Published in final edited form as:
*J Clin Epidemiol*. 2010 October ; 63(10): 1159–1163. doi:10.1016/j.jclinepi.2009.12.017.

# Cardiovascular outcomes ascertainment was similar using blinded and unblinded adjudicators in a national prospective study

**Gaurav Parmar, MD, MPH, CPH**[1], **Pallavi Ghuge, MD, MPH**[1], **Jewell H. Halanych, MD, MSc**[1], **Ellen Funkhouser, DrPH**[1], and **Monika M. Safford, MD**[1]

[1] University of Alabama at Birmingham, Birmingham, AL

## Abstract

**OBJECTIVE—**Observational studies can avoid biases by blinding medical records to characteristics of interest prior to outcomes adjudication. However, blinding is costly. We assessed the effect of blinding race and geography on outcomes ascertainment.

**STUDY DESIGN AND SETTING—**The REasons for Geographic And Racial Differences in Stroke-Myocardial Infarction (REGARDS-MI) Study is an ancillary study to the REGARDS national prospective cohort study including 30,228 participants. The primary characteristics of interest are race and geography, and the prespecified acceptable agreement rate between adjudicators is set at > 80%. We selected 116 suspected cardiovascular events that underwent adjudication with usual blinding. At least 3 months later, cases were readjudicated without blinding race and geographic location of the patient. We assessed differences in outcomes ascertainment using Cohen s κ statistic and agreement rates.

**RESULTS—**Agreement between the blinded and unblinded reviews was good to excellent for all four outcomes. κ statistics were 0.80 (chest pain), 0.85 (heart failure), 0.86 (revascularization) and 0.74 (MI) ($p<0.0001$ for all). Within each outcome, agreement rates were similar for race and geographic groups (agreement 83–100%).

**CONCLUSION—**In observational studies, blinding medical record review for outcomes ascertainment for some types of patient characteristics may be an unwarranted expense.

### Keywords

Blinding; Quality; Unblinding; Masking; Hospital records review; Medical records; disparities

## Introduction

Unbiased outcomes assessment is an important objective in observational research studies. Knowing the predictor status of participants at the time of outcomes assessment could introduce biases that are often subconscious [1]. Blinding investigators to the primary predictors of interest can avoid introducing such biases [2].

Corresponding Author address: Gaurav M Parmar, Medical Towers 520 B, 1717 11th Avenue South, Birmingham, AL 35294-4410, Phone: (205)934-5923 Fax: (205)974-7405, gparmar@dopm.uab.edu.

However, when predictors relate to patient characteristics that are ubiquitous in hospital records used to detect outcomes, considerable resources may be required to achieve adequate blinding. Two such characteristics are race and geography. In fact, some studies with a strong interest in racial differences have not blinded to characteristics such as race, for example, the Women s Health Initiative [3–5] and the Translating Research Into Action for Diabetes study [6,7]. Empiric studies that have evaluated biases introduced by unblinded endpoint reviews have not been reported to our knowledge, but such knowledge could guide resource utilization decisions.

We report the agreement between blinded and unblinded medical record review for cardiovascular outcomes ascertainment in a large, prospective cohort study whose main predictors of interest are race and geography.

## Methods

The REasons for Geographic And Racial Differences in Stroke-Myocardial Infarction (REGARDS-MI) Study is an ancillary study to the REGARDS national cohort study, and is following 30,228 African Americans and European Americans prospectively for cardiovascular events. Details of the study are described elsewhere [8]. Briefly, recruitment was conducted from 2003–2007 using commercially available lists and a combination of mail and telephone contact to recruit English-speaking, community-dwelling adults aged 45 and older living in the 48 contiguous US. By design, half of the sample was recruited from the 8 Stroke Belt states (North Carolina, South Carolina, Georgia, Tennessee, Mississippi, Alabama, Louisiana and Arkansas). Baseline data collection included telephone surveys and in-home exams, and living participants are telephoned every 6 months and asked if they were hospitalized for a stroke or a heart-related condition. Medical records are then retrieved for hospitalized chest pain, MI, revascularization and heart failure. The study was approved by the Institutional Review Board of the University of Alabama at Birmingham.

