Racial Disparities in Operative Procedure Time

The Influence of Obesity

Jeffrey H. Silber, M.D., Ph.D.,* Paul R. Rosenbaum, Ph.D.,† Richard N. Ross, M.S.,‡
Orit Even-Shoshan, M.S.,§ Rachel R. Kelz, M.D.,‖ Mark D. Neuman, M.D.,#
Caroline E. Reinke, M.D.,** Justin M. Ludwig, M.A.,†† Fabienne A. Kyle, B.A.,‡‡
Dale W. Bratzler, D.O., M.P.H.,§§ Lee A. Fleisher, M.D.‖‖‖

ABSTRACT

Background: Using Pennsylvania Medicare claims from 1995 to 1996, the authors previously reported that anesthesia procedure length appears longer in blacks than whites. In a new study using a different and larger data set, the authors now examine whether body mass index (BMI), not available in Medicare claims, explains this difference. The authors also examine the relative contributions of surgical and anesthesia times.

> ### What We Already Know about This Topic
>
> - Operations reportedly take longer in black than white Medicare patients
> - The authors asked whether body mass index explains the difference
>
> ### What This Article Tells Us That Is New
>
> - White patients with similar body mass index having general or orthopedic surgery were matched to comparable black patients within the same hospitals
> - After matching, blacks had induction-to-recovery room duration 7 min longer than whites ($P = 0.0019$) and 6-min reflect cut-to-close time ($P = 0.0032$)

* Professor, Center for Outcomes Research, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania; The Departments of Pediatrics and Anesthesiology and Critical Care, The University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania; Department of Health Care Management, The Wharton School, The University of Pennsylvania, Philadelphia, Pennsylvania; and The Leonard Davis Institute of Health Economics, The University of Pennsylvania, Philadelphia, Pennsylvania. † Professor, Department of Statistics, The Wharton School, The University of Pennsylvania, and The Leonard Davis Institute of Health Economics, The University of Pennsylvania. ‡ Senior Statistician, †† Statistical Programmer, ‡‡ Research Assistant, Center for Outcomes Research, The Children's Hospital of Philadelphia. § Associate Director, Center for Outcomes Research, The Children's Hospital of Philadelphia, and The Leonard Davis Institute of Health Economics, The University of Pennsylvania. ‖ Assistant Professor, ** Instructor, Department of Surgery, The University of Pennsylvania School of Medicine. # Assistant Professor, ‖‖‖ Professor, Department of Anesthesiology and Critical Care, The University of Pennsylvania School of Medicine, and The Leonard Davis Institute of Health Economics, The University of Pennsylvania. §§ Professor, Oklahoma University Health Sciences Center, College of Public Health and College of Medicine, Oklahoma City, Oklahoma, and Oklahoma Foundation for Medical Quality, Oklahoma City, Oklahoma.

Address correspondence to Dr. Silber: Center for Outcomes Research, The Children's Hospital of Philadelphia, 3535 Market Street, Suite 1029, Philadelphia, Pennsylvania 19104. silberj@wharton.upenn.edu. Information on purchasing reprints may be found at www.anesthesiology.org or on the masthead page at the beginning of this issue. ANESTHESIOLOGY's articles are made freely accessible to all readers, for personal use only, 6 months from the cover date of the issue.

Methods: The Obesity and Surgical Outcomes Study of 47 hospitals throughout Illinois, New York, and Texas abstracted chart information including BMI on elder Medicare patients (779 blacks and 14,596 whites) undergoing hip and knee replacement and repair, colectomy, and thoracotomy between 2002 and 2006. The authors matched all black Medicare patients to comparable whites and compared procedure lengths.

Results: Mean BMI in the black and white populations was 30.24 and 28.96 kg/m², respectively ($P < 0.0001$). After matching on age, sex, procedure, comorbidities, hospital, and BMI, mean white BMI in the comparison group was 30.1 kg/m² ($P = 0.94$). The typical matched pair difference (black–white) in anesthesia (induction to recovery room) procedure time was 7.0 min ($P = 0.0019$), of which 6 min reflected the surgical (cut-to-close) time difference ($P = 0.0032$). Within matched pairs, where the difference in procedure times was greater than 30 min between patients, blacks more commonly had longer procedure times (Odds = 1.39; $P = 0.0008$).

Conclusions: Controlling for patient characteristics, BMI, and hospital, elder black Medicare patients experienced slightly but significantly longer procedure length than their

> ⊕ Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are available in both the HTML and PDF versions of this article. Links to the digital files are provided in the HTML text of this article on the Journal's Web site (www.anesthesiology.org).

closely matched white controls. Procedure length difference was almost completely due to surgery, not anesthesia.

IN our previous work studying hospitals and patients in Pennsylvania, we reported that operative procedure length as measured through Medicare anesthesia claims seems longer in blacks than whites undergoing the same general and orthopedic surgical procedures.[1] After adjusting for the surgical procedure, patient comorbidities, and hospital, we previously observed a 5.5-min black–white difference in procedure length (95% CI, 3.8–7.1; $P < 0.0001$). This difference varied among hospitals; some institutions displayed a 16-min difference whereas others displayed no difference at all.[1] These observations had a number of possible explanations, with unobserved confounding being an important consideration.

