Economics, University of Michigan, Ann Arbor, MI

b. National Bureau of Economic Research, Cambridge, MA

c. Departments of Health Evidence & Policy, New York, NY

d. New York City Department of Health and Mental Hygiene, New York, NY

e. Departments of Health Evidence & Policy and Pediatrics, Mount Sinai School of Medicine, New York, NY

f. Department of Society, Health, and Human Development, Harvard School of Public Health, Boston, MA.

1. Introduction

Researchers often use logistic regression to isolate the effect of specific factors on a dichotomous outcome, controlling for confounders (Hosmer and Lemeshow, 1989; Lee 1981). Results from logistic regression models can be presented in a variety of ways, such as an odds ratio, a risk ratio, or a risk difference. Odds ratios are easily calculated from the logit model and thus are commonly reported. Although many researchers consider the risk ratio to be a more intuitively understood and interpreted measure of relative effect, accurate conversion from the odds ratio to the risk ratio has proven to be difficult (Holland, 1989; Greenland and Holland, 1991; Zhang and Yu, 1998; McNutt et al., 2003). To create a practical solution to this longstanding issue, Kleinman and Norton (2009) developed and validated regression risk analysis (RRA), a simple and general method to estimate adjusted risk ratios (ARR) and adjusted risk differences (ARD) directly from nonlinear models. Using logistic regression as an example, they showed that regression risk analysis produces precise and accurate ARR and ARD estimates, both for simple data sets and ones with substantial confounding. Regression risk analysis reproduces Mantel-Haenszel ARR and ARD estimates for logit models with categorical confounders. It also produces highly accurate results for models with continuous covariates and does not experience computational difficulties that other methods do in such scenarios.

This paper extends the work of Kleinman and Norton (2009) in several important ways. First, we demonstrate that regression risk analysis is applicable not only to logistic regression, but also to other nonlinear probability models, such as the probit and multinomial and ordered logit models, and to cases involving complex survey design. Second, we provide an exemplar model, with data from the National Survey of Children's Health (NSCH), to show that ARR and ARD estimates can be easily generated using our Stata program. Third, we discuss two methods for measuring variance for regression risk analysis: the delta method and bootstrapping.

Fourth, we examine the effect of regression risk analysis on the choice of measure and implications of that choice within the context of comparative effectiveness research. Because regression risk analysis generates accurate risk ratio and risk difference estimates easily from nonlinear models, it presents researchers with a choice of how to report their results. This led us to contemplate what would be desirable practice for deciding between these measures. We noticed that in studies indirectly comparing two or more treatments, the ranking of treatments may vary across measure choices. That is, results reported as odds ratios may rank treatments differently than results reported as risk ratios. We have termed this phenomenon *rank reversal*. In this paper, we will describe the circumstances in which rank reversal does and does not occur and provide an example using a recent meta-analysis by Matchar and colleagues (2008). We will also discuss the implications of rank reversal for research and policy decisions and recommend best practice.

2. Methods

Logistic regression is a common approach to estimating models with binary dependent variables. When running such models, results can be reported in various ways, with the most common being odds ratios, risk ratios, and risk differences. Odds ratios are accurate measures of relative effect size if interpreted correctly. However, research

shows that confusion can arise when understanding odds ratios, and many people mistakenly interpret these results as risk ratios or relative risks (Kladman 1990; Teuber 1990; Altman, Deeks, and Sackett, 1998; Bier 2001). The unintuitive nature of the odds ratio and frequent confusion with the risk ratio has resulted in many articles calling for the end of reporting odds ratios (e.g. Spiegelman and Hertzmark, 2005).

As researchers, two primary goals in our work are to first conduct high-quality research and second to communicate the results of our research accurately and effectively. The adjusted risk ratio (ARR) and adjusted risk difference (ARD) are measures that satisfy these goals and can be estimated simply after running a logistic regression model.