Endpoint ascertainment is modeled on other studies including the Look Ahead Study [9], Women s Health Initiative [3], the Atherosclerosis Risk in Communities Study [10] and the Multi-Ethnic Study of Atherosclerosis [11]. Each potential event is reviewed independently by 2 experts, and disagreements are resolved by consensus by committee. If individual adjudicator agreement with the final outcome falls below 80%, retraining is undertaken. Adjudication uses a standardized approach, and at the time of this study, the team had been calibrated so that agreement rates had been >80% for each of the 8 adjudicators for at least one year.

Each medical record was blinded to race and geography manually, requiring on average 50 minutes per record. Geographic location required the most time because it appeared on most pages of the record (including participant and hospital addresses, zip codes, name of hospital, area code in phone numbers, etc.). Because of the ubiquitous nature of these data elements in hospital records, double staff review was necessary to achieve consistent results and only occasional errors.

A single group, uncontrolled, pre-post design was used to analyze the effect of blinding. We selected 116 available medical records that had been adjudicated using the blinded approach, independently adjudicated by 2 of the 8 adjudicators. These records were then readjudicated without blinding by one of the 2 adjudicators that originally reviewed the case at least 3 months later.

We reported agreement rates (AR) [(a+d) / (a+b+c+d)] (Table 1) for all four major outcomes. Within each outcome, we also reported agreement rates among African Americans, European Americans, Southerners and non-Southerners. As agreement rates do

not take into account the agreement that would have been expected due solely to chance [12,13], the level of agreement between the two reviews was also calculated using Cohen s κ statistic for each outcome separately. This statistic adjusts for agreement occurring by chance with possible values between −1 to +1 where '0' can be interpreted as no agreement above that expected by chance, '−1' means complete disagreement and '+1' means almost perfect agreement. Based on Landis-Koch and Fleiss JL, agreement was classified as poor (κ <0.40), fair to good (κ 0.40 to 0.75), and excellent (κ > 0.75) [12,14]. We also calculated Positive Percent Agreement (PPA) [a / (a+c)], Negative Percent Agreement (NPA) [d / (b +d)] (see table 1) considering blinded review as reference. Percent agreements were calculated because κ does not take into account the degree of disagreement (cells b and c in table-1), that is, all disagreement is treated equally as total disagreement [15,16]. PPA and NPA are analogous to sensitivity and specificity of a screening test respectively. Data were analyzed using SAS version 9.1 (Cary, North Carolina).

## Results

Among the 116 cases, 51 (44%) were from African American patients and 87 (75%) were from the Southern region. The κ statistics for agreement between blinded and unblinded reviews were 0.80 for chest pain, 0.85 for heart failure, 0.86 for coronary revascularization and 0.74 for myocardial infarction (Table 2). Thus, the agreement between the blinded and unblinded reviews was in the good to excellent range for all outcomes [17,18]. The standard errors for these κ statistics were near 0.06 for all, and p-values were all < 0.0001 (Table 2).

All outcomes were combined to see the maximum effect of misclassification and it revealed no difference with overall κ statistic being 0.82 with same AR, PPA and NPA of 91% (Table 2).

Within each outcome, the AR, PPA and NPA were similar across race and geographic strata for all four outcomes, and all were >80% (Table 3).

## Discussion

This study found no evidence that review of unblinded medical records for race and geography predictors introduced biases in cardiovascular outcomes ascertainment. The considerable resources required to blind medical records to these characteristics may not be warranted. Cost savings could be considerable, as our estimate of 50 minutes per record translates to nearly 2 full-time equivalent staff in a study the size of the REGARDS study.

Given the resources required, surprisingly little empiric evidence exists to justify the practice of blinding to sociodemographics in prospective cohort studies. Our findings suggest that reports on racial differences in cardiovascular outcomes from studies like the Women s Health Initiative, and the Translating Research Into Practice for Diabetes studies, which did not blind records to race, are unlikely to have been biased.

