Terms of Clearance: **Learner's Perception Survey, 2900-0691**

Kashner_Henley_Golden_Effects ACGME

Studying the Effects of ACGME Duty Hours Limits on Resident Satisfaction: Results From VA Learners' Perceptions Survey

This document is in reference to the non-response bias analysis requested by OMB. OMB made a second request for the lead Statistician to expound on their previous response, by providing more detail.

Studying the Effects of ACGME Duty Hours Limits on Resident Satisfaction: Results From VA Learners' Perceptions Survey

T. Michael Kashner, PhD, JD, Steven S. Henley, MS, Richard M. Golden, PhD, John M. Byrne, DO, Sheri A. Keitz, MD, PhD, Grant W. Cannon, MD, Barbara K. Chang, MD, MA, Gloria J. Holland, PhD, David C. Aron, MD, Elaine A. Muchmore, MD, Annie Wicker, and Halbert White, PhD

Abstract

Background
As the Accreditation Council on Graduate Medical Education (ACGME) deliberates over further limiting duty hours of graduate medical education (GME) trainees, few large-scale studies have shown residents to be satisfied with the effect the 2003 standards have had on clinical care, education outcomes, or working environments. This study measures the effect of the 2003 duty hours limits on resident-reported satisfaction with GME training during their rotations through the Department of Veterans Affairs (VA) medical centers from 2001 through 2007.

Method
Self-reported satisfaction with clinical care and education environments were assessed by comparing responses to VA's annual Learners' Perceptions Survey administered before 2003 with responses administered after 2003. To measure duty hours effects on satisfaction, before–after differences were adjusted for covariate biases modeled after an exhaustive covariate search with 10-fold cross-validation. Because nonteaching controls are not available in satisfaction studies, we used a robust differencing variable technique to control before–after differences for trend biases in the simultaneous presence of missing data and possible model misspecification.

Results
There were 19,605 responders. Adjusting for covariate and trend biases, after the 2003 ACGME standards, 25% more residents in medicine specialties reported satisfaction with VA clinical environment and 11% more with VA preceptors and faculty. For surgery, 33% more residents reported satisfaction with VA clinical environment and 12% more with VA preceptors and faculty. Satisfaction with working environment was mixed.

Conclusions
The 2003 ACGME duty hours standards were associated with improved satisfaction for resident clinical training and learning environments.

Acad Med. 2010; 85:1130–1139.

Standards governing duty hours limits have generally been considered necessary in graduate medical education (GME) to protect the safety of both patients[1] and residents.[2] Resident sleep deprivation as a result of long duty hours has been linked to higher rates of medical errors,[3] poorer clinical performance,[4] adverse events,[5] and attentional failures[6] in observational, pre–post, and experimental studies. Longer duty hours have also been linked to resident motor-vehicle-related injuries,[7] obstetric complications,[8] depression,[9] burnout,[10] poorer quality of life[11] and neuropsychological performance,[12] including memory loss and reduced response times.[13]

In response to these safety concerns, the Accreditation Council for Graduate Medical Education (ACGME) implemented mandatory standards on July 1, 2003, that limited duty hours for medical residents in accredited U.S. GME programs.[14] Although benefits from ACGME duty hours limits continue to be debated,[15] few studies have described how duty hours limits may be affecting clinical training environments, trainee learning, resident access to preceptors and faculty, and resident education.[16] For instance, residents have complained that mandatory duty hours rules interfere with continuity of care,[17] increase cross-coverage errors,[18] shift the education focus away from professionalism,[19] create fear that new regulations will add additional training years,[20] and cause frustration when residents are faced with heavy workloads and must reconcile actual hours against ACGME duty hours rules.[21] Underscoring these concerns are the contrasting missions of the teaching hospitals, who need staff to provide professional care; faculty, who balance attending, practice, service, and research responsibilities; and residents, who need access to faculty and supervised clinical experiences to properly prepare them to enter independent practice.

To assess the effects of ACGME duty hours rules on training environments, researchers have surveyed residents using post and pre–post survey designs. Post surveys were administered after ACGME duty hours rules became mandatory. These surveys asked residents and fellows[10,22–25] about their views of the success or failure of the mandatory standards. Although informative about

Please see the end of this article for information about the authors.

Correspondence should be addressed to Dr. Kashner, Department of Psychiatry, University of Texas Southwestern Medical Center at Dallas, 5323 Harry Hines Blvd., Dallas, TX 75390-9086; telephone: (214) 648-4608; fax: (214) 648-4612; e-mail: michael.kashner@va.gov.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text; simply type the URL address into any Web browser to access this content. Clickable links to this material are provided in the HTML text and PDF of this article on the journal's Web site (www.academicmedicine.org).

how residents perceived duty hours limits, postsurvey results are often colored by memory loss, cohort confounds when all of the responders who have prelimits experiences have become upper-level residents by the time the survey is administered, and reporting biases when residents mimic faculty attitudes and beliefs in their survey responses.

Pre–post designs[26–28] compare responses to surveys that were administered in 2003 and earlier with responses to the same surveys readministered in 2004 and later. Pre–post designs are subject to covariate biases whenever responders who took the survey prelimits (2003 and earlier) differ significantly from responders who took the survey postlimits (2004 and later). Pre–post designs are also subject to trend biases whenever naturally occurring time trends in the data confound pre–post differences. Covariate biases have been addressed by computing outcomes that are adjusted to reflect the influences of variations in responder characteristics. Trend biases are addressed by difference-in-differences methods[29] where effect sizes are computed by subtracting the pre–post difference in mean responses among physician residents rotating through "effect" settings minus the pre–post difference in mean responses computed for comparable residents who rotated through "control" settings. Control settings have been identified as (1) nonteaching hospitals where duty hours limits are irrelevant,[30–32] (2) training programs in teaching hospitals where duty hours limits were openly not enforced,[26] or (3) responders whose duty schedules were not changed, for whatever reason, by duty hours limits. Facility-level controls are limited to outcomes that can be observed in both teaching and nonteaching settings, such as patient outcomes and medical errors, and are thus not practical for resident satisfaction surveys. Program-level controls are often difficult to implement because few program directors openly defy ACGME standards. Responder-level controls can be identified by asking respondents if the 2003 duty hours limits had any impact on their actual duty schedules. However, such questions were not answerable before 2003, when ACGME duty hours limits were first implemented. We call this the "missing-data problem."

For this report, we introduce and apply a methodology that uses responder-level controls to assess the influence of the 2003 mandatory ACGME duty hours limits on how physician residents perceived their clinical training environments in the Department of Veterans Affairs (VA) medical centers between July 1, 2000 and June 30, 2007. The study addresses covariate confounds, trend biases, and missing-data problems in three important aspects. First, we used the Learners' Perceptions Survey (LPS), a structured interview administered annually by the VA Office of Academic Affiliations (VA-OAA) to residents rotating through VA medical centers. Second, respondents were classified into effects or control groups based on LPS survey questions that asked respondents whether duty hours limits actually changed their hours worked during scheduled VA rotations. Third, we adjusted for covariate and trend biases using a robust differencing variable technique, an advanced statistical method designed to handle the missing-data problem caused by failing to identify controls among pre-2003 responders.

Method

Data collection

We obtained resident satisfaction data from the VA LPS, which has been described elsewhere.[33,34] Elements of each satisfaction domain and ACGME duty hours limits questions are listed in List 1. The analyses for this study were conducted for administrative purposes by, and were under the direct supervision of, the VA-OAA, under review by OMB Information Collection (#2900-0691) approved for VA Form #10-0439, for all data collected through January 2010.

Used as a performance metric since 2001, VA-OAA has administered the LPS annually to all trainees who rotate through VA medical centers. To reflect overall satisfaction, respondents rate "clinical training … on a scale from 0 to 100, where 100 is a perfect score and 70 is a passing score." We dichotomized overall satisfaction responses from all available surveys (2001–2007) into satisfied ($\geq$70) or otherwise ($<$70), where 70 was defined by VA as a "passing" score.

Respondents also rated each of five domains on a five-point scale (List 1).

List 1

Elements Comprising Satisfaction in the Veterans Affairs Annual Learners' Perceptions Survey

1. *Faculty/Preceptors Domain*
   - Clinical skills
   - Teaching ability
   - Interest in teaching
   - Research mentoring
   - Accessibility
   - Approachability
   - Feedback timeliness
   - Evaluation fairness
   - Role models
   - Mentoring
   - Patient-oriented
   - Faculty quality
   - Evidence-based practice
2. *Learning Environment Domain*
   - Time with patients
   - Supervision
   - Autonomy
   - Noneducational "scut" work
   - Interdisciplinary approach
   - Preparation for clinical practice
   - Future training
   - Business aspects
   - Learning time
   - Access to specialty expertise
   - Teaching conferences
   - Care quality
   - Patient safety
   - Spectrum of patient problems
   - Patient diversity
3. *Clinical Environment Domain*
   - Work hours
   - Number inpatients admitted for your care
   - Outpatients seen
   - Timely availability of outpatient appointments
   - Timely performance of necessary procedures/surgeries
   - Timely admission of patient
   - Ability to use emerging therapies/pharmaceuticals
   - How well physicians and nurses work together
   - How well physicians and ancillary staff work together
   - Tests done in timely manner on weekends
   - Tests done in timely manner at nights
   - Accessing patient records
   - Backup system for electronic records
   - Amount of paperwork
   - Ability to work within system to get best care for patients

(*Continues*)

List 1
(*Continued*)

4. *Working Environment Domain*
   - Morale of faculty
   - Support staff
   - Peer group
   - Services of laboratory
   - Radiology
   - Ancillary/support staff
   - Library
   - Call schedule
   - Computerized patient record system
   - Computer access
   - Internet access
   - Orientation program
   - Workspace
5. *Physical Environment Domain*
   - Convenience of facility location
   - Parking
   - Personal safety
   - Availability of phones
   - Needed equipment
   - Food services
   - Equipment maintenance
   - Facility maintenance
   - Lighting
   - Heating and air conditioning
   - Cleanliness/housekeeping
   - Call rooms
6. *Accreditation Council for Graduate Medical Education Duty Hours Limits Question*
   - Colleague support
   - Personal reward
   - Relationship with patients
   - Appreciation of work by faculty
   - Appreciation of work by patient
   - Balance of professional and personal life
   - Work enjoyment
   - Job stress
   - Work fatigue
   - Continuity of care
   - Patient-care responsibility
   - Quality of care
   - Enhancement of clinical skills

For these analyses, to compute the odds that a resident reported being satisfied, we dichotomized responses into "satisfied" (very satisfied, somewhat satisfied) and "otherwise" (neither satisfied nor dissatisfied, somewhat dissatisfied, very dissatisfied). Since 2001, domains included satisfaction with clinical faculty/preceptors, learning, working, and physical environments. A fifth domain, clinical environment, was added with the 2003 survey.