2.1 Estimating adjusted risk ratios and adjusted risk differences

Kleinman and Norton (2009) developed the adjusted risk ratio (ARR) and adjusted risk difference (ARD) as valid measures of relative and absolute risk. The measures are calculated using two predicted probabilities from the model of interest. In the case of a binary explanatory variable of interest, one predicted probability is estimated when the covariate of interest equals the non-exposed value (typically zero) while the other predicted probability is estimated when the covariate of interest equals the exposed value (typically one). For example, if the explanatory variable of interest is exposure to some treatment or risk, then after estimating the model one can predict the predicted probability as if exposed or as if unexposed for each observation. The ARR and ARD are calculated using these predicted probabilities averaged across the entire sample. The ARR is the ratio of the mean predicted probabilities, and the ARD is the difference of the mean predicted probabilities.

$$ARR = \frac{P_1}{P_0} = \frac{\frac{1}{N} \sum_{i=1}^N risk_i(X_i | \text{as if exposed})}{\frac{1}{N} \sum_{i=1}^N risk_i(X_i | \text{as if unexposed})}$$

$$ARD = P_1 - P_0 = \frac{1}{N} \sum_{i=1}^N risk_i(X_i | \text{as if exposed}) - \frac{1}{N} \sum_{i=1}^N risk_i(X_i | \text{as if unexposed})$$

These equations can be generalized to apply to continuous covariates, subsamples of interest, and weights from complex survey data. When calculating ARR and ARD for continuous variables of interest, any two policy-relevant values can be chosen over which to estimate these measures. Estimating ARR and ARD for a particular subsample also changes the sample size over which the predicted probabilities are calculated. Introducing complex survey weights alters the equations for ARR and ARD such that the individual predicted probabilities are weighted accordingly before being averaged across the sample.

Calculating ARR and ARD is not restricted to logistic regression with a binary outcome. Our research has extended the application of these measures to both logit and probit models with binary, multinomial, and ordered outcomes. Below, we focus on the estimation of ARR and ARD after running a multinomial logistic regression as a representative example.

2.2 Multinomial logistic regression

Multinomial logit models have outcomes that consist of multiple, discrete, and unordered categories. Examples of this type of outcome are kinds of health insurance, modes of transportation, and college majors. In the multinomial logit model, the formula for the probabilities of each category for categories $j = 1, \dots, J - 1$ is

$$\Pr(y = j, j \neq J | X, x) = \frac{e^{(\beta_x^j x + X\beta^j)}}{\sum e^{(\beta_x^j x + X\beta^j)} + 1}$$

and the probability of the J^{th} category is

$$\Pr(y = J|X, x) = \frac{1}{\sum e^{(\beta_x^J x + x\beta^J)} + 1}$$

After running a multinomial logit model, we can compute J distinct ARR and ARDs, one for each outcome category. Compute the predicted probabilities for any or all of the outcome categories, changing the explanatory variable of interest as described in section 2.1. Then take the ratio for the ARR or the difference for the ARD.

2.3 Complex survey design

Complex survey design affects coefficient estimation and their standard errors. In turn, complex survey design affects ARRs, ARDs, and their standard errors. Incorporating elements such as stratification, clustering, and weighting in the estimation of a model are necessary to generate accurate measures of ARRs and ARDs.

3. Calculating ARRs and ARDs after running nonlinear models

In this section we will use data from the National Survey of Children's Health (NSCH) to demonstrate the calculation and interpretation of ARRs and ARDs after running a multinomial logit model. It is important to note that these models are for illustrative purposes only, and the reader is not to infer causality from the results from these simple models.

A program created for Stata generates estimates of ARRs, ARDs, and delta-method standard errors for covariates of interest after running a specified model. Our program automatically incorporates complex survey design, estimates these measures for subsamples of interest, and runs for logit and probit models with binary, multinomial, and ordered outcomes. This Stata program is available upon request.

3.1 NSCH data

The National Survey of Children's Health (NSCH) is sponsored by the Child and Maternal Health Bureau of the U.S. Department of Health and Human Services. This telephone survey is administered nationwide by the National Center for Health Statistics using the State and Local Area Integrated Telephone Survey (SLAITS) and collects health and health care information about children under 18 years old. First administered in 2003, this survey is conducted every four years.