We discussed the practice of blinding with the adjudicators after the study was completed to obtain their qualitative feedback. They reported that whereas the complexity of the task of ascertaining the outcome quickly absorbed their attention, they also noted that areas in blinded records sometimes caused attention to rest where it normally would not. For example, most records describe European Americans as being "white", but use the much longer word "African American" to describe blacks. The dominant black mark over these words was often difficult to overlook. Therefore, blinding may inadvertently call attention to the race of the individual, resulting in the exact opposite of its desired intent.

Our study was small and limited in scope due to practical constraints. Moreover, findings may not generalize to other CVD outcomes. A larger number of each outcome may have provided more stable estimates. Nevertheless, the consistent pattern of high agreement across the outcomes and the strata is noteworthy. The misclassification we observed in this study was consistent with past experience in major epidemiologic studies such as the Women s Health Initiative, and was above the pre-specified goal of ≥80% agreement in our study. As in all observational studies, caution in ascribing the observed effects to blinding is prudent.

## Conclusion

This study provides preliminary evidence that outcomes ascertainment for cardiovascular endpoints may be similar for blinded or unblinded medical records when race and geography are the main characteristics of interest. If confirmed, the results suggest that removal of blinding in similar settings could save significant time and human resources.

## Acknowledgments

Support provided by R01 HL80477 from the National Heart, Lung and Blood Institute, and by a cooperative agreement U01 NS041588 from the National Institute of Neurological Disorders and Stroke, National Institutes of Health, Department of Health and Human Services. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Neurological Disorders and Stroke or the National Institutes of Health. The authors acknowledge the participating REGARDS investigators and institutions for their valuable contributions: The University of Alabama at Birmingham, Birmingham, Alabama (Study PI, Statistical and Data Coordinating Center, Survey Research Unit): George Howard DrPH, Leslie McClure PhD, Virginia Howard PhD, Libby Wagner MA, Virginia Wadley PhD, Rodney Go PhD, Monika Safford MD, Ella Temple PhD, Margaret Stewart MSPH, J. David Rhodes BSN; University of Vermont (Central Laboratory): Mary Cushman MD; Wake Forest University (ECG Reading Center): Ron Prineas MD, PhD; Alabama Neurological Institute (Stroke Validation Center, Medical Monitoring): Camilo Gomez MD, Susana Bowling MD; University of Arkansas for Medical Sciences (Survey Methodology): LeaVonne Pulley PhD; University of Cincinnati (Clinical Neuroepidemiology): Brett Kissela MD, Dawn Kleindorfer MD; Examination Management Services, Incorporated (In-Person Visits): Andra Graham; Medical University of South Carolina (Migration Analysis Center): Daniel Lackland DrPH; Indiana University School of Medicine (Neuropsychology Center): Frederick Unverzagt PhD; National Institute of Neurological Disorders and Stroke, National Institutes of Health (funding agency): Claudia Moy PhD.

## References

1. Day SJ, Altman DG. Statistics notes: blinding in clinical trials and other studies. BMJ 2000 Aug 19–26;321(7259):504. [PubMed: 10948038]
2. Grimes DA, Schulz KF. Cohort studies: marching towards outcomes. Lancet 2002 Jan 26;359(9303):341–5. [PubMed: 11830217]
3. Curb JD, McTiernan A, Heckbert SR, Kooperberg C, Stanford J, Nevitt M, Johnson KC, Proulx-Burns L, Pastore L, Criqui M, Daugherty S. WHI Morbidity and Mortality Committee. Outcomes ascertainment and adjudication methods in the Women's Health Initiative. Ann Epidemiol 2003 Oct;13(9 Suppl):S122–8. [PubMed: 14575944]
4. Pradhan AD, Manson JE, Rossouw JE, Siscovick DS, Mouton CP, Rifai N, Wallace RB, Jackson RD, Pettinger MB, Ridker PM. Inflammatory biomarkers, hormone replacement therapy, and incident coronary heart disease: prospective analysis from the Women's Health Initiative observational study. JAMA 2002 Aug 28;288(8):980–7. [PubMed: 12190368]
5. Hsia J, Aragaki A, Bloch M, LaCroix AZ, Wallace R. WHI Investigators. Predictors of angina pectoris versus myocardial infarction from the Women's Health Initiative Observational Study. Am J Cardiol 2004 Mar 15;93(6):673–8. [PubMed: 15019867]
6. Brown AF, Gregg EW, Stevens MR, Karter AJ, Weinberger M, Safford MM, Gary TL, Caputo DA, Waitzfelder B, Kim C, Beckles GL. Race, ethnicity, socioeconomic position, and quality of care for adults with diabetes enrolled in managed care: the Translating Research Into Action for Diabetes (TRIAD) study. Diabetes Care 2005 Dec;28(12):2864–70. [PubMed: 16306546]