The current article aims to address some weaknesses of the previous study. As Medicare claims did not record body mass index (BMI), we could not directly determine whether differences in BMI by race were causing the observed racial differences in procedure length as reported in our previous work. In addition, our previous work did not closely adjust for secondary procedures. It further used a regression approach that made assumptions about the form of the model used to estimate the racial differences. Last, our previous work did not examine the implications of the 5.5-min average disparity when the disparity may not be uniform among patients.

Using a new data source that includes chart abstraction as well as Medicare claims, we now ask whether there remains a difference in procedure time. In the current study, we carefully match on BMI (unobservable in our previous analysis); estimated procedure length that takes into consideration types of secondary procedures; comorbidities; source of admission; and the hospital where the procedure was performed. We also examined whether the apparent disparity is associated more with the anesthesiology team or the surgical team by asking whether the disparity occurs during the cut-to-close period of the procedure or before and after that period. Finally, we address the clinical importance of the size of the procedure length disparity observed in our analysis.

Materials and Methods

Study Overview

The current study reports on racial differences in procedure length using a special data set developed to examine the influence of obesity on surgical outcomes. It differs from our previous work on racial disparity in procedure time[1,2] in three important ways: (1) we used multivariate matching to compare racial differences in procedure length controlling for the hospital, rather than using m-estimation (a form of regression used in our previous work). The advantage of this approach is that we do not need to make assumptions about model form. (2) We concentrated on just five categories

of surgery (*vs.* 40 in our previous work), therefore, we can more precisely adjust for differences in procedure type; and (3) we augmented the typical Medicare claims data with chart-derived BMI information and recorded both anesthesia time (induction to recovery room)[2,3] and surgical time (cut-to-close).[2,3]

As previously described,[3–6] the Obesity and Surgical Outcomes Study comprised 47 hospitals throughout Illinois, New York, and Texas where Medicare claims data were merged with chart abstraction for general and orthopedic procedures. Medicare patients aged 65–80 yr were identified undergoing one of the following five types of surgery between 2002 and 2006: (1) hip replacement or revision excluding fracture (International Classification of Diseases, Ninth Revision, Clinical Modification [ICD-9-CM] Principal Procedure codes 81.51–81.53); (2) knee replacement or revision (ICD-9-CM Principal Procedure codes 81.54, 81.55); (3) colectomy for cancer (ICD-9-CM Principal Procedure codes 45.7–45.79, 45.8 and ICD-9-CM Principal Diagnosis codes 153–153.9, 154–154.8, 230.3–6); (4) colectomy not for cancer (ICD-9-CM Principal Procedure codes 45.7–45.79, 45.8 and ICD-9-CM Principal Diagnosis codes 562.1–562.13); and (5) thoracotomy (ICD-9-CM Principal Procedure codes 32–32.9).

Hospitals were contacted by the Oklahoma Foundation for Medical Quality and requested to abstract between 300 and 400 prespecified charts to collect baseline informations including BMI, admission vital signs, and laboratory tests, and information on the surgical procedure. All data collected were deidentified and merged with encrypted Medicare claims files and sent to the study investigators for analysis. Approval was obtained from The Children's Hospital of Philadelphia Institutional Review Board (the Institutional Review Board associated with the Principal Investigator of the study) as well as hospital-specific Institutional Review Boards when requested.

Statistical Analysis

Overview. The Obesity and Surgical Outcomes Study included 15,914 elder surgical patients in Medicare of which 779 (4.9%) were black. For each patient, we obtained procedure length through chart abstraction, using Medicare claims when occasional chart data elements were missing.[3]

The Matching Algorithm. Matching was performed using the algorithm MIPMatch.[7] We performed two matches, one without BMI and one with. Both matched exactly on hospital and procedure group, so that each pair of patients (1 black and 1 white) was matched within the same hospital and had exactly the same procedure group. MIPMatch allowed us to force black and white matched groups to have nearly the same frequencies of patient comorbidities, ICD-9-CM procedures within procedure groups, and gender (a requirement known as "near-fine balance").[8–10] Subject to the requirements of an exact match for procedure group and hospital, plus near-fine balance for comorbidities and procedures, we minimized the

total distance within matched pairs, a requirement known as optimal matching.[8] The matches achieved an exact match on ICD-9-CM principal procedure in over 95% of all pairs (see appendix, Supplemental Digital Content 1, http://links.lww.com/ALN/A922).

The distance included age, sex, procedure, comorbidities such as diabetes, heart failure, previous myocardial infarction, and arrhythmias (see appendix, Supplemental Digital Content 1, http://links.lww.com/ALN/A922), a propensity score for black race, a risk score[11] for death, and a predicted time score. The second match controlled for all of these variables and also included BMI. We used the propensity score as one of many variables to match on. Specifically, we found whites with propensity scores similar to those of blacks. It has been shown that when matching on the propensity score, one will also tend to match on the independent variables making up the propensity score.[12–14] Unlike the propensity score, which is computed with the study data, the risk score must be computed with an independent sample of patients outside the study population, in order to be able to compare outcomes after the matching process.[11]

The time score, like the risk score, was also based on a regression model to predict a patient's procedure time given their principal and secondary procedures, but not race. As time is continuous and the measure of procedure length from claims occasionally has large error (see the study by Silber *et al.*[3]), we used m-estimation.[15,16] As with the risk score, we fit the time score model only in patients who were not part of the study population because when we match on this variable, we want to compare times across black and white matched patients.