We use the 2007 NSCH to illustrate the calculation and interpretation of adjusted risk ratios (ARR) and adjusted risk differences (ARD) in a multinomial logistic regression. In our model the outcome of interest is the primary location where the sampled child receives health care ("*usc*" in our model, representing usual source of care). We modified this variable such that the four most prevalent locations were maintained as separate locations, and we dropped all other locations (consisting of less than one percent of the responses). The primary explanatory variable of interest is the child's type of health insurance. In the NSCH, insurance status is divided into private insurance, public insurance, and no insurance. Approximately 66 percent of the sample has private insurance, 26.8 percent of the sample has public insurance, and 7.2 percent are uninsured.

The set of explanatory variables in our model includes the health insurance variables and a basic set of demographic variables. We construct age categories to correspond to children in different developmental and educational phases. Children up to age four consist of one category, children ages five to eleven create the second age category, and children ages twelve to seventeen make up the final category. Individuals report their race as being white, black, multi-racial, or other. Due to a sizeable number of individuals not reporting their race, a separate variable is constructed that denotes the race variable is missing for those observations. The race and ethnicity variables are constructed similarly; these variables equal one if the child is identified as part of a particular race or ethnicity and equals zero otherwise. The resulting sample size of variables with non-missing values on the variables of interest is 86,913 observations.

The NSCH was administered with a complex sampling structure. Observations were stratified at the state level, and the primary sampling unit is the household. Probability weights are used in the analysis to generate a sample that is nationally representative (Blumberg et al., 2007). The following tables show summary statistics for the explanatory variables and then for the dependent variable.

	Mean	Min	Max
ins_pub	.2862582	0	1
ins_uni	.0808009	0	1
age5_11	.3813833	0	1
age12_17	.3422308	0	1
female	.4902354	0	1
race_bl	.1440758	0	1
race_multi	.0494847	0	1
race_oth	.0522108	0	1
race_missing	.092276	0	1
hispanic	.1873219	0	1

. tab usc

Place of health care	Freq.	Percent	Cum.
1 Doctor's office	68,219	78.49	78.49
2 Hospital emergency room	1,184	1.36	79.85
3 Hospital outpatient department	1,880	2.16	82.02
4 Clinic or health center	15,630	17.98	100.00
Total	86,913	100.00	

3.2 Multinomial logit model

We estimate a multinomial logistic model to predict the probability of each of four types of usual sources of care, as a function of the insurance status and demographics. Because there are four categorical outcomes, the model estimates three parameters for each explanatory variable.

```
. svy: mlogit usc i.ins_pub i.ins_uni age5_11 age12_17 i.female race_bl race_multi
race_oth race_missing hispanic
(running mlogit on estimation sample)
```

Survey: Multinomial logistic regression

Number of strata	=	51	Number of obs	=	86913
Number of PSUs	=	86913	Population size	=	68905676
			Design df	=	86862
			F(30, 86833)	=	47.33
			Prob > F	=	0.0000

usc	Linearized			t	P> t	[95% Conf. Interval]
	Coef.	Std. Err.				

1_Dr's office (base outcome)						

2_Hosp ER						
1.ins_pub	1.142589	.2066544	5.53	0.000	.737548	1.54763
1.ins_uni	1.992841	.2199382	9.06	0.000	1.561764	2.423918
age5_11	.1875998	.1922636	0.98	0.329	-.1892351	.5644347
age12_17	.3488113	.1748101	2.00	0.046	.0061851	.6914375
1.female	.0495236	.1536241	0.32	0.747	-.2515783	.3506254
race_bl	1.433909	.1779922	8.06	0.000	1.085045	1.782772
race_multi	-.0516816	.2692711	-0.19	0.848	-.5794506	.4760875
race_oth	1.031731	.4342249	2.38	0.018	.180654	1.882808
race_missing	.6750487	.3171608	2.13	0.033	.0534164	1.296681
hispanic	.9825168	.2229067	4.41	0.000	.5456216	1.419412
_cons	-5.313072	.1761959	-30.15	0.000	-5.658414	-4.96773