The VA added an ACGME duty hours limits question beginning with the 2004 survey (List 1). The question read, "In July 2003, the Accreditation Council for Graduate Medical Education instituted changes in requirements in duty hours/scheduling for resident education. In your opinion, what effect have these changes had on your educational experience at the VA facility…?" Respondents rated their answers on a five-point scale. To adjust pre–post differences for time trends, we constructed a differencing variable by dichotomizing responses to this question to classify each responder as either a no-effect control (response of "no effect") or an effect-responder (response of "very positive," "somewhat positive," "somewhat negative," or "very negative" effect).

Several scenarios could explain why residents may have claimed that ACGME duty hours limits had no effect on their VA clinical training settings. For example, a resident may have worked a schedule of hours that was within the duty hours limits whether or not the training program was enforcing the ACGME-mandated duty hours rules. Alternatively, the training program may have ignored duty hours rules, at least during the respondent's rotation.

To adjust for trend biases, we constructed the differencing variable to combine responders who reported either positive or negative perceptions of duty hours effects. We combined the two groups because our purpose was to measure how duty hours limits influenced residents' ratings of their training environment, and not to determine how residents actually perceived duty standards. Although important, the latter research lies outside the scope of the current study.

The LPS also obtained demographic information about each respondent, including gender, training level (postgraduate year), and specialty. For our analyses, we grouped resident specialties into medicine (internal medicine, neurology, physical and rehabilitation medicine), surgery (surgery, anesthesiology), psychiatry (including psychiatric subspecialties), and ancillary care (diagnostic specialties, radiology, pathology). Beginning in 2003, the LPS began collecting information on medical school graduation (date graduated, U.S. versus foreign medical school) and the mix of patients seen during the respondent's VA rotation. We estimated patient mix from survey responses as the percentage of patients the respondent reported seeing during "an average week, at the VA…" for each of seven patient categories: 65 years of age or older, chronic mental illness, chronic medical illness, multiple medical illnesses, alcohol/substance dependence conditions, low-income socioeconomic status, and no social/family support.

Analyses

To be comparable with other studies, we computed the effect of ACGME duty hours limits on resident satisfaction for each satisfaction domain as a ratio of odds ratios (ROR) describing whether the resident reported satisfaction or otherwise. The ROR numerator is calculated for effect-responders and equals the odds that these respondents would have reported satisfaction in the postperiod divided by the odds that these same respondents would have reported satisfaction in the preperiod. The denominator is calculated in the same way, but only for no-effect controls. ROR $= 1$ indicates that pre–post changes in satisfaction rates among effect-responders were no different from the pre–post changes among no-effect controls, and suggests that no duty-limit effect on satisfaction was observed. On the other hand, ROR > 1 indicates that pre–post changes in satisfaction rates among effect-responders were greater than their no-effect counterparts, and suggests that duty hours limits were associated with higher satisfaction rates. Similarly, ROR $<$ 1 suggests that duty hours limits led to decreased satisfaction rates.

ROR is adjusted for both covariate and trend biases using a robust differencing variable technique that extends difference-in-differences analyses[29] to logistic regression[35] by (1) using resident-level training environments as the unit of analyses, (2) identifying control respondents to adjust for trend biases, (3) accounting for missing data without imputation noise, (4) performing an exhaustive model search to adjust for covariate biases, and (5) computing estimates of effect sizes and confidence intervals (CIs) that are robust to model misspecification (see Mathematical Appendix,

Table 1

Description of Veterans Affairs Learners' Perceptions Survey Respondents by Reporting Period, 2001–2007*

	No. (%) reporting from all periods	No. (%) reporting from pre–duty hours limits period	No. (%) reporting from post–duty hours limits period
Gender	18,323	6,781	11,542
Female	7,102 (39)	2,639 (39)	4,463 (39)
Medical school†	14,177	2,571	11,606
U.S.	10,575 (75)	1,947 (76)	8,628 (74)
Entered residency	14,006	2,616	11,390
Gap†‡	2,174 (16)	411 (16)	1,763 (16)
Training level	19,605	6,964	12,641
PGY-1	5,498 (28)	1,910 (27)	3,588 (28)
PGY-2	4,446 (23)	1,554 (22)	2,892 (23)
PGY-3	4,289 (22)	1,523 (22)	2,766 (22)
PGY-4	2,964 (15)	1,080 (16)	1,884 (15)
PGY-5	1,444 (7)	541 (8)	903 (7)
PGY-6	706 (4)	275 (4)	431 (3)
PGY-7	258 (1)	81 (1)	177 (1)
Medical specialty	17,833	6,692	11,141
Medicine§	11,990 (67)	4,731 (70)	7,259 (65)
Surgery	3,804 (21)	1,433 (21)	2,371 (21)
Psychiatry§	1,615 (9)	402 (6)	1,213 (11)
Ancillary§	424 (2)	126 (2)	298 (3)
Patients seen each week†			
Over 65 years of age§	14,183	2,521	11,662
<10%	406 (3)	31 (1)	375 (3)
10%–24%	700 (5)	83 (3)	617 (5)
25%–49%	1,518 (11)	175 (7)	1,343 (11)
50%–74%	4,833 (34)	792 (31)	4,041 (34)
75%–89%	5,310 (37)	1,125 (45)	4,185 (37)
90%–100%	1,416 (10)	315 (13)	1,101 (10)
Mental illness§	14,126	2,464	11,662
<10%	2,304 (16)	512 (21)	1,792 (15)
10%–24%	3,703 (26)	683 (28)	3,020 (26)
25%–49%	3,359 (24)	540 (22)	2,819 (24)
50%–74%	2,553 (18)	397 (16)	2,156 (19)
75%–89%	1,418 (10)	209 (8)	1,209 (10)
90%–100%	789 (6)	123 (5)	666 (6)
Medical illness§	14,165	2,503	11,662
<10%	292 (2)	28 (1)	264 (2)
10%–24%	301 (2)	48 (2)	253 (2)
25%–49%	771 (5)	104 (4)	667 (6)
50%–74%	2,299 (16)	325 (13)	1,974 (17)
75%–89%	5,017 (35)	888 (36)	4,129 (35)
90%–100%	5,485 (39)	1,110 (44)	4,375 (38)
Multiple illnesses§	14,154	2,493	11,661
<10%	255 (2)	19 (1)	236 (2)
10%–24%	274 (2)	57 (2)	217 (2)
25%–49%	858 (6)	121 (5)	737 (6)

(Continues)

Supplemental Digital Content 1, http://links.lww.com/ACADMED/A19).

To account for covariate biases between pre- and postperiods, and effect- and control responders, we adjusted satisfaction outcomes based on responder characteristics and clinic experience. To account for nonlinear associations, we used a maximum-likelihood recoding strategy to transform all continuous and ordinal variables into binary covariates. Specifically, each continuous and ordinal variable was independently dichotomized using nonparametric, bootstrapped, maximum-likelihood cut-point estimates for each of the five domains and overall satisfaction score.[36] For each dependent variable, we determined a model containing the most predictive covariates from an exhaustive model search[37] using the generalized Akaike information criteria[38] based on data from postlimits periods. We then validated these empirically motivated models using a 10-fold cross-validation approach.[39]

Next, we constructed a theoretically motivated model to contain three specific variables. First, a period indicator variable assumed a value of zero if the respondent answered the LPS survey in prelimits years (2001–2003), or a value of one if the respondent had answered the survey during postlimits years (2004–2007). Second, a differencing variable was constructed to assume a value of zero if the respondent was a no-effect control, and a value of one if the respondent reported either a positive or negative effect to the ACGME duty hours limits question (effect-responder). Third, a period × differencing variable interaction term was computed by multiplying the period indicator and differencing variables for each respondent.