A second match was performed, which was similar to the first, but added BMI as a matching variable describing each patient. Likewise, we asked whether the procedure time difference we previously observed between blacks and whites would persist when we found white patients with very similar BMIs to their matched black patients. If the disparity in procedure length vanished, this would have suggest that differences in procedure length between blacks and whites previously reported were due to BMI differences, and not due to other potential causes of interest to policy makers concerned with disparities.

All data except BMI and procedure length were obtained from Medicare claims. BMI was abstracted from the chart, and procedure length will be reported based on our "best estimate" analysis,[3] which uses a measure of procedure length that combined the anesthesia bill with the abstracted length to produce a "best" measure of procedure length as described in our previous work.[3] Using only anesthesia claims produced very similar results for anesthesia procedure length (not shown), but claims do not include BMI or surgical cut-to-close time.

As has been suggested by Rubin *et al.*,[17,18] matching was performed first, without viewing outcomes. After matching was completed, we then analyzed procedure time differences across blacks and whites.

Statistical Tests

Balance on observed variables after matching was appraised using standard two-sample checks that contrast achieved balance with the magnitude of covariate balance anticipated from completely random assignment.[19] For each matching variable, we reported the "standardized difference" for group comparisons before and after matching, which represents the standardized mean difference among groups, using the SD of the pooled cases and controls.[19,20] For example, the standardized difference for age would be calculated as follows, where $\mu_{age,black}$ and $\mu_{age,white}$ are the mean ages of the black cases and matched white controls; $s^2_{age,black}$ and $s^2_{age, all white}$ are the variances of the black cases and all white potential controls. The standardized difference is then ($\mu_{age, black} - \mu_{age,white}$) divided by the square root of [($s^2_{age,black} + s^2_{age, all white}$)/2]. A usual rule of thumb is to try to achieve standardized differences below 0.2 or a fifth of an SD.[12,19–21]

We also compared covariate balance attained by matching with the covariate balance anticipated from complete randomization using two-sample randomization tests, specifically the Wilcoxon rank sum test for continuous covariates and the Fisher exact test for binary covariates.

When testing the hypothesis of no difference in outcomes between the matched black and white patients, the widely used Wilcoxon signed-rank statistic[22] was calculated, together with its corresponding CI and point estimate, the so-called Hodges–Lehmann estimate.[22] Also reported is the median. For binary outcomes, the McNemar statistic[23] was used. The Kruskal–Wallis test was used to examine differences in procedure time across procedure groups.[22] Findings were considered significant if *P* value is less than 0.05 (two-sided). We utilized the software package R for all statistical tests. R is available as Free Software under the terms of the Free Software Foundation's GNU General Public License in source code form.##

Results

The Quality of the Matches

Results on the quality of the two matches were displayed in table 1. We first present the quality of Match 1, the match that included age, sex, hospital, procedure, comorbidities, a risk score, a propensity score to be black, and a time score, but did not include BMI. We then present Match 2, a similar match that also includes BMI. As can be seen, the quality of the match did not vary with or without matching on BMI. Furthermore, the quality of the matches was uniformly excellent, as demonstrated by standardized difference after matching, which were all considerably less than 0.10 SDs for all variables and statistically insignificant.

Table 1. Matching Quality

Variable Name	Mean Cases/Black	Mean All Controls	Standardized Difference Before Match (SDs)
BMI	30.24	28.96	0.19
Age	72.05	72.78	−0.18
Sex	0.36	0.42	−0.13
Hip replacement	0.195	0.254	−0.14
Knee replacement	0.306	0.370	−0.14
Colectomy (for cancer)	0.250	0.145	0.27
Colectomy (no cancer)	0.089	0.073	0.06
Thoracotomy	0.161	0.159	0.00
Hx diabetes	0.38	0.22	0.35
Hx CHF	0.20	0.11	0.25
Hx MI	0.04	0.06	−0.08
Hx arrhythmia	0.20	0.22	−0.04
Risk of death	0.023	0.016	0.18
Propensity to be black	0.13	0.05	0.95
Time score, min	163.6	160.6	0.11

The two matches are presented here, without BMI (Match 1) and with BMI (Match 2), all matching exactly on the hospital. There were 779 cases (black patients), to which 779 white patients (controls) were selected among 14,596 whites (all controls). Standardized difference refers to the average difference in terms of units of SDs; standardized difference before matching compares the 779 black cases with all 14,596 white controls, standardized difference after matching compares the 779 black cases with the 779 white controls selected through matching. Risk score is the probability of death at 30 days and includes the following variables: age, sex, race, comorbidities, procedure, transfer-in status, and admission from ER. In Match 2, obesity status is also included in the risk score. Propensity score is the probability of being black and includes the following variables: age, sex, comorbidities, procedure, hospital, transfer-in status, and admission from ER. In the match with BMI, BMI is also included in the propensity score. Time score: reported in minutes. Based on principle procedure and secondary procedures from a model developed in patients in hospitals not in the 47 hospital Obesity and Surgical Outcomes Study. Procedure groupings were exactly matched for hip procedure, knee procedure, colectomy for cancer, colectomy not for cancer and thoracotomy.

* P value reflects the differences between cases and matched control groups using the Wilcoxon rank sum statistic for continuous variables, and the chi-square test for categorical variables.

BMI = body mass index; CHF = congestive heart failure; ER = emergency room; Hx = history; MI = miocardial infarction.