3_Hosp OPD						
1.ins_pub	.5343329	.1416113	3.77	0.000	.256776	.8118898
1.ins_uni	.8920434	.2027603	4.40	0.000	.4946349	1.289452
age5_11	.035345	.1694218	0.21	0.835	-.2967202	.3674103
age12_17	.2538672	.1727587	1.47	0.142	-.0847384	.5924727
1.female	.0106483	.1328544	0.08	0.936	-.2497452	.2710418
race_bl	1.184395	.1304534	9.08	0.000	.9287077	1.440083
race_multi	1.142207	.3137291	3.64	0.000	.5273005	1.757113
race_oth	1.727319	.2241178	7.71	0.000	1.28805	2.166588
race_missing	1.220938	.2280895	5.35	0.000	.7738846	1.667992
hispanic	.261648	.1345126	1.95	0.052	-.0019954	.5252915
_cons	-4.514257	.1871833	-24.12	0.000	-4.881134	-4.147379

4_Clinic/Cntr						
1.ins_pub	1.017277	.0571559	17.80	0.000	.9052515	1.129301
1.ins_uni	1.296523	.0825516	15.71	0.000	1.134722	1.458323
age5_11	-.0454325	.0642795	-0.71	0.480	-.1714198	.0805548
age12_17	.130426	.0649003	2.01	0.044	.0032219	.25763
1.female	.1138293	.0513247	2.22	0.027	.0132332	.2144253
race_bl	.31956	.0660122	4.84	0.000	.1901767	.4489433
race_multi	.0091794	.1241035	0.07	0.941	-.2340624	.2524213
race_oth	.4852826	.1148665	4.22	0.000	.2601453	.7104199
race_missing	.6239109	.1039805	6.00	0.000	.4201101	.8277117
hispanic	1.005137	.0758997	13.24	0.000	.8563738	1.153899
_cons	-2.463513	.0593282	-41.52	0.000	-2.579796	-2.34723

The coefficients from a multinomial model are hard to interpret directly. Instead, we calculate ARR and ARD for three covariates: public insurance, uninsured, and female. For each covariate, we calculate an ARR and ARD for each outcome. The delta-method standard errors are shown in parentheses. For the following output, the usual source of care for outcome 1 is the Doctor's office, for outcome 2 is the Hospital Emergency Room, for outcome 3 is a Hospital Outpatient Department, and for outcome 4 is a Clinic or Community Health Center.

```
. adjrr ins_pub
ARR(outcome 1) = 0.8067 (0.0108)
ARD(outcome 1) = -0.1605 (0.0094)
ARR(outcome 2) = 2.1979 (0.4253)
ARD(outcome 2) = 0.0170 (0.0047)
ARR(outcome 3) = 1.2918 (0.1721)
ARD(outcome 3) = 0.0061 (0.0032)
ARR(outcome 4) = 2.0180 (0.0837)
ARD(outcome 4) = 0.1373 (0.0091)
```

```
. adjrr ins_uni
ARR(outcome 1) = 0.6935 (0.0195)
ARD(outcome 1) = -0.2455 (0.0159)
```

```

ARR(outcome 2) = 4.1614 (0.8167)
ARD(outcome 2) = 0.0497 (0.0108)
ARR(outcome 3) = 1.5064 (0.2887)
ARD(outcome 3) = 0.0112 (0.0060)
ARR(outcome 4) = 2.1451 (0.1102)
ARD(outcome 4) = 0.1846 (0.0163)

```

```

. adjrr female
ARR(outcome 1) = 0.9817 (0.0090)
ARD(outcome 1) = -0.0144 (0.0072)
ARR(outcome 2) = 1.0163 (0.1491)
ARD(outcome 2) = 0.0003 (0.0029)
ARR(outcome 3) = 0.9845 (0.1259)
ARD(outcome 3) = -0.0004 (0.0029)
ARR(outcome 4) = 1.0853 (0.0406)
ARD(outcome 4) = 0.0145 (0.0066)

```

To interpret the insurance variables, we must remember that the omitted insurance variable in the multinomial model is private insurance; the results for the other insurance variables are understood with respect to this reference category. The results for the publicly insured and the uninsured have similar patterns as to where individuals receive care, with a larger effect size observed among the uninsured compared to the publically insured children.