We constructed a final model by combining the terms that made up the theoretically and empirically motivated models. All models included a constant term. We then used a nonnested model selection test[40–43] to compare the fit of all three models. If the final model fit the data better than either theoretically or empirically motivated models, then duty hours limits effects were estimated by exponentiating the estimated coefficient to the period × differencing variable interaction term to the final model. To semantically interpret the interaction

Table 1
(*Continued*)

	No. (%) reporting from all periods	No. (%) reporting from pre–duty hours limits period	No. (%) reporting from post–duty hours limits period
50%–74%	2,505 (18)	357 (14)	2,148 (18)
75%–89%	5,338 (38)	967 (39)	4,371 (38)
90%–100%	4,924 (35)	972 (39)	3,952 (34)
Alcohol/substance abuse§	14,145	2,483	11,662
<10%	889 (6)	152 (6)	737 (6)
10%–24%	3,029 (21)	566 (23)	2,463 (21)
25%–49%	4,310 (31)	680 (27)	3,630 (31)
50%–74%	3,776 (27)	712 (29)	3,064 (26)
75%–89%	1,738 (12)	302 (12)	1,436 (12)
90%–100%	403 (3)	71 (3)	332 (3)
Low income	14,123	2,461	11,662
<10%	445 (3)	75 (3)	370 (3)
10%–24%	1,905 (14)	318 (13)	1,587 (14)
25%–49%	4,356 (31)	730 (30)	3,626 (31)
50%–74%	4,047 (29)	724 (29)	3,323 (29)
75%–89%	2,649 (19)	485 (20)	2,164 (19)
90%–100%	721 (5)	129 (5)	592 (5)
No social support	14,112	2,450	11,662
<10%	969 (7)	173 (7)	796 (7)
10%–24%	3,404 (24)	572 (23)	2,832 (24)
25%–49%	4,622 (33)	781 (32)	3,841 (33)
50%–74%	3,350 (24)	596 (24)	2,754 (24)
75%–89%	1,475 (11)	269 (11)	1,206 (10)
90%–100%	292 (2)	59 (2)	233 (2)
Satisfaction outcome			
Summary score	18,391	6,764	11,627
70 or above	16,349 (90)	5,998 (89)	10,351 (90)
Clinical preceptor§	18,917	6,587	12,330
Very satisfied	8,593 (45)	2,664 (40)	5,929 (48)
Somewhat satisfied	7,909 (42)	2,901 (44)	5,008 (41)
Neither	1,175 (6)	498 (8)	677 (6)
Somewhat dissatisfied	861 (5)	371 (6)	490 (4)
Very dissatisfied	379 (2)	153 (2)	226 (2)
Learning§	18,923	6,709	12,214
Very satisfied	5,449 (29)	1,475 (22)	3,974 (33)
Somewhat satisfied	9,457 (50)	3,599 (54)	5,858 (48)
Neither	2,123 (11)	912 (14)	1,211 (10)
Somewhat dissatisfied	1,416 (8)	545 (8)	871 (7)
Very dissatisfied	478 (3)	178 (3)	300 (3)
Clinical†§	14,391	2,508	11,883
Very satisfied	3,221 (22)	336 (13)	2,885 (24)
Somewhat satisfied	6,400 (45)	1,149 (46)	5,251 (44)
Neither	2,452 (17)	555 (22)	1,897 (16)
Somewhat dissatisfied	1,726 (12)	337 (13)	1,389 (12)
Very dissatisfied	592 (4)	131 (5)	461 (4)

(*Continues*)

term, we assumed that the adjusted impact of the differencing variable on satisfaction is invariant with time (see Mathematical Appendix, Supplemental Digital Content 1, http://links.lww.com/ACADMED/A19).

The LPS did not ask respondents about duty hours limits in prelimits periods when ACGME rules were not enforced (missing-data problem). However, the concept of a no-effect control in prelimits periods is still relevant. Although one can only speculate about the actions of VA staff during 2001–2003, it is possible that some residents were assigned to work schedules that complied naturally with the duty hours rules. Residents may also have been supervised by attending physicians who *would* have done business as usual and ignored duty hours limits had such rules been mandatory. Assigning values to the differencing variable for prelimits responders is treated as a missing-at-random problem. That is, by knowing the year of the survey, one knows whether the value of the differencing variable is missing.[44] Concerning missing data, the period × differencing variable interaction term will always equal "zero" during prelimits periods. Thus, only the differencing variable as a main effects term will have missing data. Rather than using imputation, we computed maximum-likelihood estimates by taking into account all possible patterns of values for the missing data. Missing values among covariates caused by later additions to the LPS survey were also treated in this way. Thus, model coefficients were computed directly from the observable component of the data without imputation noise.

Final models were tested for fit, the presence of model misspecification, and multicollinearity. Because of the potential for misspecification, robust estimation methods that are valid in the presence of model misspecification were used to compute both parameters and CIs.[29,45,46]

Results

Table 1 presents characteristics and satisfaction scores for 19,605 LPS physician resident responders classified by reporting period. Variation in responder characteristics underscores the need to adjust for covariate biases. Compared with all residents in ACGME-accredited programs in 2008–2009,[47] the

Table 1
(*Continued*)

	No. (%) reporting from all periods	No. (%) reporting from pre–duty hours limits period	No. (%) reporting from post–duty hours limits period
Physical§	18,319	6,635	11,684
Very satisfied	4,516 (25)	1,289 (19)	3,227 (28)
Somewhat satisfied	8,403 (46)	3,065 (46)	5,338 (46)
Neither	3,112 (17)	1,308 (20)	1,804 (15)
Somewhat dissatisfied	1,861 (10)	795 (12)	1,066 (9)
Very dissatisfied	427 (2)	178 (3)	249 (2)
Working§	18,748	6,669	12,079
Very satisfied	4,672 (25)	1,182 (18)	3,490 (29)
Somewhat satisfied	8,749 (47)	3,211 (48)	5,538 (46)
Neither	2,826 (15)	1,259 (19)	1,567 (13)
Somewhat dissatisfied	1,890 (10)	773 (12)	1,117 (9)
Very dissatisfied	611 (3)	244 (4)	367 (3)
Duty hours limits¶	N/A	N/A	10,653
No effect	N/A	N/A	2,585 (24)
Effect	N/A	N/A	8,068 (76)
Very negative	N/A	N/A	96 (1)
Somewhat negative	N/A	N/A	465 (4)
Somewhat positive	N/A	N/A	3,916 (37)
Very positive	N/A	N/A	3,591 (34)

* Total sample (n = 19,605) includes those reporting during pre– duty hours limits periods representing academic years 2001 (n = 1,752), 2002 (n = 2,531), and 2003 (n = 2,681) and those reporting during post– duty hours limits periods representing academic years 2004 (n = 2,793), 2005 (n = 3,101), 2006 (n = 3,792), and 2007 (n = 2,955).
† First introduced beginning with the FY2003 survey.
‡ Time between graduating from medical school and beginning residency program is greater than four years.
§ Indicates statistically significant at $P < .05$ based on two-sided Pearson chi-square test.
¶ First introduced beginning with the FY2004 survey.

LPS sample had slightly fewer females at 7,102 of 18,323 (39%) versus 48,823 of 108,176 (45%) residents in ACGME-accredited programs, had fewer international medical school graduates at 3,602 of 14,177 (25%) versus 29,488 of 108,176 (27%) ACGME residents, and fewer first-year residents at 5,498 of 19,605 (28%) versus 38,404 of 108,176 (36%) ACGME residents.

Table 2 reports estimates of duty hours limits effects measured as an ROR based on the robust differencing variable technique. The wide CIs reflect the uncertainty associated with working with incomplete datasets.

Overall, respondents tended to report higher satisfaction with their VA clinical training environment when duty hours limits applied. For instance, respondents overall were 2.46 times (95% CI [1.49, 4.05], $P < .001$) more likely to report satisfaction with VA as a clinical training environment under duty hours limits than without such standards. These findings held across each of the five domains, for all residents taken together, and for medicine residents only. Surgery residents tended to report higher levels of satisfaction only for clinical faculty or preceptors and clinical environment. Estimates for ancillary and psychiatry specialties were inconclusive.

To understand its relevance to education, we recalculated ROR estimates of duty hours limits effect sizes (Table 2) to reflect the adjusted estimate of the percentage of respondents who would change their response from "not satisfied" to "satisfied" as duty hours limits became mandatory (Table 3). The largest change occurred in the clinical environment domain. For surgery residents (ROR = 9.10, 95% CI [2.62, 31.61], $P = .0005$), satisfaction rates for clinical environments increased from a prelimits period rate of 60% (Table 1) to an expected 93% under mandatory limits, adjusted to reflect differences in the mix of respondents and other time trends in the data. That is, we estimate that 33 out of 100 respondents, who otherwise would not have been satisfied, would have reported satisfaction under mandatory duty hours limits. For medicine (ROR = 3.46, 95% CI [1.37, 8.70], $P = .0084$), the prelimits period satisfaction rate of 58% increased to 83%, for an adjusted net increase of 25% under the mandatory duty hours rules. Similarly, these data suggest that an expected 12% more surgery residents and 11% more medicine residents would have reported satisfaction with faculty or preceptors under ACGME mandatory duty hours rules than without such rules.

To show the importance of adjusting for covariate mix and time trends, the unadjusted pre–post period change in satisfaction with VA training environments is OR = 1.00 (95% CI [0.91, 1.11], $P = .96$). There was also little adjusted cross-sectional difference in overall satisfaction between effect-responders and no-effect controls (OR = 1.12, $P = .66$). This finding is comparable with those of studies showing few differences in patient outcomes between teaching and nonteaching VA hospitals.[48] Females were generally more likely to report overall satisfaction for VA training (OR = 1.12, $P = .038$) as well as clinical (OR = 1.09, $P = .039$) and working (OR = 1.15, $P < .001$) environments. The higher rates of satisfaction among females are consistent with other surveys.[49] Respondents who reported that 50% or more of the patients they saw were without family support, or were substance abusers, were only 56% (OR = 0.56, $P < .0001$) and 73% (OR = 0.73, $P < .0001$), respectively, as likely to report satisfaction with VA clinical training environments as their counterparts who saw fewer than 50% of such patients.