The Influence of BMI on Racial Differences in Procedure Length

In table 2, for both matches (Match 1 without matching on BMI and Match 2 including BMI), we examined the differences in length of procedure as defined by (1) a best estimate of anesthesia procedure time based on the chart and Medicare claim (induction to recovery room);[1–3] (2) surgical procedure time as determined from the chart (this is the cut-to-close time);[1–3] and finally, (3) anesthesia induction/emergence time which we defined as the difference between anesthesia procedure time (1) and surgical time (2). For each procedure time reported, we provided the mean, median, and the Hodges–Lehmann statistic which provides a measure of the typical time for each group, as well as a 95% CI.

The time differentials between black and white Medicare patients were similar across matches, and all differences in time between black and white matched pairs were highly significant. For example, using the match that did not include BMI and our best estimate of the anesthesia procedure time based on claim and chart, we observed that black patient procedures took about 6.5 min (95% CI, 2–10.5) longer than matched white patients, controlling for the same hospital, procedure, and comorbidities. When we included BMI into the matching algorithm (Match 2), we observed a similar finding of a 7-min difference

(95% CI, 2.5–11.5 min). We further examined whether procedure groups (knee, hip, colectomy for cancer, colectomy not for cancer, and thoracotomy) displayed different patterns of disparity. The black-minus-white pair differences in time were not significantly different among the five procedure groups (Kruskal–Wallis test: $P = 0.9319$ in non-BMI matched pairs and 0.4784 in BMI matched pairs). We next asked what would the disparity have been if we only exact matched on principal procedure, and no other variables. The Hodges–Lehmann estimate for the difference between blacks and whites in anesthesia time was 12.5 min (95% CI, 7–18; $P < 0.0001$), surgical time difference was 10.5 min (95% CI, 5.5–15.5; $P < 0.0001$), and induction/emergence time difference 1.5 min (95% CI, 0–3.5; $P = 0.0863$). Finally, the Obesity and Surgical Outcomes Study was not designed to make comparisons among hospitals, and the number of blacks in individual hospitals is too small to yield much statistical power for such a comparison. We did take the black-minus-white matched pair differences in operative time and compare them among hospitals using the Kruskal–Wallis test, and by this test the black–white difference in operative time did not differ significantly among hospitals (P value is 0.53 in the match with BMI and 0.72 in the match without BMI); however, it is difficult to know what to make of failing to find a difference when the power is low.

Table 1. *(Continued)*

Match 1 (without BMI)			Match 2 (with BMI)		
Mean Controls/White	Standardized Difference After Match (SDs)	P Value*	Mean Controls/White	Standardized Difference after Match (SDs)	P Value*
29.53	0.10	0.1113	30.07	0.03	0.9436
71.95	0.03	0.6859	71.90	0.04	0.5295
0.36	0.00	1.0000	0.36	0.00	1.0000
0.195	0.00	1.0000	0.195	0.00	1.0000
0.306	0.00	1.0000	0.306	0.00	1.0000
0.250	0.00	1.0000	0.250	0.00	1.0000
0.089	0.00	1.0000	0.195	0.00	1.0000
0.161	0.00	1.0000	0.161	0.00	1.0000
0.37	0.04	0.4963	0.36	0.05	0.4019
0.20	0.00	1.0000	0.18	0.03	0.6059
0.03	0.05	0.2883	0.04	0.01	1.0000
0.23	−0.08	0.1977	0.23	−0.07	0.2680
0.024	−0.03	0.8031	0.023	−0.01	0.6128
0.13	0.03	0.9721	0.13	0.01	0.7310
163.6	0.00	0.9339	163.6	0.00	0.9193

Which Components of the Operation Are Contributing to the Disparity?

When we examined racial differences in surgical time (cut-to-close time) and the anesthesia induction/emergence time (the difference between anesthesia procedure time and surgical time), we saw that racial differences in surgical time are almost equal to the differences in anesthesia procedure time (measuring induction to recovery room). For example, in table 2, using Match 2 with BMI, the Hodges–Lehmann estimate for anesthesia procedure time difference was 7 min in contrast to 6 min (95% CI, 2–9.5 min) for surgical time. The Hodges–Lehmann estimate of the black–white difference in anesthesia induction/emergence time, calculated from individual pairs as the difference between anesthesia time and surgical time, was 1 min (95% CI, −0.5 to 3 min).

Understanding the Clinical Importance of a 7-min Gap

One may reasonably ask whether a typical difference in procedure time of 7 min between black and white patients, although statistically significant, is clinically important. To better understand the implications of a time differential, we asked whether blacks were more likely to have longer procedure times than whites over various ranges of procedure

time differences. For example, it may be the case that there is always a gap between black and white patients, and that the typical 7-min gap is distributed rather evenly across patients. On the other hand, it may be the case that generally there is little difference between black and white procedure time, but in situations where there are large differences, blacks have longer times. This would be a more concerning pattern from a clinical perspective, suggesting that the average 7-min difference potentially signals a more important clinical problem.

To study this question, we first examined the unpaired distributions of procedure length in black and white patients. Figure 1 consists of two quantile–quantile plots[24] (one for each match) of black and white procedure times, omitting two extreme patients from each plot for display purposes. Points on the line of identity would suggest that the black and white quantiles were similar. We found that black patient procedure lengths were longer than white patients as the procedure lengths increased, suggesting that the typical black–white difference of seven is not distributed evenly across patients.