The interpretation of the ARR for outcome 1 (care typically received at a doctor's office) by insurance status is straightforward. Those publicly insured are 19.33% less likely to receive care at a doctor's office than those with private insurance. On average, children who are uninsured are 30.65% less likely to receive care in this setting than the privately insured, holding all else equal. In terms of absolute differences, children who are publicly insured receive care from doctor's offices 16.05 percentage points less often than those privately insured. In comparison, the uninsured receive care from doctor's offices 24.55 percentage points less often than those privately insured. One can readily estimate that the uninsured receive care at private doctor's offices 8.5 percentage points (16.05 – 24.55) less often than those with public insurance.

The uninsured and the publicly insured receive care far more frequently in hospital emergency rooms and clinics and health centers than the privately insured. On average, uninsured children receive care in hospital emergency rooms approximately 316% more often than privately insured children. In comparison, publicly insured children receive care in this setting approximately 120% more often than privately insured children.

In examining the results from the variable *female*, girls and boys receive care at similar rates in hospital settings. Differences in location of care by sex emerge at doctor's offices (outcome 1) and clinics and health centers (outcome 4). Girls receive care from clinics and health centers 8.53% more than boys, on average. In comparison, boys receive care from doctor's offices 1.83% more than girls. In terms of absolute differences, girls are 1.45 percentage points more likely to go to clinics and health centers and they are 1.44 percentage points less likely to go to doctor's offices than boys, on average.

3.3 Complex survey design

In the above model, we specified the complex survey design of the NSCH before estimating the results. Identifying the stratification, sampling structure, and weighting of observations affects the model's results. If we had run the previous model without the survey commands, we would have estimated the following results for the ARRs and ARDs for the public insurance variable.

```

. adjrr ins_pub
ARR(outcome 1) = 0.8412 (0.0046)
ARD(outcome 1) = -0.1292 (0.0038)
ARR(outcome 2) = 2.6797 (0.1773)

```

ARD(outcome 2) = 0.0165 (0.0014)
ARR(outcome 3) = 1.4690 (0.0763)
ARD(outcome 3) = 0.0090 (0.0013)
ARR(outcome 4) = 1.6600 (0.0265)
ARD(outcome 4) = 0.1037 (0.0037)

Without incorporating the complex survey design, publicly insured children are estimated to receive care in doctor's offices, hospital emergency rooms, and hospital outpatient departments more frequently than they do. Publicly insured children also appear to receive care in clinics and health centers less frequently than the surveyed population does. In many cases, these differences are quite sizeable. Without adjusting for the appropriate survey characteristics, results may lead to different conclusions and policy implications than the data actually suggest.

4. Variance estimation

Although the literature has identified several standard ways to present results from logistic regression, typically as odds ratios or risk ratios, it has not settled on a way to calculate the variance associated with risk ratio estimates. We describe our investigation of the relative advantages of the two most common approaches to measuring variance as it applies to regression risk analysis: the delta method, which models a first-order Taylor-series expansion, and bootstrapping, a numeric (non-model based) approach to estimating variance.

The delta method expands a function of a random variable about its mean with Taylor-series approximation. The variation generated by the delta method comes from two sources: uncertainty in the estimated coefficients and differences in the distribution of covariates across the sample. Most major software programs automatically compute delta-method standard errors. When they are not automatically computed, the analyst must compute them, and this method can take time to debug but is fast to run.

In contrast, bootstrapping is relatively simple to program but can take a long time to run, particularly for large data sets. Confidence intervals are estimated directly from the bootstrap and may be asymmetric; standard errors may be estimated from the width of the confidence interval. Asymmetric confidence intervals are more appropriate when the probabilities are close to zero or one, when ceiling and floor effects bind.

As noted the ARR has two sources of variation—uncertainty in the estimated coefficients and differences in the distribution of covariates across the sample—that added together represent the total variance. The first component can be calculated from the covariance matrix of the estimated logistic regression coefficient and a vector that accounts for non-linearities in the ARR formula. The second component is the variance of the predicted risk given exposure, normalized by the average predicted risk given no exposure.