Discussion

Using advanced statistical techniques to adjust for trend and covariate biases, we found that the 2003 ACGME standards significantly and materially enhanced learning satisfaction rates for medicine and surgery residents rotating through VA medical centers. The statistical tools, along with our large sample size and robust survey, provided a comprehensive estimate of the impact of duty hours

Table 2

Effect of Accreditation Council for Graduate Medical Education Duty Hours Limits on Resident Satisfaction With Clinical Rotations Through Veterans Affairs Medical Centers Between 2001 and 2007

	Effect size*	95% CI	P†	GAIC/2n‡	Condition number§	No.
Overall clinical training						
All specialties¶	2.46	1.49–4.05	.0004	0.341	1214	16,774
Medicine¶	2.98	1.80–4.95	.0000	0.341	915	11,315
Surgery	1.26	0.14–11.46	.8349	0.348	6214	3,552
Psychiatry	**					1,508
Ancillary	1.42	0.03–73.54	.8612	0.341	2391	399
Clinical faculty/preceptors						
All specialties¶	2.94	1.84–4.72	.0000	0.472	1060	16,394
Medicine¶	3.48	2.09–5.79	.0000	0.366	979	11,047
Surgery¶	4.76	1.68–13.49	.0034	0.401	971	3,468
Psychiatry††	0.83	0.48–1.42	.4918	0.346	81048	1,489
Ancillary	1.63	0.06–46.56	.7762	0.435	1421	390
Learning environment						
All specialties¶	2.23	1.47–3.38	.0001	0.498	1650	17,236
Medicine¶	2.49	1.59–3.90	.0001	0.507	1239	11,616
Surgery¶	2.32	0.79–6.75	.1237	0.495	1787	3,649
Psychiatry	2.21	0.42–11.55	.3471	0.441	2359	1,563
Ancillary††	0.62	0.17–2.35	.4899	0.493	40805	408
Clinical environment						
All specialties¶	3.93	2.05–7.55	.0000	0.598	5037	12,935
Medicine¶	3.46	1.37–8.70	.0084	0.603	5663	8,464
Surgery¶	9.10	2.62–31.61	.0005	0.632	7587	2,802
Psychiatry	4.90	0.99–24.33	.0520	0.504	2714	1,348
Ancillary	0.05	0.00–4.07	.1789	0.543	100964	321
Physical environment						
All specialties¶	2.02	1.27–3.21	.0029	0.587	2406	16,710
Medicine¶	2.21	1.34–3.67	.0020	0.588	1772	11,294
Surgery	2.66	0.89–7.94	.0798	0.591	2580	3,517
Psychiatry	0.80	0.03–22.81	.8976	0.577	11357	1,505
Ancillary¶††	0.01	0.00–0.22	.0036	0.547	715327	394
Working environment						
All specialties¶	2.95	2.04–4.27	.0000	0.571	1547	17,081
Medicine¶	3.23	2.13–4.90	.0000	0.578	1393	11,500
Surgery	1.84	0.60–5.58	.2843	0.601	2652	3,622
Psychiatry¶	5.42	1.83–16.07	.0023	0.473	1339	1,554
Ancillary	4.37	0.27–70.62	.2985	0.503	1459	405

* Exponentiation of period × differencing variable interaction parameter (i.e., ratio of odds ratios), with effect size, confidence interval, standard error, and model fit computed using robust missing data methods[46,66] (see also Mathematical Appendix).

† Computed from robust standard errors.[45,46,65]

‡ Estimate of model fit.[38]

§ Measure of multicollinearity computed as the maximum condition number (CN) for the Hessian and outer product gradient of the variance/covariance matrices where CN = λ_{max}(matrix)/λ_{min}(matrix)| where λ = eigenvalues for the corresponding matrix.

¶ $P < .01$.

** Cannot be estimated.

†† Misspecified model.[45,46,65]

limits on residents' satisfaction with their educational environment. Understanding these effects can provide useful information to government agencies, accrediting bodies, teaching hospitals, and program directors in assessing the effects of duty hours limits, and to understand how residents' satisfaction with their training environments can improve as duty hours limits rules are enforced.

These findings were consistent with subanalyses conducted across domain elements, and when satisfaction scales were "cut" at different levels. However, our results both compared to and contrasted with those of previous studies. Specifically, these findings are consistent with reported associations between reduced work hours and residents' perceptions of more time to read and learn independently,[24,28,50] greater attending supervision,[28,51] and attending physicians' increased role in patient care.[52] In contrast, these findings differ from postsurveys[22–25,53] and pre–post surveys[26–28] that reported clinical experiences and patient-care quality remained unchanged, or even worsened, with fewer duty hours.

There are several possible reasons for the disparity between these survey findings and ours. First, the robust differencing variable technique applied here was designed to adjust for time trends using respondent-level controls with pre–post survey data. Such corrections, in fact, had an important effect on our study findings. For example, we found no ACGME duty hours limits effect on satisfaction rates (OR = 1.00, 95% CI [0.91, 1.11], $P = .96$) with LPS data when effect sizes were based entirely on unadjusted pre–post differences. Adjusting for time trends alone, the estimated effect size increased to an ROR of 2.13 (95% CI [1.27, 3.58], $P = .004$), and to 2.46 (95% CI [1.49, 4.05], $P < .001$) (Table 2) when estimates were further adjusted to account for differences in responder mix across periods and duty hours limits effect settings.

A second explanation for the discrepancies may involve differences in survey designs. For purposes of identifying control respondents, the LPS survey asked responders to rate satisfaction about current clinical rotations and whether duty hours limits (including limits on schedules and shifts) had an effect (good or bad) on the respondent's actual VA training environment. Postsurvey designs often focused on previous clinical training experiences and actual hours worked, which are subject to underreporting biases.[54]

Table 3

Adjusted* Estimates in Satisfaction Rates for Medicine and Surgery Resident Respondents to the Veterans Affairs Learners' Perceptions Survey (LPS) After Accreditation Council for Graduate Medical Education Duty Hours Limits, 2001–2007

	Percentage of respondents satisfied before duty hours limits†	Adjusted percentage of respondents satisfied after duty hours limits‡	Change in respondents who report satisfaction	
			Estimate	95% CI
Overall satisfaction				
Medicine	88.4	95.8	7.4%	4.8% to 9.0%
Surgery	89.9	91.8	1.9%	−34.4% to 9.1%
Faculty/ preceptor				
Medicine	84.5	95.0	10.5%	7.4% to 12.4%
Surgery	84.2	96.2	12.0%	5.8% to 14.4%
Learning				
Medicine	74.1	87.7	13.6%	7.9% to 17.7%
Surgery	79.1	89.8	10.7%	−4.2% to 17.1%
Clinical				
Medicine	57.9	82.6	24.7%	7.4% to 34.4%
Surgery	59.9	93.1	33.2%	19.7% to 38.0%
Physical				
Medicine	64.2	79.9	15.7%	6.4% to 22.6%
Surgery	68.6	85.3	16.7%	−2.6% to 25.9%
Working				
Medicine	64.6	85.5	20.9%	14.9% to 25.3%
Surgery	66.4	78.4	12.0%	−12.2% to 25.3%

* Adjusted for covariate and trend biases.
† Computed as the percentage of respondents for academic years 2001, 2002, or 2003 who reported "very satisfied" or "satisfied" on the LPS survey, computed from data in Table 1.
‡ Adjusted percentage of respondents [p_2] satisfied during post–duty hours periods computed from effect sizes (Table 2) [R] and the percentage of respondents satisfied during pre–duty hours limits periods (column 1 of this table) [p_1], or $g = [p_1 / (1 - p_1)] * [R]$, and $p_2 = g / (1 + g)$.

A third difference may be attributable to the sample and the survey design. One-third of the nation's residents rotate through VA medical centers under VA affiliation agreements with 107 U.S. medical schools,[55] with VA second only to Medicare and Medicaid as the largest funder of residency training in the United States.[56] Although VA teaching medical centers likely differ from non-VA teaching hospitals, this is the largest survey of physician resident satisfaction to date and involves a variety of facility sizes and medical school affiliations in diverse geographic areas across the United States. Furthermore, the confidential LPS survey is administered by a federal agency under strict rules of confidentiality enforced under federal oversight by the Office of Management and Budget. Promoted as an administrative tool designed to improve VA as a clinical training environment,[33,34] the LPS survey began with the 2001 academic year, three years before duty hours limits were first implemented, and one full year after full implementation of VA's quality improvement initiatives had been completed.[57,58]

Fourth, by classifying respondents individually into "effect" respondents and "no-effect" controls, we avoided aggregation errors created when respondents were grouped by educational program or facility. Overall, 36% of LPS respondents claimed that duty hours limits did not impact their VA clinical rotations during postlimits academic years (2004–2007). Such reports occurred across programs, specialties, and facilities, indicating the diversity of experiences residents encountered within the same programs and teaching facilities.