7. McEwen LN, Kim C, Karter AJ, Haan MN, Ghosh D, Lantz PM, Mangione CM, Thompson TJ, Herman WH. Risk factors for mortality among patients with diabetes: the Translating Research Into Action for Diabetes (TRIAD) Study. Diabetes Care 2007 Jul;30(7):1736–41. Epub 2007 Apr 27. [PubMed: 17468353]
8. Howard VJ, Cushman M, Pulley L, Gomez CR, Go RC, Prineas RJ, Graham A, Moy CS, Howard G. The reasons for geographic and racial differences in stroke study: objectives and design. Neuroepidemiology 2005;25(3):135–43. [PubMed: 15990444]
9. Ryan DH, Espeland MA, Foster GD, Haffner SM, Hubbard VS, Johnson KC, Kahn SE, Knowler WC, Yanovski SZ. Look AHEAD Research Group. Look AHEAD (Action for Health in Diabetes): design and methods for a clinical trial of weight loss for the prevention of cardiovascular disease in type 2 diabetes. Control Clin Trials 2003 Oct;24(5):610–28. [PubMed: 14500058]
10. The ARIC investigators. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. Am J Epidemiol 1989 Apr;129(4):687–702. [PubMed: 2646917]
11. Bild DE, Bluemke DA, Burke GL, Detrano R, Diez Roux AV, Folsom AR, Greenland P, Jacob DR Jr, Kronmal R, Liu K, Nelson JC, O'Leary D, Saad MF, Shea S, Szklo M, Tracy RP. Multi-ethnic study of atherosclerosis: objectives and design. Am J Epidemiol 2002 Nov 1;156(9):871–81. [PubMed: 12397006]
12. Fleiss, JL. Statistical methods for Rates and Proportions. 2. New York: John Wiley and Sons; 1981. p. 216p. 218
13. Hunt RJ. Percent agreement, Pearson's correlation, and kappa as measures of inter-examiner reliability. J Dent Res 1986 Feb;65(2):128–30. [PubMed: 3455967]
14. PMLandis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977 Mar;33(1):159–74. [PubMed: 843571]
15. Feinstein AR, Cicchetti DV. High agreement but low kappa: I. The problems of two paradoxes. J Clin Epidemiol 1990;43(6):543–9. [PubMed: 2348207]
16. Cicchetti DV, Feinstein AR. High agreement but low kappa: II. Resolving the paradoxes. J Clin Epidemiol 1990;43(6):551–8. [PubMed: 2189948]
17. Hollander JE, Go S, Lowery DW, Wolfson AB, Pollack CV, Herbert M, Mower WR, Hoffman JR. Interrater reliability of criteria used in assessing blunt head injury patients for intracranial injuries. Acad Emerg Med 2003 Aug;10(8):830–5. [PubMed: 12896882]
18. Stevens MW, Gorelick MH, Schultz T. Interrater agreement in the clinical evaluation of acute pediatric asthma. J Asthma 2003 May;40(3):311–5. [PubMed: 12807175]

**Table 1**

Agreement possibilities for comparison of blinded and unblinded review.

| | | Blinded Review | |
|---|---|---|---|
| | | **Outcome Present** | **Outcome Absent** |
| Unblinded Review | Outcome Present | A | B |
| | Outcome Absent | C | D |