In future work, we will use Monte Carlo simulations that incorporate variation and confounding to demonstrate and specify the advantages and disadvantages of these two approaches to estimating uncertainty and variance for adjusted risk measures from regression risk analysis. We will also propose decision rules to help investigators choose how to calculate and describe uncertainty in specified circumstances.

5. Rank Reversal

Patient-centered outcomes research helps people make informed health care decisions and allows their voice to be heard in assessing the value of health care options. This research answers patient-focused questions:

- “Given my personal characteristics, conditions and preferences, what should I expect will happen to me?”
- “What are my options and what are the benefits and harms of those options?”
- “What can I do to improve the outcomes that are most important to me?”
- “How can the health care system improve my chances of achieving the outcomes I prefer?”

To answer these questions, patient-centered outcomes research assesses the benefits of interventions. It focuses on outcomes that people notice and care about such as survival, function, symptoms, and health-related quality of life; and incorporates a wide variety of settings and diversity of participants (see www.pcori.org).

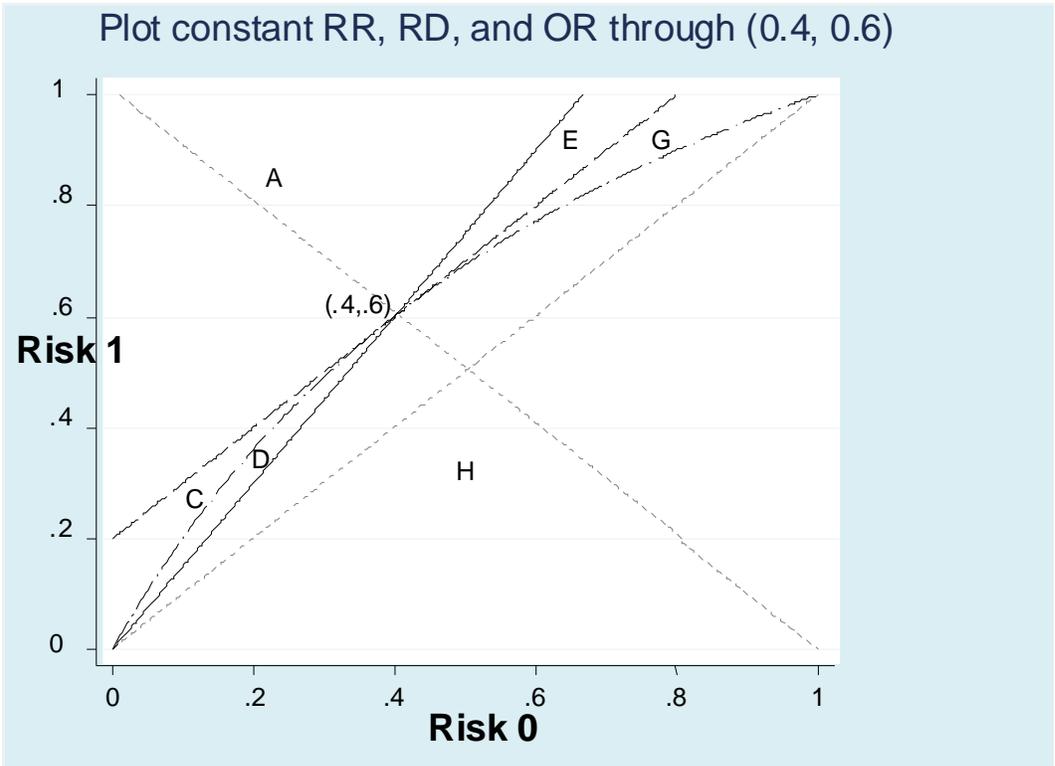
Comparative effectiveness research is the generation and synthesis of evidence that compares the benefits and costs of alternative methods to prevent, diagnose, treat, and monitor a clinical condition or to improve the delivery of care. The purpose of comparative effectiveness research is to assist consumers, clinicians, purchasers, and policy makers to make informed decisions that will improve health care at both the individual and population levels (see www.iom.org). The key elements of this definition are the comparison of effective interventions, the study of patients in typical day-to-day clinical care, and the aim of tailoring decisions to the needs of individual patients.