Finally, it may not be unusual to find "no-effect" environments after 2003 because some training programs had failed on occasion to adhere to mandatory duty hours rules. In one study, respondents reported exceeding the 80-hour rule at least once during six months in surgical (89%) and nonsurgical (74%) specialties while underreporting their work hours to their program directors (73% and 38%, respectively).[54] In a national survey of interns after ACGME implementation, 67% reported working shifts beyond the 30-hour rule, 43% more than the 80-hour rule, and 44% less than the one-in-seven day rule.[59] Despite having regulated resident duty hours since 1989, New York State found 54 of the state's 82 teaching hospitals were in noncompliance.[60]

The present study has certain limitations. VA clinic rotations may not necessarily represent experiences at non-VA locations. Second, respondents may not know when duty hours limits affected their training environments, thus leading to overreporting of "no effect" on the ACGME duty hours limits question. However, overreporting "no-effect" would bias estimates of duty hours limits effect sizes toward zero. Third, it is unknown whether resident satisfaction with clinical training is related to objective measures of education outcomes, such as in-service competencies examinations, board scores, and attending physician evaluations. Fourth, covariates we used to adjust for differences in respondent mix may not have controlled for all relevant factors that drive satisfaction rates. The study did not address why satisfaction may have changed, but this shift could be explained by many factors in addition to duty hours limits, including changes in workload, work life,[61] resident cross-coverage, night-float systems, redistribution of workload, reassignment of noneducational tasks to midlevel and lower-level providers,[62] clinical schedules that minimize sleep interruption,[63] or reduced in-house on-call duties. Fifth, the results are based on resident perceptions and may not necessarily reflect true differences in the quality of patient care or the effectiveness of the teaching environment. Finally, it is unknown whether further restrictions on duty schedules will continue to improve resident satisfaction.

Conclusions

In summary, applying advanced statistical methods to robust survey data, we found the 2003 ACGME mandatory duty hours limits were associated with improved training satisfaction rates. With the prospect that ACGME may adopt new standards for resident duty hours,[16] education researchers may wish to consider using the LPS survey design with robust differencing analyses to assess the impact of new standards across U.S. teaching hospitals.[64]

Dr. Kashner is with the Office of Academic Affiliations, Veterans Health Administration, Department of Veterans Affairs, Washington, DC, and is professor, Department of Medicine, Loma Linda University Medical School, Loma Linda, California, and Department of Psychiatry, University of Texas Southwestern Medical Center at Dallas, Dallas, Texas.

Mr. Henley is president, Martingale Research Corporation, Plano, Texas.

Dr. Golden is professor of cognitive science and engineering, School of Behavioral and Brain Sciences, University of Texas at Dallas, Richardson, Texas.

Dr. Byrne is associate chief of staff for education, Jerry L. Pettis Memorial VA Medical Center, Loma Linda VA Healthcare System, Loma Linda, California.

Dr. Keitz is chief of medical service, Miami VA Healthcare System, Miami, Florida, and professor, Department of Medicine, University of Miami, Miami, Florida.

Dr. Cannon is associate chief of staff for academic affiliations, George E. Wahlen VA Medical Center, Salt Lake City, Utah, and Thomas E. and Rebecca D. Jeremy Presidential and Endowed Chair for Arthritis Research, School of Medicine, University of Utah, Salt Lake City, Utah.

Dr. Chang is director of medical and dental education, Office of Academic Affiliations, Veterans Health Administration, Department of Veterans Affairs, Washington, DC.

Dr. Holland is special assistant for policy and planning, Office of Academic Affiliations, Veterans Health Administration, Department of Veterans Affairs, Washington, DC.

Dr. Aron is associate chief of staff for education, VA Senior Scholar, Louis Stokes Cleveland DVA Medical Center, Cleveland, Ohio, and professor of medicine, epidemiology and biostatistics, and organizational behavior, Weatherhead School of Management, Case Western Reserve University, Cleveland, Ohio.

Dr. Muchmore is associate chief of staff for education, VA Medical Center, San Diego, California, and professor of clinical medicine and vice chair for education, University of California at San Diego School of Medicine, San Diego, California.

Ms. Wicker is data coordinator, Jerry L. Pettis Memorial VA Medical Center, Loma Linda VA Healthcare System, Loma Linda, California.

Dr. White is Chancellor's Associates Distinguished Professor of Economics, Department of Economics, University of California at San Diego, La Jolla, California.

Acknowledgments: Sincere gratitude is expressed for the very generous review, academic direction, and support from Malcolm Cox, MD, chief academic affiliations officers, and Karen M. Sanders, MD, deputy chief academic affiliations officer, Veterans Health Administration, Department of Veterans Affairs (Washington, DC).

Gratitude is also expressed for the guidance and direction from Tetyana K. Kashner, MD, physician resident at the Pennsylvania State University Hershey School of Medicine (Hershey, Pennsylvania); Linda Godleski, MD, associate professor of psychiatry at the Yale University School of Medicine (West Haven, Connecticut); Catherine P. Kaminetzky, MD, assistant professor of medicine at the Duke University School of Medicine (Durham, North Carolina); Susan Kirsh, MD, associate professor of medicine at Case Western Reserve University School of Medicine (Cleveland, Ohio); and Edward H. Livingston, MD, professor of surgery at the University of Texas Southwestern Medical Center (Dallas, Texas).

Gratitude is also expressed for administrative and data management support from Keith Hoffman, database administrator at the Veterans Health Administration Allocation Resource Center (Braintree, Massachusetts); Robert S. Hinson, executive assistant, and Linda McInturff, Evert Melander, Cynthia Miller, and Dilpreet Singh from Veterans Health Administration Office of Academic Affiliations (Washington, DC); and Christopher T. Clarke, director, and David S. Bernett, Terry V. Kruzan, George E. McKay, and Laura Stefanowycz from the Department of Veterans Affairs Office of Academic Affiliations Data Management Center (St. Louis, Missouri).

Funding/Support: This study was funded in part by the Small Business Innovation Research (SBIR) program from the National Cancer Institute (NCI) (R44CA139607; PI: S.S. Henley) and the National Institute on Alcohol Abuse and Alcoholism (NIAAA) (R43AA014302, R43AA013670, R43/44AA013768, R43/44AA013351, R43/44AA011607; PI: S.S. Henley), and by the Department of Veterans Affairs' Health Services Research and Development Service (SHP #08-164; PI: T.M. Kashner).

Other disclosures: None.

Ethical approval: The analyses for this study were conducted for administrative purposes by, and were under the direct supervision of, the Office of Academic Affiliations, Veterans Health Administration, Department of Veterans Affairs (VA-OAA), under review by OMB Information Collection (#2900-0691) approved for VA Form #10-0439, for all data collected through January 2010.

Disclaimer: The opinions expressed herein do not necessarily reflect the views of the Department of Veterans Affairs or its affiliates, the National Cancer Institute, or the National Institute of Alcohol Abuse and Alcoholism.

References

1 Fletcher KE, Davis SQ, Underwood W, Mangrulkar RS, McMahon Jr LF, Saint S. Systematic review: Effects of resident work hours on patient safety. Ann Intern Med. 2004;141:851–857.
2 Woodrow SI, Segouin C, Armbruster J, Hamstra SJ, Hodges B. Duty hours reforms in the United States, France and Canada: Is it time to refocus our attention on education? Acad Med. 2006;81:1045–1051.
3 Landrigan CP, Rothschild JM, Cronin JW, et al. Effect of reducing interns' work hours on serious medical errors in intensive care units. N Engl J Med. 2004;351:1838–1848.
4 Veasey S, Rosen R, Barzansky B, Rosen I, Owens J. Sleep loss and fatigue in residency training: A reappraisal. JAMA. 2002;288:1116–1123.
5 Szklo-Coxe M. Are residents' extended shifts associated with adverse events? PloS Med. 2006;3:2194–2196.
6 Barger LK, Ayas NT, Cade BE, et al. Impact of extended-duration shifts on medical errors, adverse events, and attentional failures. PloS Med. 2006;3:2440–2448.
7 Barger LK, Cade BE, Ayas NT, et al. Extended work shifts and the risk of motor vehicle crashes among interns. N Engl J Med. 2005; 352:125–134.
8 Grunebaum A, Minkoff H, Blake D. Pregnancy among obstetricians: A comparison of births before, during and after residency. Am J Obstet Gynecol. 1987;157: 79–83.
9 Reuben DB. Psychologic effects of residency. South Med J. 1983;76:380–383.
10 Golub JS, Weiss PS, Ramesh AK, Ossoff RH, Johns MM 3rd. Burnout in residents of otolaryngology–health and neck surgery: A national inquiry into the health of residency training. Acad Med. 2007;82:596–601.
11 Fletcher KE, Underwood W 3rd, Davis SQ, Mangrulkar RS, McMahon LF Jr, Saint S. Effects of work hour reduction on residents' lives: A systematic review. JAMA. 2005;294: 1088–1100.
12 Rollinson DC, Rathlev NK, Moss M, et al. The effects of consecutive night shifts on neuropsychological performance of interns in the emergency department: A pilot study. Ann Emerg Med. 2003;41:400–406.
13 Hart RP, Buchsbaum DG, Wade JB, Hamer RM, Kwentus JA. Effect of sleep deprivation on first-year residents' response times, memory and mood. J Med Educ. 1987;62: 940–942.
14 Philibert I, Friedmann P, Williams WT; ACGME Work Group on Resident Duty Hours. Accreditation Council for Graduate Medical Education. New requirements for resident duty hours. JAMA. 2002;288:1112–1114.
15 Yoon HH. Adapting to duty-hour limits—Four years on. N Engl J Med. 2007;356:2668–2670.
16 Ulmer C, Wolman DM, Johns MME, eds. Resident Duty Hours: Enhancing Sleep, Supervision, and Safety. Washington, DC: Institute of Medicine; 2008.
17 Okie S. An elusive balance—Residents' work hours and the continuity of care. N Engl J Med. 2007;356:2665–2667.
18 Petersen LA, Brennan TA, O'Neil AC, Cook EF, Lee TH. Does housestaff discontinuity of

care increase the risk of preventable adverse events? Ann Intern Med. 1994;121:866–872.