**Table 2**

Measures of agreement between blinded and unblinded review (n=116)

| | A (n) | B (n) | C (n) | D (n) | PPA* | NPA* | AR | Kappa *K (SE)* | *95% Confidence Interval* | *p-value* |
|---|---|---|---|---|---|---|---|---|---|---|
| Chest Pain | 75 | 5 | 5 | 31 | 94% | 86% | 91% | 0.80 (0.06) | 0.68 – 0.92 | <0.0001 |
| Myocardial Infarction | 35 | 5 | 9 | 67 | 80% | 93% | 88% | 0.74 (0.07) | 0.61 – 0.87 | <0.0001 |
| Heart Failure | 36 | 6 | 2 | 72 | 95% | 92% | 93% | 0.85 (0.05) | 0.75 – 0.95 | <0.0001 |
| Coronary Revascularization | 29 | 3 | 1 | 18 | 97% | 86% | 92% | 0.84 (0.08) | 0.68 – 0.99 | <0.0001 |
| ALL OUTCOMES COMBINED | 175 | 17 | 19 | 188 | 91% | 91% | 91% | 0.82 (0.03) | 0.76 – 0.88 | <0.0001 |

A = Both reviews identified outcome present

B = Unblinded review identified outcome present and blinded review identified outcome absent.

C = Blinded review identified outcome present and unblinded review identified outcome absent.

D = Both reviews identified outcome absent

PPA: Positive Percent Agreement = [A / (A+C)] * 100 (see Table 1)

NPA: Negative Percent Agreement = [D / (B+D)] * 100 (see Table 1)

AR: Percent Agreement Rate = [(A+D) / (A+B+C+D)] * 100 (see Table 1)

SE: Standard Error of measurement

* Blinded outcomes were used as reference

## Table 3

Measures of agreement stratified by race and geographic location (n=116)

| | PPA* | NPA* | AR | Kappa *K (SE)* |
|---|---|---|---|---|
| CHEST PAIN | | | | |
| RACE: | | | | |
| African American (n=51) | 97% | 86% | 92% | 0.84 (0.08) |
| European American (n=65) | 92% | 87% | 91% | 0.75 (0.10) |
| GEOGRAPHY | | | | |
| South (n=87) | 95% | 88% | 93% | 0.84 (0.06) |
| Non-South (n=29) | 89% | 80% | 86% | 0.69 (0.14) |
| HEART FAILURE | | | | |
| RACE: | | | | |
| African American (n=51) | 100% | 87% | 92% | 0.84 (0.08) |
| European American (n=65) | 89% | 96% | 94% | 0.85 (0.07) |
| GEOGRAPHY | | | | |
| South (n=87) | 93% | 92% | 92% | 0.82 (0.06) |
| Non-South (n=29) | 100% | 95% | 97% | 0.92 (0.07) |
| CORONARY REVASCULARIZATION | | | | |
| RACE: | | | | |
| African American (n=51) | 71% | 98% | 94% | 0.74 (0.14) |
| European American (n=65) | 100% | 93% | 95% | 0.90 (0.06) |
| GEOGRAPHY | | | | |
| South (n=87) | 91% | 94% | 93% | 0.83 (0.07) |
| Non-South (n=29) | 100% | 100% | 100% | 1.00 (0.00) |
| MYOCARDIAL INFARCTION | | | | |
| RACE: | | | | |
| African American (n=51) | 76% | 97% | 90% | 0.77 (0.10) |
| European American (n=65) | 81% | 89% | 86% | 0.71 (0.09) |
| GEOGRAPHY | | | | |
| South (n=87) | 85% | 92% | 90% | 0.78 (0.07) |
| Non-South (n=29) | 60% | 95% | 83% | 0.59 (0.16) |

PPA: Positive Percent Agreement = [A / (A+C)] * 100 (see Table 1)

NPA: Negative Percent Agreement = [D / (B+D)] * 100 (see Table 1)

AR: Percent Agreement Rate = [(A+D) / (A+B+C+D)] * 100 (see Table 1)

SE: Standard Error of measurement

* Blinded outcomes were used as reference