Indirect comparisons are used when one wishes to compare the relative size of the effect of an intervention used in two studies, each of which has its own baseline or comparison. Indirect comparisons are frequently used in comparative effectiveness research to rank order the magnitude of effect of several interventions. We will show that such rankings are dependent upon the choice of outcomes measure beyond what has been previously described. The rank order is dependent upon the specific choice of relative measure, odds ratio versus risk ratio. We will demonstrate that this phenomenon occurs and explain it.

Even with more well-controlled evidence, decisions must still balance benefits, risks and costs in the context of uncertainty. Uncertainty lingers for a variety of reasons: samples in trials are non-representative; there is limited patient heterogeneity; many studies are underpowered for secondary endpoint; and there is often a short duration of follow-up. Attempting to diminish these uncertainties with statistical control most commonly employs logistic regression reporting odds ratios even though patient-centered outcomes are often common.

The phenomenon that we describe—rank reversal—does not occur when comparisons are direct, that is when treatments A and B are compared to same Baseline (Risk 0). In this circumstance there will be no rank reversal. This property can be proved with basic calculus as the first derivatives must all be positive with constant Risk 0. In contrast, for indirect comparisons in which Treatment A is compared to Baseline in Study 1 and Treatment B is compared to Baseline in Study 2, rank reversal may occur.

We can demonstrate this phenomenon using pairs of probabilities displayed on a unit square. First create a unit square and identify any point on that square. We can then identify distinct lines for which the value of RR, OR, and RD remain constant while passing through that point. We call these line isoquants. The figure demonstrates isoquants through the point (0.4, 0.6).



As you can see, these lines carve the plane into several distinct areas. Each area is defined by the level of agreement between the ranks as determined by a comparison of the rank of (0.4, 0.6) and the points in that area using the various measures, RR, OR, and RD.

Area	RR	RD	OR	Agree?
A	R1	R1	R1	Yes
B	R1	R1	R0	OR different
C	R1	R0	R1	RD different
D	R1	R0	R0	RR different
E	R0	R1	R1	RR different
F	R0	R1	R0	RD different
G	R0	R0	R1	OR different
H	R0	R0	R0	Yes

When summing the areas on the unit square, we find that 90.5 % of the area has no rank reversal, while the OR is the exception for 2.4%, the RR for 4.3%, and the RD for 2.4%.

Moving from simulations to real findings, we consider an excellent meta-analysis from the *Annals of Internal Medicine* (Matchar et al., 2008). This meta-analysis compiles the findings from a number of studies and summarizes them in a way that allows us to calculate both the odds and the risk ratio and the risk difference. We present these data highlighting five exemplar studies.

The table below clearly illustrates the phenomenon of rank reversal in the ranking of the effect size of the study. We recognize that such comparisons represent an artificial construct, but the mathematics of rank reversal is the same regardless of whether one wishes to compare the effect size shown by a study or the effectiveness of various interventions.

Study	RD	RR	OR	OR Rank	RR Rank	RD Rank
Mogensen	0.10	1.14	1.76	1	5	1
Saito	0.09	1.38	1.57	2	2	3
Hasford	0.10	1.24	1.51	3	3	1
Larochelle	0.03	1.39	1.43	4	1	7
Fogari	0.08	1.15	1.38	5	4	4

We extend current understanding by demonstrating empirically how and when the selection of measure may affect decisions that are based on evidence regarding comparative effectiveness or impact. Although our findings are equally applicable in public health and policy settings, for the sake of clarity we use the language and choose examples from comparative effectiveness research.

What has not been fully appreciated, and what we have demonstrated herein, is the extent to which the rank order of interventions may flip, depending upon whether or not they are ranked by RR or OR, and describe how and when the RD offers a complementary perspective. This is not simply a mathematical curiosity. Clinical and other decisions are being made every day, relying upon comparative effectiveness research that orders the effectiveness of treatments in ways that invite misinterpretation.