19 Ratanawongsa N, Bolen S, Howell EE, Kern DE, Sisson SD, Larriviere D. Residents' perceptions of professionalism in training and practice: Barriers, promoters and duty hour requirements. J Gen Intern Med. 2006; 21:758–763.

20 Gopal RK, Carreira F, Baker WA, et al. Internal medicine residents reject longer and gentler training. J Gen Intern Med. 2007;22: 102–106.

21 Choi D, Dickey J, Wessel K, Girard DE. The impact of the implementation of work hour requirements on residents' career satisfaction, attitudes and emotions. BMC Med Educ. 2006;6:53–58.

22 Biller CK, Antonacci AC, Pelletier S, et al. The 80-hour work guidelines and resident survey perceptions of quality. J Surg Res. 2006;135: 275–281.

23 Kort KC, Pavone LA, Jensen E, Haque E, Newman N, Kittur D. Resident perceptions of the impact of work-hour restrictions on health care delivery and surgical education: Time for transformational change. Surgery. 2004;136:861–871.

24 Lin GA, Beck DC, Steward A, Garbutt JM. Resident perceptions of the impact of work hour limitations. J Gen Intern Med. 2007;22: 969–975.

25 Myers JS, Bellini LM, Morris JB, et al. Internal medicine and general surgery residents' attitudes about the ACGME duty hours regulations: A multicenter study. Acad Med. 2006;81:1052–1058.

26 Jagsi R, Shapiro J, Weissman JS, Dorer DJ, Weinstein DF. The educational impact of ACGME limits on resident and fellow duty hours: A pre–post survey study. Acad Med. 2006;81:1059–1068.

27 Goitein L, Shanafelt TD, Wipf JE, Slatore CG, Back AL. The effects of work-hour limitations on resident well-being, patient care, and education in an internal medicine residency program. Arch Intern Med. 2005;165:2601–2606.

28 Lund KJ, Teal SB, Alvero R. Resident job satisfaction: One year of duty hours. Am J Obstet Gynecol. 2005;193:1823–1826.

29 Bertrand M, Duflo E, Mullainathan S. How much should we trust differences-in-differences estimates? Q J Econ. 2004;119: 249–275.

30 Horwitz LI, Kosiborod M, Lin Z, Krumholz HM. Changes in outcomes for internal medicine inpatients after work-hour regulations. Ann Intern Med. 2007;147:97–103.

31 Shetty KD, Bhattacharya J. Changes in hospital mortality associated with residency work-hour regulations. Ann Intern Med. 2007;147:73–80.

32 Volpp KG, Rosen AK, Rosenbaum PR, et al. Mortality among patients in VA hospitals in the first 2 years following ACGME resident duty hour reform. JAMA. 2007;298:984–992.

33 Keitz SA, Holland GJ, Melander EH, Bosworth HB, Pincus SH. The Veterans Affairs Learners' Perception Survey: The foundation for educational quality improvement. Acad Med. 2003;78:910–917.

34 Cannon GW, Keitz S, Holland G, et al. Factors determining medical student and resident satisfaction during VA-based training: Findings from the VA Learners' Perception Survey. Acad Med. 2008;83:611–620.

35 Mullahy J. Interaction Effects and Difference-in-Difference Estimation in Loglinear Models. Cambridge, Mass: National Bureau of Economic Research; 1999. Technical working paper #245.

36 Efron B, Tibshirani RJ. An Introduction to the Bootstrap. New York, NY: Chapman & Hall; 1998.

37 Furnival GM, Wilson RW. Regression by leaps and bounds. Technometrics. 1974;16: 499–511.

38 Bozdogan H. Akaike's information criterion and recent developments in information complexity. J Math Psychol. 2000;44:62–91.

39 Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning. New York, NY: Springer-Verlag; 2001.

40 Rivers D, Vuong Q. Model selection tests for nonlinear dynamic models. Econom J. 2002; 5:1–39.

41 Vuong QH. Likelihood ratio tests for model selection and non-nested hypotheses. Econometrica. 1989;57:307–333.

42 Golden RM. Statistical tests for comparing possibly misspecified and nonnested models. J Math Psychol. 2000;44:153–170.

43 Golden RM. Discrepancy risk model selection test theory for comparing possibly misspecified or nonnested models. Psychometrika. 2003;68:229–249.

44 Molenberghs G, Beunckens C, Sotto C, Kenward MG. Every missingness not at random model has a missingness at random counterpart with equal fit. J R Stat Soc Series B Stat Methodol. 2008;70:371–388.

45 White H. Maximum likelihood estimation of misspecified models. Econometrica. 1982;50: 1–25.

46 Golden RM, Henley SS, White H, Kashner TM. Maximum likelihood estimation for misspecified models with missing data: Theory. Manuscript, 2010.

47 Brotherton SE, Etzel SI. Graduate medical education, 2008–2009. JAMA. 2009;302:1357–1372.

48 Khuri SF, Najjar SF, Daley J, et al. Comparison of surgical outcomes between teaching and nonteaching hospitals in the Department of Veterans Affairs. Ann Surg. 2001;234:370–383.

49 Zonia SC, LaBaere RJ, Stommel M, Tomaszewski DD. Resident attitudes regarding the impact of the 80-duty hours work standards. J Am Osteopath Assoc. 2005; 105:307–313.

50 Vaughn DM, Stout CL, McCampbell BL, et al. Three-year results of mandated work hour restrictions: Attending and resident perspectives and effects in a community hospital. Am Surg. 2008;74:542–546.

51 Kaafarani HM, Itani KM, Petersen LA, Thornby J, Berger DH. Does resident hours reduction have an impact on surgical outcomes? J Surg Res. 2005;126:167–171.

52 Harrison R, Allen E. Teaching internal medicine residents in the new era: Inpatient attending with duty hour regulations. J Gen Intern Med. 2006;21:447–452.

53 Coverdill JE, Adrales GL, Finlay W, et al. How surgical faculty and residents assess the first year of the Accreditation Council for Graduate Medical Education duty-hour restrictions: Results of a multi-institutional study. Am J Surg. 2006;191:11–16.

54 Carpenter RO, Austin MT, Tarpley JL, Griffin MR, Lomis KD. Work-hour restrictions as an ethical dilemma for residents. Am J Surg. 2006;191:527–532.

55 Leeman J, Kilpatrick K. Inter-organizational relationships of seven Veterans Affairs Medical Centers and their affiliated medical schools: Results of a multiple-case-study investigation. Acad Med. 2000;75:1015–1020.

56 Chang BK, Murawsky J, Feldman J, et al. Resident education in VA outpatient clinics: Continuity and advanced clinic access implementation. Fed Practitioner. 2007;24: 35–36, 39–41, 44–46, 54.

57 Asch SM, McGlynn EA, Hogan MM, et al. Comparison of quality of care for patients in the Veterans Health Administration and patients in a national sample. Ann Intern Med. 2004;141:938–945.

58 Jha AK, Perlin JB, Kizer KW, Dudley RA. Effect of the transformation of the Veterans Affairs Health Care System on the quality of care. N Engl J Med. 2003;348:2218–2227.

59 Landrigan CP, Barger LK, Cade BE, Ayas NT, Czeisler CA. Interns' compliance with Accreditation Council for Graduate Medical Education work-hour limits. JAMA. 2006; 296:1063–1070.

60 State Health Department Cites 54 Teaching Hospitals for Resident Working Hour Violations. Albany, NY: New York State Department of Health; June 26, 2002.

61 Vidyarthi AR, Katz PP, Wall SD, Wachter RM, Auerbach AD. Impact of reduced duty hours on residents' educational satisfaction at the University of California, San Francisco. Acad Med. 2006;81:76–81.

62 Whang EE, Mello MM, Ashley SW, Zinner MJ. Implementing resident work hour limitations: Lessons from the New York State experience. Ann Surg. 2003;237:449–455.

63 Ogden PE, Sibbitt S, Howell M, et al. Complying with ACGME resident duty hours restrictions: Restructuring the 80-hour workweek to enhance education and patient safety at Texas A&M/Scott & White Memorial Hospital. Acad Med. 2006;81:1026–1031.

64 Byrne JM, Loo LK, Giang D. Monitoring and improving resident work environment across affiliated hospitals: A call for a national resident survey. Acad Med. 2009; 84:199–205.

MATHEMATICAL APPENDIX

"ROBUST DIFFERENCING VARIABLE TECHNIQUE"

This Appendix contains a description of the Robust Differencing Variable Technique (RDV), a statistical method used to estimate ACGME mandatory duty-hour limits effects on resident satisfaction while simultaneously controlling for covariate- and trend-biases. Extending traditional difference-in-differences approaches (DD)[1, 2] and accepted methodologies,[3-7] RDV was necessary to compute effect sizes on these data for two reasons. First, information classifying respondents into "effect" and "no-effect" control settings was missing for pre-limits periods. That is, the ACGME duty-hour limits effect question was not asked during pre-limits periods for the 2001-2003 LPS surveys. Second, the final estimating model needed to adjust for covariate-biases may be misspecified.[3, 6– 11] Using the results from Golden et al.,[11] RDV estimators and statistical tests can be shown to be asymptotically unbiased in the presence of MNAR missing data and model misspecification. ACGME duty-hour limits effects may be properly inferred from model estimates provided the impact of setting on respondent satisfaction ratings is a log-linear, time-invariant, function.