We then rank ordered all black–white matched pairs by the absolute value of the difference in procedure length within each pair. Some pairs had small differences, others large. For categories of absolute difference (between 0 and 10 min, 10 and 30 min, and greater than 30 min), we reported in table 3 the odds that

Table 2. Study Results: Time of Procedure by Race

Time, min	Anesthesia Procedure Time (Induction to Recovery Room)		Surgical Procedure Time (Cut-to-close)		Induction/Emergence Time (Anesthesia Time–Surgical Time)	
Black time						
Mean	170.9		123.1		47.8	
SD	67.1		62.3		21.4	
Median	160		110		45	
25th, 75th	(125, 199.5)		(80, 149.5)		(35, 57.5)	
Hodges–Lehmann	162.5		115		47	
95% CI	(158.5–167)		(111.5–119)		(46–48.5)	
	Match 1 Without BMI	Match 2 With BMI	Match 1 Without BMI	Match 2 With BMI	Match 1 Without BMI	Match 2 With BMI
White time						
Mean	162.7	163.1	115.4	116.1	47.3	47.0
SD	60.0	62.0	53.8	57.3	22.6	21.8
Median	150	150	102	102	45	45
25th, 75th	(120, 191)	(121, 191)	(75, 141)	(77, 141)	(35, 55)	(35, 56)
Hodges–Lehmann	156.5	156.0	109.5	109.5	46.5	46.5
95% CI	(152.5–160.5)	(152.0–160.0)	(105.5–113.0)	(106.0–113.5)	(45–47.5)	(45.5–48)
Black–white time						
Mean	8.1**	7.8**	7.7**	7.0*	0.5	0.8
SD	71.7	74.3	65.8	69.4	26.9	27.5
Median	6	5	6	5	0	0.5
25th, 75th	(–30, 39)	(–25, 43)	(–25, 37)	(–26, 39)	(–14, 15)	(–13, 15)
Hodges–Lehmann	6.5**	7**	6**	6**	0.5	1
95% CI	(2–10.5)	(2.5–11.5)	(2–9.5)	(2–9.5)	(–1, 2)	(–0.5, 3)

For each match (Match 1 without BMI and Match 2 with BMI), we present three time-based results: (1) anesthesia time (induction to recovery room); (2) surgical time (cut-to-close); and (3) induction/emergence time (which we define as the difference between anesthesia time and surgical time for each individual patient). The Hodges–Lehmann statistic provides a measure of the typical time for each group, as well as a 95% CI. This is more stable than the mean time. *P* values are provided to test whether the differences between black and white procedure times are different from zero.

* *P* < 0.01; ** *P* < 0.005.

BMI = body mass index.

the member of the pair with the longer time was a black patient (as compared with a white patient). For the match that did not include BMI, the overall odds was 1.22 (95% CI, 1.06–1.41; *P* = 0.0067). When time differences between members of a pair were greater than 30 min, the odds that the longer patient was black was 1.29 (1.07–1.56; *P* = 0.0070). Similarly, when BMI was included in the match, the overall odds that the black patient had the longer procedure time was 1.23 (1.07–1.42; *P* = 0.0047). When time differences between members of a pair were greater than 30 min, the odds that the longer patient was black was 1.39 (1.15–1.69; *P* = 0.0008). In brief, the 7-min typical gap between black and white matched pairs translated into an increased odds in which blacks have large (*i.e.,* >30 min) differences in procedure length than their matched white controls.

Discussion

In our previous research using Medicare claims,[1] we observed that the average difference in procedure time between black

and white patients was 5–7 min, and for some hospitals the difference was considerably greater. In that study, we used Medicare claims, and consequently could not adjust for the influence of BMI. However, obesity may increase the length of a procedure[1,25–27] and can be a challenge for both the anesthesia care team and the surgeon.[28–32] It would have been reassuring, from a disparities perspective, whether BMI differences between blacks and whites could explain the difference in procedure time. However, they did not.