In conclusion, we demonstrate that the choice of measure is consequential when indirectly comparing effect size. Even when restricting ourselves to relative measures, the rank order may depend upon the choice of OR or RR, with the RD providing complementary information. We demonstrate these phenomena in the interest of enhancing the understanding of CER studies, for the failure to appreciate rank reversal may lead to incorrect conclusions.

6. Conclusion

Regression risk analysis (RRA) generates intuitive and straightforward measures for interpreting the results from logistic regressions. In this paper, we expand on the work of Kleinman and Norton (2009) by extending RRA to

other nonlinear probability models and models with complex survey design. Using data from the National Survey of Children's Health, we present the results of a simple multinomial logit model and illustrate the calculation and interpretation of ARR and ARD in this context. We emphasize the ease of estimating these measures using our Stata program. When calculating ARR and ARD, consideration must be given to the variance estimation associated with these measures. We examine two variance estimation approaches, the delta method and bootstrapping, and propose future research to simulate the relative benefits of each option. Finally, we demonstrate the importance of the risk measure chosen when communicating results where indirect comparisons are estimated. The potential for *rank reversal*, a phenomenon where the ranking of treatments is contingent upon the choice of risk measure, must be acknowledged when reporting results from comparative effectiveness research and making conclusions. RRA provides important tools for interpreting the results from nonlinear models. Understanding the benefits of these measures as well as identifying scenarios where treatment rankings may be sensitive to the measure chosen will result in more informed researchers and higher quality research.

7. References

- Altman, D.G., Deeks, J.J., and Sackett, D.L. (1998). Odds Ratios Should Be Avoided When Events Are Common [Letter]. *British Medical Journal*, **317**, p. 1318.
- Bier, V.M. (2001). On the State of the Art: Risk Communication to the Public. *Reliability Engineering & System Safety*, **71**, pp. 139-150.
- Blumberg, S.J., Foster, E.B., Frasier, A.M., et al. (2007). Design and Operation of the National Survey of Children's Health. National Center for Health Statistics. Vital Health Stat 1. Forthcoming.
- Greenland, S. and Holland, P. (1991). Estimating Standardized Risk Differences from Odds Ratios. *Biometrics*, **47**, pp. 319–322.
- Holland, P. (1989). A Note on the Covariance of the Mantel-Haenszel Log-Odds-Ratio Estimator and the Sample Marginal Rates. *Biometrics*, **45**, pp. 1009-1016.
- Hosmer, D.W. and Lemeshow, S. (1989). Applied Logistic Regression. New York: John Wiley & Sons.
- Klaidman, S. (1990). How Well the Media Report Health Risk. *Daedalus*, **119**, pp. 119-132.
- Kleinman, L.C. and Norton, E.C. (2009). What's the Risk? A Simple Approach for Estimating Adjusted Risk Ratios from Nonlinear Models Including Logistic Regression. *Health Services Research*, **44**, pp. 288-302.
- Lee, J. (1981). Covariance Adjustment of Rates Based on the Multiple Logistic Regression Model. *Journal of Chronic Diseases*, **34**, pp. 415–426.
- Matchar, D.B., McCrory, D.C., Orlando, L.A., and Patel, M.R. (2008). Systematic Review: Comparative Effectiveness of Angiotensin-converting Enzyme Inhibitors and Angiotensin II Receptor Blockers for Treating Essential Hypertension. *Annals of Internal Medicine*, **148**, pp. 6–29.
- McNutt, L., Wu, C., Xue, X., and Hafner, J.P. (2003). Estimating the Relative Risk in Cohort Studies and Clinical Trials of Common Outcomes. *American Journal of Epidemiology*, **157**, pp. 940–943.
- Spiegelman, D. and Hertzmark, E. (2005). Easy SAS Calculations for Risk or Prevalence Ratios and Differences. *American Journal of Epidemiology*, **162**, pp. 199-200.
- Teuber, A. (1990). Justifying Risk. *Daedalus*, **119**, pp. 235-254.
- Zhang, J. and Yu, K.F. (1998). What's the Relative Risk?: A Method of Correcting the Odds Ratio in Cohort Studies of Common Outcomes. *Journal of the American Medical Association*, **280**, pp. 1690–1691.