Modeling Assumptions

<u>Notation.</u>

The random variable Y is a binary random variable that takes on values such that:

$$Y = \begin{cases} 0 & if \quad respondent\ is\ "not\ satisfied" \\ 1 & if \quad respondent\ is\ "satisfied" \end{cases}$$

For the binary *period* covariate d_1, respondents were administered the survey in either pre- (d_1 = 0) or post- (d_1 = 1) duty-hour limits periods. The ACGME duty-hour limits question was asked only during post-limits periods (d_1 = 1). For the binary *setting* covariate d_2, the respondent reported on the LPS survey that ACGME duty-hour limits had "no effect" (d_2 =0) or "effect" (d_2 =1) on their clinical training environment. The variables

$$d_1 = \begin{cases} 0 & if \quad \text{pre-mandatory limits period} \\ 1 & if \quad \text{post-mandatory limits period} \end{cases}$$

and

$$d_2 = \begin{cases} 0 & if \quad \text{no effect setting} \\ 1 & if \quad \text{effect setting} \end{cases}$$

are always included in all models. In some cases, k additional covariates $\mathbf{x} = [x_1, ..., x_k]$, are also included to improve predictive performance.

<u>Data Generating Process Assumptions.</u>

Let the $k+3$-dimensional vector $\boldsymbol{o}_i \equiv \left(y^i, d_1^i, d_2^i, \mathbf{x}^i\right)$ denote the ith observation, $i = 1,\ldots,n$. The $k+3$-dimensional binary vector $\boldsymbol{h}_i$ will be used to specify which elements in $\boldsymbol{o}_i$ are observable by setting the jth element of $\boldsymbol{h}_i$ equal to zero when the jth element of $\boldsymbol{o}_i$ is not observable; and setting the remaining elements of $\boldsymbol{h}_i$ equal to the value of one, $i = 1,\ldots,n$. It is assumed that $(\boldsymbol{o}_1, \boldsymbol{h}_1), ..., (\boldsymbol{o}_n, \boldsymbol{h}_n)$ is a realization of a sequence of n independent and identically distributed random variables. It is additionally assumed that y^i and d_1^i are always observable.

Researcher's Probability Model of the Complete Data.

(A1) Let $\boldsymbol{\beta} \equiv \begin{bmatrix} \beta_{1,0} & \beta_{2,0} & \left(\beta_{1,2}+\beta_{2,1}\right) & \left(\boldsymbol{\beta}_x\right)^T & \beta_0 \end{bmatrix}^T$ be a $k+4$-dimensional column vector where

$\boldsymbol{\beta}_x$ is a k-dimensional column vector. Let $p\left(Y=1 \mid d_1, d_2, \mathbf{x}; \boldsymbol{\beta}\right)$ be defined such that:

$$\log\left(\frac{p\left(Y=1 \mid d_1, d_2, \mathbf{x}; \boldsymbol{\beta}\right)}{p\left(Y=0 \mid d_1, d_2, \mathbf{x}; \boldsymbol{\beta}\right)}\right) = \beta_1 d_1 + \beta_2 d_2 + \mathbf{x}^T \boldsymbol{\beta}_x + \beta_0$$

$$\beta_1 \equiv \beta_{1,2} d_2 + \beta_{1,0}$$
$$\beta_2 \equiv \beta_{2,1} d_1 + \beta_{2,0}$$

By using the definitions of β_1 and β_2 we obtain:

$$\log\left(\frac{p\left(Y=1 \mid d_1, d_2, \mathbf{x}; \boldsymbol{\beta}\right)}{p\left(Y=0 \mid d_1, d_2, \mathbf{x}; \boldsymbol{\beta}\right)}\right) = \left(\beta_{1,2} d_2 + \beta_{1,0}\right) d_1 + \left(\beta_{2,1} d_1 + \beta_{2,0}\right) d_2 + \mathbf{x}^T \boldsymbol{\beta}_x + \beta_0$$

$$= \beta_{1,0} d_1 + \beta_{2,0} d_2 + \left(\beta_{1,2}+\beta_{2,1}\right) d_1 d_2 + \mathbf{x}^T \boldsymbol{\beta}_x + \beta_0$$
$$= \boldsymbol{\beta}^T \begin{bmatrix} d_1 & d_2 & d_1 d_2 & \mathbf{x}^T & 1 \end{bmatrix}^T.$$

Assumption A1 states that the researcher is modeling the data generating process as a logistic regression model[12] with dependent binary variable Y and covariates $d_1, d_2, d_1 d_2$, and k-dimensional covariate vector **x** when no data is missing. Also note that, ignoring the experimental context and considering the above expression from a purely formal perspective, the interaction term $\beta_{1,2}$ specifies how the impact of D_2 is influenced by D_1 while the interaction term $\beta_{2,1}$ specifies how the impact of D_1 is influenced by D_2.

When no data are missing, it is not necessary for the researcher to specify the joint distribution of the covariates. For the more general case, however, when maximum likelihood estimation in the presence of general types of data decimation mechanisms is desired, it is necessary that the researcher model the joint distribution of the covariates that are not fully observable. Let $x_{j,miss}$ denote the value of the *j*th covariate that contains missing data. Such a covariate will be referred to as a *partially observable covariate*. Ibrahim et al.[13-15] have proposed to model the covariate distribution of the partially observable covariates as a product of one-dimensional parametric conditional distributions so that:

$$p_0\left(d_2, \mathbf{x}\right) = p_0\left(d_2\right) p_0\left(x_1\right) \prod_{j=2}^{k} p_o\left(x_j \mid x_{j-1}, ..., x_1\right).$$

In addition, make the stronger assumption that the joint distribution, $p_0\left(d_2, \mathbf{x}\right)$, of the partially observable covariates may be expressed as:

(A2) Let $p_0\left(d_2, \mathbf{x}_{miss}\right) = p_0\left(d_2\right) \prod_{k} p_o\left(x_{k,miss}\right)$.

Assumption A2 states that the additional partially observable covariates in **x** will only be included in the model if they provide a source of information that is not redundant with the information source d_2 (i.e., $p_0\left(\mathbf{x}_{miss} \mid d_2\right) = p\left(\mathbf{x}_{miss}\right)$). In addition, A2 states that the *j*th partially observable covariate was added to the model only if it provided a source of information that was not redundant with the previous *j-1* partially observable covariates included in the model (i.e., $p_0\left(x_j \mid x_{j-1}, ..., x_1\right) = p\left(x_j\right)$).

It is important to emphasize that while the covariate modeling distribution A2 may not be completely satisfied in practice, our empirical investigations have shown that this choice of covariate prior resulted in the development of missing data probability models that did not evidence any signs of model misspecification. Moreover, if A2 does not hold, Golden et al.[11] provide explicit regularity conditions on the researcher's complete data model that ensures the asymptotic consistency of all estimators and statistical test results based upon the missing data probability model.

Researcher's Model of the Decimation Mechanism Assumed to be "Ignorable."

(A3) Assume Y, D_1, and some subset (possibly an empty subset) of the covariates X are observable.

(A4) Let $p\left(\boldsymbol{h}^i \mid y^i, d_1^i, d_2^i, \mathbf{x}^i\right) = p\left(\boldsymbol{h}^i \mid y^i, d_1^i, \mathbf{x}_{obs}^i\right)$ where $\mathbf{x}_{obs}^i$ denotes the covariates which are observable for the *i*th data record, $i = 1,\ldots,n$.

Assumption A4 states that that the researcher's model of the missing data has the ignorability property as defined by Golden et al.[11] (see also Little and Rubin[16] for a review). Such a property is highly desirable since estimators and statistical tests derived from an ignorable missing data model will not be biased by different forms of the resulting *data decimation mechanism model* $p\left(\boldsymbol{h}^i \mid y^i, d_1^i, \mathbf{x}_{obs}^i\right)$. Thus, because of the ignorability assumption, it is not necessary to provide a more specific specification of $p\left(\boldsymbol{h}^i \mid y^i, d_1^i, \mathbf{x}_{obs}^i\right)$.

(A5) Assume $p\left(\boldsymbol{h}^i \mid y^i, d_1^i, \mathbf{x}_{obs}^i\right)$ satisfies the constraint that whenever $d_1^i = 0$ that the value of d_2^i is not observable.

Assumption A5 shows how the decimation mechanism $p\left(\boldsymbol{h}^i \mid y^i, d_1^i, \mathbf{x}_{obs}^i\right)$ is used to represent two fundamentally distinct types of "missingness". First, we have missingness since the questionnaire in the pre-program phase (i.e., the case where $d_1 = 0$) differed from the post-program questionnaire by not including the "duty-hour limits" question that is the basis for determining the distribution of D_2. Second, we have missingness in the post-program phase (i.e., the case where $d_1 = 1$) when the question about "duty-hour limits" does in fact exist because it is possible that the distribution of D_2 may not be observable due to various factors (e.g., participants chose to not answer that question and so on). Both of these two types of missingness may be simultaneously modeled using the decimation mechanism $p\left(\boldsymbol{h}^i \mid y^i, d_1^i, \mathbf{x}_{obs}^i\right)$ since this mechanism is functionally dependent upon the observed value of $d_1^i, i = 1,\ldots,n.$ Indeed, if the decimation mechanism $p\left(\boldsymbol{h}^i \mid y^i, d_1^i, \mathbf{x}_{obs}^i\right)$ were not dependent on d_1^i (i.e., $p\left(\boldsymbol{h}^i \mid y^i, d_1^i, \mathbf{x}_{obs}^i\right) = p\left(\boldsymbol{h}^i \mid y^i, \mathbf{x}_{obs}^i\right)$) and given A5, it follows that the variable d_2^i must be eliminated from the model.