Our current study on a different population of Medicare patients, using matching instead of regression and including chart review to obtain BMI, reports a disparity similar in magnitude to our previous report. Although the typical disparity of 7 min does not seem large, the processes that lead us to be able to observe these significant differences may be very different for blacks and whites. Any conclusions depend on a careful comparison of black and white patients, the procedures they had, and potential confounding factors not adjusted for in our previous work, such as
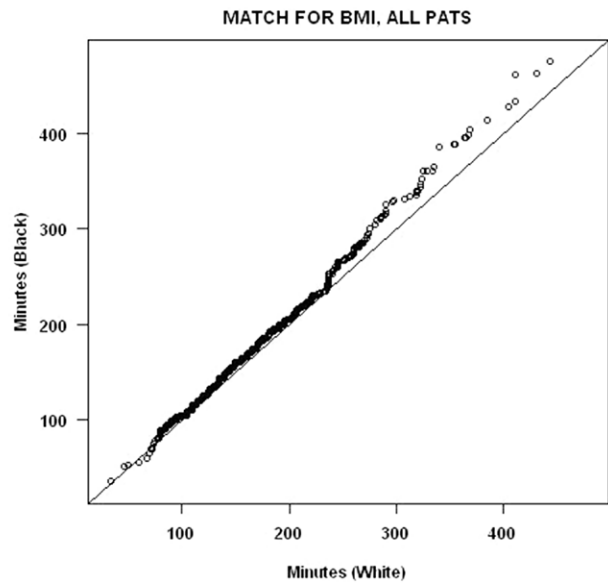
Match 1 QQ plot:
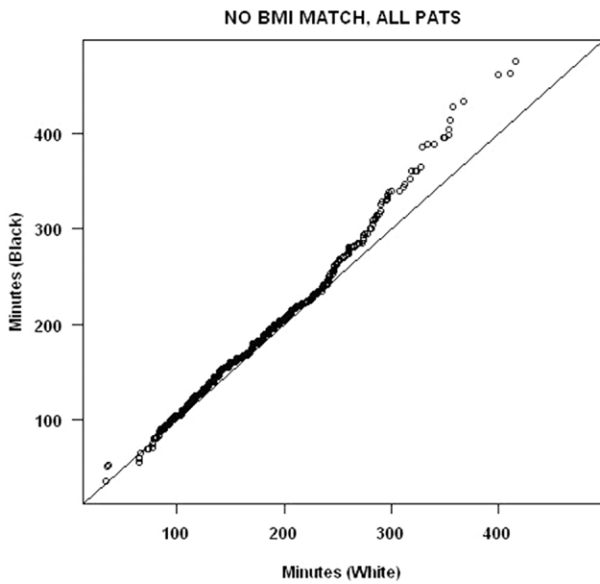
Match 2 QQ plot:

Fig. 1. Quantile–Quantile (QQ) plots. We present two QQ plots, one for Match 1(does not include BMI) and one for Match 2 (with BMI). For both plots, we rank all whites in order of total procedure length (*i.e.*, anesthesia time from induction to recovery room) on the x-axis and time for all blacks on the y-axis. Each point represents a white–black "pair" matched only on the quantile. For scaling purposes, two patients with outlier procedure times were removed from each plot. For both Match 1 and Match 2 samples, we observe the same pattern, a nonparallel divergence from the identity line, suggesting that the disparity we observe in procedure length does not reflect merely a constant increase in length for all black patients. BMI = body mass index; PATS = patients.

BMI. Furthermore, the 7-min gap translated into a 29–39% increase in the odds that when differences of greater than 30 min occurred within a matched pair, it was the black patient with the longer procedure.

Medicare claims do not include BMI, so if black–white differences in BMI explained differences in procedure times, then Medicare data could not be used to study disparities in procedure times. Table 1 shows that (1) the black–white difference in mean BMI was not large before any matching, (2) matching for Medicare comorbidities, surgical procedures, and hospital removed about half of the small initial difference in BMI without using BMI. Our later results show that completely adjusting for BMI as measured by chart abstraction did not remove the black–white disparity in procedure times. We conclude that Medicare claims can be used to study racial disparities despite the absence of BMI in Medicare claims.

Table 3. Likelihood of Longer Surgical Time in Black–White Matched Pairs, for All Procedures

	Pairs where Black Patient had Longer Time	Pairs where White Patient had Longer Time	Odds of Black Patient Taking Longer than White Patient	95% CI	P Value
Match 1: no BMI: absolute difference in anesthesia time, min*					
0 ≤ time < 10	46	47	0.9787	(0.652–1.470)	0.9174
10 ≤ time < 30	122	103	1.1845	(0.911–1.540)	0.2058
30 ≤ time	251	194	1.2938	(1.073–1.560)	0.0070
Overall	419	344	1.2180	(1.056–1.405)	0.0067
Match 2: w/BMI: absolute difference in anesthesia time, min†					
0 ≤ time < 10	51	56	0.9107	(0.623–1.331)	0.6290
10 ≤ time < 30	118	105	1.1238	(0.864–1.462)	0.3843
30 ≤ time	249	179	1.3911	(1.148–1.686)	0.0008
Overall	418	340	1.2294	(1.065–1.419)	0.0047

* Eleven pairs with ties were excluded (1.41% of 779 matched pairs); number of pairs in final analysis = 763. † Sixteen pairs with ties were excluded (2.05% of 779 matched pairs); number of pairs in final analysis = 758.
BMI = body mass index.

As can be seen from table 1, both populations were remarkably similar in composition. Importantly, our matching produced almost exactly similar predicted procedure times. The predicted times were virtually the same between blacks and whites, yet observed differences after matching were about 7 min.

Studying BMI and chart time information provided a window into what were unobservable variables from our previous study. In the current study, we have shown that BMI did not explain why procedure length of black is greater than that of white. One interesting clue as to why (or when) the disparity occurred did emerge from table 2. We observed that 6 min of the 7-min gap seen in the match that included BMI was found in the cut-to-close surgical time interval. The typical black–white difference in anesthesia induction/emergence time in this example was only 1 min based on the Hodges–Lehmann point estimate. In other words, almost all the time differential between blacks and whites is found during surgical procedure time. The contribution to the racial disparity from the anesthesia team was small and not statistically significant.