Note that if the binary variable $d_1^i = 0$ and the binary variable d_2^i is not observable, then the interaction term $d_1^i d_2^i = 0$ and is observable. To see this, note that (without any loss in generality) it may be assumed that the data generating process generates a complete data record and then subsequently decimates the complete data record. In the situation where the complete data record has $d_1^i = 0$ it is always the case that $d_1^i d_2^i = 0$. However, in the case where $d_1^i = 1$ and d_2^i is not observable, then the interaction term $d_1^i d_2^i$ must be defined as not observable since the value of $d_1^i d_2^i$ cannot be logically inferred without observing the value of d_2^i.

Semantic Interpretation of Interaction Term for the Complete Data Case.

Let $\rho\left(d_1, d_2 \mid \mathbf{x};\boldsymbol{\beta}\right) \equiv \dfrac{p\left(y=1 \mid d_1, d_2, \mathbf{x};\boldsymbol{\beta}\right)}{p\left(y=0 \mid d_1, d_2, \mathbf{x};\boldsymbol{\beta}\right)}$.

Let $r\left(d_1, d_2 \mid \mathbf{x};\boldsymbol{\beta}\right) \equiv \log\left(\rho\left(d_1, d_2 \mid \mathbf{x};\boldsymbol{\beta}\right)\right) = \left[d_1 \quad d_2 \quad d_1 d_2 \quad \mathbf{x}^T \quad 1\right]\boldsymbol{\beta}$.

Following standard methods (see Page 11 of Section V in Mullahy[2]), in the special case where no data is missing it follows that the "ratio of ratios" measures the impact of duty-hour limits on the dependent variable while controlling for the effects of time-trends and other covariates. In particular, the Ratio of Ratios (ROR) formula is defined as:

$$ROR \equiv \frac{\rho\left(d_1=1, d_2=1 \mid \mathbf{x};\boldsymbol{\beta}\right) / \rho\left(d_1=0, d_2=1 \mid \mathbf{x};\boldsymbol{\beta}\right)}{\rho\left(d_1=1, d_2=0 \mid \mathbf{x};\boldsymbol{\beta}\right) / \rho\left(d_1=0, d_2=0 \mid \mathbf{x};\boldsymbol{\beta}\right)}.$$

The log ROR may be rewritten as:

$$\log ROR = \left(r\left(d_1=1, d_2=1 \mid \mathbf{x};\boldsymbol{\beta}\right) - r\left(d_1=0, d_2=1 \mid \mathbf{x};\boldsymbol{\beta}\right)\right) - \left(r\left(d_1=1, d_2=0 \mid \mathbf{x};\boldsymbol{\beta}\right) - r\left(d_1=0, d_2=0 \mid \mathbf{x};\boldsymbol{\beta}\right)\right) =$$

$$\left(\begin{bmatrix}\beta_{1,0}\\ \beta_{2,0}\\ \beta_{1,2}+\beta_{2,1}\\ \boldsymbol{\beta}_x\\ \beta_0\end{bmatrix}^T \begin{bmatrix}1\\1\\1\\ \mathbf{x}\\1\end{bmatrix} - \begin{bmatrix}\beta_{1,0}\\ \beta_{2,0}\\ \beta_{1,2}+\beta_{2,1}\\ \boldsymbol{\beta}_x\\ \beta_0\end{bmatrix}^T \begin{bmatrix}0\\1\\0\\ \mathbf{x}\\1\end{bmatrix}\right) - \left(\begin{bmatrix}\beta_{1,0}\\ \beta_{2,0}\\ \beta_{1,2}+\beta_{2,1}\\ \boldsymbol{\beta}_x\\ \beta_0\end{bmatrix}^T \begin{bmatrix}1\\0\\0\\ \mathbf{x}\\1\end{bmatrix} - \begin{bmatrix}\beta_{1,0}\\ \beta_{2,0}\\ \beta_{1,2}+\beta_{2,1}\\ \boldsymbol{\beta}_x\\ \beta_0\end{bmatrix}^T \begin{bmatrix}0\\0\\0\\ \mathbf{x}\\1\end{bmatrix}\right) = \beta_{1,2}+\beta_{2,1}.$$

However, for the experimental context considered here, the presence or absence of the duty limit effect D_2 does not influence how the program implementation indicator factor D_1 impacts respondent satisfaction rating implying that: $\beta_{2,1}=0$. Given the *identifiability assumption* that $\beta_{2,1}=0$, it follows from the above analysis of the ROR that:

$$ROR = \exp\left(\beta_{1,2}+\beta_{2,1}\right) = \exp\left(\beta_{1,2}+0\right) = \exp\left(\beta_{1,2}\right)$$

and we may write:

$$\begin{aligned}\boldsymbol{\beta} \equiv \left[\beta_{1,0} \quad \beta_{2,0} \quad \left(\beta_{1,2}+\beta_{2,1}\right) \quad \boldsymbol{\beta}_x \quad \beta_0\right]^T &= \left[\beta_{1,0} \quad \beta_{2,0} \quad \left(\beta_{1,2}+0\right) \quad \boldsymbol{\beta}_x \quad \beta_0\right]^T \\ &= \left[\beta_{1,0} \quad \beta_{2,0} \quad \beta_{1,2} \quad \boldsymbol{\beta}_x \quad \beta_0\right]^T\end{aligned}.$$

Thus, the interaction term coefficient has the semantic interpretation of measuring how the program implementation factor D_1 influences the impact of the duty limit effect D_2 on respondent satisfaction rating.

Missing Data Theory Results.

As described by Golden et al.,[11] the maximum likelihood estimate $\hat{\boldsymbol{\beta}}_n$ of a possibly misspecified missing data model with an ignorable decimation mechanism consistent with assumptions A1-A5 may be computed using the negative log-likelihood:

$$l_n(\boldsymbol{\beta}) \equiv -n^{-1} \sum_{i=1}^{n} \left(\sum_{d_2^i, \mathbf{x}_{miss}^i} \log p\left(y^i \mid d_1^i, d_2^i, \mathbf{x}_{obs}^i, \mathbf{x}_{miss}^i; \boldsymbol{\beta}\right) p_0\left(d_2^i, \mathbf{x}_{miss}^i\right) \right)$$

by setting: $\hat{\boldsymbol{\beta}}_n = \arg\min l_n\left(\hat{\boldsymbol{\beta}}_n\right)$. We refer to $\hat{\boldsymbol{\beta}}_n$ as the RDV maximum likelihood estimate and $\hat{l}_n \equiv l_n\left(\hat{\boldsymbol{\beta}}_n\right)$ as the RDV negative log-likelihood.

Moreover, the missing data theory of Golden et al.[11] formally establishes that the RDV maximum likelihood estimate $\hat{\boldsymbol{\beta}}_n$ is an asymptotically consistent estimator with an asymptotic Gaussian distribution even if a model satisfying A1-A5 is misspecified and even if the missing data generating process is of the most general type (i.e., the data generating process is type MNAR). Furthermore, the methods of Golden et al.[11] were used to derive new asymptotically consistent RDV odds ratio estimators and new asymptotically consistent RDV statistical tests which are valid in the presence of both model misspecification and MNAR statistical environments.

References

1. Bertrand M, Duflo E, Mullainathan S. How much should we trust differences-in-differences estimates? Quarterly Journal of Economics 2004;119:249-275.

2. Mullahy J. Interaction effects and difference-in-difference estimation in loglinear models. Technical working paper #245. Cambridge MA: National Bureau of Economic Research, 1999.

3. White H. Estimation, Inference, and Specification Analyses. New York: Cambridge University Press, 1994.

4. Robins JM, Wang N. Inference for imputation estimators. Biometrika. 2000;87:113-124.

5. Qin J, Zhang B. Empirical-likelihood-based difference-in-differences estimators. Journal of the Royal Statistical Society B. 2008;70 Pt 2:329-349.

6. Golden RM. Making correct statistical inferences using a wrong probability model. Journal of Mathematical Psychology 1995;39:3-20.

7. Golden RM. Mathematical Methods for Neural Network Analysis and Design. Cambridge MA: MIT Press, 1996.

8. Golden RM. Statistical tests for comparing possibly misspecified and nonnested models. Journal of Mathematical Psychology 2000;44:153-170.

9. Golden RM. Discrepancy risk model selection test theory for comparing possibly misspecified or nonnested models. Psychometrika 2003;68:165-332.

10. White H. Maximum likelihood estimation of misspecified models. Econometrica 1982;50:1-25.

11. Golden RM, Henley SS, White H, Kashner TM. Maximum likelihood estimation for misspecified models within missing data: Theory. Manuscript, 2010.

12. Hosmer DW, Lemeshow S. Applied Logistic Regression, Second Edition. New York: John Wiley & Sons, 2000.

13. Ibrahim JG, Chen H, Lipsitz SR, Herring AH. Missing-data methods for generalized linear models: A comparative review. Journal of the American Statistical Association 2005;100(469):332-346.

14. Lipsitz SR, Ibrahim JG. A conditional model for incomplete covariates in parametric regression models. Biometrika 1996;83(4):916-922.

15. Ibrahim JG, Lipsitz SR, Chen MH. Missing covariates in generalized linear models when the missing data mechanism is nonignorable. Journal of the Royal Statistical Society Series B 1999;61:173-190.

16. Little RJA and Rubin DB. Statistical Analysis with Missing Data. 2nd Edition. New York: Wiley and Sons, 2002.