Finally, our study sheds light on the importance of the typical 7-min gap between black and white patients. From a clinical perspective, a 7-min gap may be less concerning if distributed evenly across all pairs, but this was not the case. Instead, from table 3, we saw that when there were relatively small differences in procedure time between blacks and whites, there was no significant increase in the odds of blacks or whites experiencing the longer procedure. However, for matched pairs where there was greater than a half hour difference in procedure length between patients, the odds that the black patient had the longer time than the matched white control were significantly increased suggesting that the longer procedure length was associated more often with the black patient than the white patient. Our results suggest that usually there is little difference in procedure time (as seen in fig. 1 and table 3), but when there is a disparity, the problem is a considerable one. The reported typical difference of 7 min for all blacks must imply that for the typical black patient experiencing the disparity, the procedure time difference from whites is much larger than 7 min, as most blacks did not experience a disparity.

A limitation true for any observational study is the possibility that unobserved factors may have accounted for the finding of interest. As such, in our original Surgical Outcomes Study article, the finding of procedure length differences by race[1] may have been influenced by some unobservables. The current report asks if the adjustment method was improved and some unobservables become observable, would the racial differences in procedure length disappear. The current study improves on the previous study in the following four ways: (1) the study examines procedure length in three states not studied before—Texas, Illinois, and New York. It therefore represents an independent sample on which validation of the previous finding could be made, and indeed, we found the disparity to be similar. (2) We now adjust for some variables that were unobservables in the previous study. Most important is the addition of BMI to the adjustment, which was collected through chart review—something not possible in the previous study. We also match on time score; (3) a third important improvement in the current study is the use of multivariate matching rather than regression, thereby not requiring assumptions regarding the form of a model used for adjustment. We observed excellent matches, as reported in table 1. Using this method, we could also match on a propensity score, a procedure time score, BMI, and other variables simultaneously, in part because we had a very large pool of whites in which to find matches to the black population. Finally, (4) through chart review, we could examine whether the procedure length differences were occurring between cut and close or between induction to cut and close to recovery room. This analysis was impossible in the previous study because procedure length was based only on the anesthesia claim, not chart review.

Still another factor that may influence our results concern variables associated with race that may also influence the disparity in procedure time, such as income and education. In previous work,[1] we did observe an income effect, in that we observed less of a disparity in higher income blacks. In this study, we did not match on income, as income is highly correlated with race[33] and it was not our intent to disentangle that connection. If the mechanism for procedure time disparity was income, education, or race, the disparity would be equally interesting in this Medicare population, as all had insurance and went to the same hospital.

The implications of identifying a clinically relevant disparity in procedure length are complex. If the disparity is real, after adjusting for patient characteristics that may prolong procedure length, then we are forced to ask why this difference is occurring. It has been shown that when procedures are performed by resident surgeons, they are longer than when performed by attending surgeons.[34,35] If blacks had a higher risk of receiving care from residents or inexperienced surgeons, this may possibly account for some of the time differential.[36] As Medicare data do not provide a separate bill for the resident surgeon, we cannot directly know whether the prolonged case was due to teaching. Furthermore, we do not know whether specific key portions of the surgical procedure were performed by the attending or the resident surgeons. Who is holding the scalpel and who is holding the retractor is impossible to know short of videotaping all procedures. However, if the etiology is differential surgical experience, making departments aware that the process of surgical selection is leading to this pattern may aid in producing a more equitable system. Determining the cause or causes of this disparity is beyond the scope of this report.

In conclusion, the racial differences in procedure time in the Medicare population were significant even after adjusting for BMI. Furthermore, the observed time differences appear to occur almost entirely during the surgical cut-to-close period, not the induction-to-emergence period. When differences within matched pairs exceeded 30 min, the black patient was

significantly more likely to be the one with the longer procedure time. This remaining significant difference in procedure time between black and white patients requires our attention.

References

1. Silber JH, Rosenbaum PR, Zhang X, Even-Shoshan O: Influence of patient and hospital characteristics on anesthesia time in medicare patients undergoing general and orthopedic surgery. ANESTHESIOLOGY 2007; 106:356–64

2. Silber JH, Rosenbaum PR, Zhang X, Even-Shoshan O: Estimating anesthesia and surgical procedure times from medicare anesthesia claims. ANESTHESIOLOGY 2007; 106:346–55

3. Silber JH, Rosenbaum PR, Even-Shoshan O, Mi L, Kyle FA, Teng Y, Bratzler DW, Fleisher LA: Estimating anesthesia time using the medicare claim: A validation study. ANESTHESIOLOGY 2011; 115:322–33

4. Silber JH, Rosenbaum PR, Kelz RR, Reinke CE, Neuman MD, Ross RN, Even-Shoshan O, David G, Saynisch PA, Kyle FA, Bratzler DW, Fleisher LA: Medical and financial risks associated with surgery in the elderly obese. Ann Surg 2012; 256:79–86

5. Zubizarreta JR, Neuman MD, Silber JH, Rosenbaum PR: Contrasting evidence within and between institutions that supply treatment in an observational study of alternative forms of anesthesia. J Am Stat Assoc 2012; 107:901–15

6. Zubizarreta JR, Reinke CE, Kelz RR, Silber JH, Rosenbaum PR: Matching for several sparse nominal variables in a case-control study of readmission following surgery. Am Stat 2011; 65:229–38

7. Zubizarreta JR: Using mixed integer programming for matching in an observational study of kidney failure after surgery. J Am Stat Assoc 2012; 107:1360–71

8. Rosenbaum PR: Part II: Matching, Design of Observational Studies. New York, Springer, 2010, pp 153–253

9. Rosenbaum PR, Ross RN, Silber JH: Minimum distance matched sampling with fine balance in an observational study of treatment for ovarian cancer. J Am Stat Assoc 2007; 102:75–83

10. Yang D, Small DS, Silber JH, Rosenbaum PR: Optimal matching with minimal deviation from fine balance in a study of obesity and surgical outcomes. Biometrics 2012; 68:628–36

11. Hansen BB: The prognostic analogue of the propensity score. Biometrika 2008; 95:481–8

12. Rosenbaum PR: Design of Observational Studies. New York, Springer, 2010, pp 43–44; 172–175; 188; 197–206

13. Rosenbaum P, Rubin D: The central role of the propensity score in observational studies for causal effects. Biometrika 1983; 70:41–55

14. Joffe MM, Rosenbaum PR: Invited commentary: Propensity scores. Am J Epidemiol 1999; 150:327–33

15. Huber PJ: Robust Statistics. Hoboken, NJ, John Wiley & Sons, 1981

16. Hampel FR, Ronchett EM, Rousseeuw PJ, Stahel WA: Chapter 6. Linear Models: Robust Estimation, Robust Statistics. The Approach Based on Influence Functions. New York, NY, John Wiley & Sons, 1986, pp 315–28

17. Rubin DB: For objective causal inference, design trumps analysis. Ann Appl Stat 2008; 2:808–40

18. Rubin DB: The design *versus* the analysis of observational studies for causal effects: Parallels with the design of randomized trials. Stat Med 2007; 26:20–36

19. Rosenbaum PR, Rubin DB: Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. Am Stat 1985; 39:33–8

20. Silber JH, Rosenbaum PR, Trudeau ME, Even-Shoshan O, Chen W, Zhang X, Mosher RE: Multivariate matching and bias reduction in the surgical outcomes study. Med Care 2001; 39:1048–64

21. Cochran WG, Rubin DB: Controlling bias in observational studies. A review. Sankhya 1973; Series A, 35:417–46

22. Hollander M, Wolfe DA: Nonparametric Statistical Methods, 2nd edition. New York, NY, John Wiley & Sons, 1999, pp 36–49; 51–5; 72–4; 190–9

23. Bishop YMM, Fienberg SE, Holland PW: Chapter 8: Analysis of Square Tables: Symmetry and Marginal Homogeneity, Discrete Multivariate Analysis: Theory and Practice. Cambridge, The MIT Press, 1975, pp 281–6

24. Wilk MB, Gnanadesikan R: Probability plotting methods for the analysis of data. Biometrika 1968; 55:1–17

25. Lee CT, Dunn RL, Chen BT, Joshi DP, Sheffield J, Montie JE: Impact of body mass index on radical cystectomy. J Urol 2004; 172(4 Pt 1):1281–5

26. Chang SS, Jacobs B, Wells N, Smith JA Jr, Cookson MS: Increased body mass index predicts increased blood loss during radical cystectomy. J Urol 2004; 171:1077–9

27. Dindo D, Muller MK, Weber M, Clavien PA: Obesity in general elective surgery. Lancet 2003; 361:2032–5

28. Bostanjian D, Anthone GJ, Hamoui N, Crookes PF: Rhabdomyolysis of gluteal muscles leading to renal failure: A potentially fatal complication of surgery in the morbidly obese. Obes Surg 2003; 13:302–5

29. Torres-Villalobos G, Kimura E, Mosqueda JL, García-García E, Domínguez-Cherit G, Herrera MF: Pressure-induced rhabdomyolysis after bariatric surgery. Obes Surg 2003; 13:297–301

30. Lebuffe G, Andrieu G, Wierre F, Gorski K, Sanders V, Chalons N, Vallet B: Anesthesia in the obese. J Visc Surg 2010; 147(5 suppl):e11–9

31. Myatt J, Haire K: Airway management in obese patients. Curr Anaesth Crit Care 2010; 21:9–15

32. Sinha AC: Some anesthetic aspects of morbid obesity. Curr Opin Anaesthesiol 2009; 22:442–6

33. Braveman PA, Cubbin C, Egerter S, Chideya S, Marchi KS, Metzler M, Posner S: Socioeconomic status in health research: One size does not fit all. JAMA 2005; 294:2879–88

34. Khuri SF, Najjar SF, Daley J, Krasnicka B, Hossain M, Henderson WG, Aust JB, Bass B, Bishop MJ, Demakis J, DePalma R, Fabri PJ, Fink A, Gibbs J, Grover F, Hammermeister K, McDonald G, Neumayer L, Roswell RH, Spencer J, Turnage RH; VA National Surgical Quality Improvement Program: Comparison of surgical outcomes between teaching and nonteaching hospitals in the Department of Veterans Affairs. Ann Surg 2001; 234:370–82

35. Campbell DA Jr, Henderson WG, Englesbe MJ, Hall BL, O'Reilly M, Bratzler D, Dellinger EP, Neumayer L, Bass BL, Hutter MM, Schwartz J, Ko C, Itani K, Steinberg SM, Siperstein A, Sawyer RG, Turner DJ, Khuri SF: Surgical site infection prevention: The importance of operative duration and blood transfusion—results of the first American College of Surgeons–National Surgical Quality Improvement Program Best Practices Initiative. J Am Coll Surg 2008; 207:810–20

36. Egbert LD, Rothman IL: Relation between the race and economic status of patients and who performs their surgery. N Engl J Med 1977; 297:90–1