Annals of Internal Medicine | REVIEW

"July Effect": Impact of the Academic Year-End Changeover on Patient Outcomes

A Systematic Review

John Q. Young, MD, MPP; Sumant R. Ranji, MD; Robert M. Wachter, MD; Connie M. Lee, MD; Brian Niehaus, MD; and Andrew D. Auerbach, MD, MPH

Background: It is commonly believed that the quality of health care decreases during trainee changeovers at the end of the academic year.

Purpose: To systematically review studies describing the effects of trainee changeover on patient outcomes.

Data Sources: Electronic literature search of PubMed, Educational Research Information Center (ERIC), EMBASE, and the Cochrane Library for English-language studies published between 1989 and July 2010.

Study Selection: Title and abstract review followed by full-text review to identify studies that assessed the effect of the changeover on patient outcomes and that used a control group or period as a comparator.

Data Extraction: Using a standardized form, 2 authors independently abstracted data on outcomes, study setting and design, and statistical methods. Differences between reviewers were reconciled by consensus. Studies were then categorized according to methodological quality, sample size, and outcomes reported.

Data Synthesis: Of the 39 included studies, 27 (69%) reported mortality, 19 (49%) reported efficiency (length of stay, duration of procedure, hospital charges), 23 (59%) reported morbidity, and 6 (15%) reported medical error outcomes; all studies focused on inpatient settings. Most studies were conducted in the United States. Thirteen (33%) were of higher quality. Studies with higher-quality designs and larger sample sizes more often showed increased mortality and decreased efficiency at time of changeover. Studies examining morbidity and medical error outcomes were of lower quality and produced inconsistent results.

Limitations: The review was limited to English-language reports. No study focused on the effect of changeovers in ambulatory care settings. The definition of changeover, resident role in patient care, and supervision structure varied considerably among studies. Most studies did not control for time trends or level of supervision or use methods appropriate for hierarchical data.

Conclusion: Mortality increases and efficiency decreases in hospitals because of year-end changeovers, although heterogeneity in the existing literature does not permit firm conclusions about the degree of risk posed, how changeover affects morbidity and rates of medical errors, or whether particular models are more or less problematic.

Primary Funding Source: National Heart, Lung, and Blood Institute.

Ann Intern Med. 2011;155:309-315. **www.annals.org**

For author affiliations, see end of text.

All organizations experience turnover when employees leave and are replaced. Outside health care, such workforce transitions generally occur throughout the year and, at any one time, typically affect a small number of workers. Previous studies, mostly from the economics literature, have identified several factors that mediate the effect of turnover on an organization's performance, including the nature of the task (1), the degree of hierarchy within the organization (2), whether the turnover is voluntary (3) and occurs in a predictable manner (1), and the absolute level of turnover (4).

Teaching hospitals are among the few organizations (others being military units in combat and political administrations) that experience "cohort turnover": the exit of many workers coupled with a similarly sized entry of new workers at a single point in time. Cohort turnover is thought to lead to decreased productivity due to disruption in operations (5) and the loss of tacit knowledge held by the more experienced, departing workers (6). Teaching hospitals encounter cohort turnover among housestaff when experienced trainees depart at the same time that a new group of trainees enters. This annual changeover affects more than 100 000 housestaff in the United States (7) and 32 000 in Europe (8). As a result, the average experience of the teaching hospital's workforce abruptly declines, established teams are disrupted, and many of the remaining trainees are promoted and assume new roles in the care delivery process. Because of concerns that cohort turnover of residents may have a deleterious effect on patients, this transition has been called the "August killing season" in the United Kingdom and the "July phenomenon" or "July effect" in the United States (9, 10).

Several studies have examined whether patient outcomes are worse during the first months of the academic

See also:

Print

Editors' Notes 310
Editorial comment. 331
Summary for Patients. I-38

Web-Only

Appendix Tables
Conversion of graphics into slides

Context

Does mass housestaff turnover, which typically occurs during summer, adversely affect outcomes of patients admitted to teaching hospitals?

Contribution

This systematic review describes 39 studies conducted in inpatient settings that examined the effect of the academic year–end trainee changeover on patient outcomes. Larger and higher-quality studies showed increased mortality and decreased efficiency of care associated with year-end changeover.

Caution

Many studies did not account for time trends and clinical characteristics of patients. Few examined medical errors or morbidity outcomes.

Implication

Changeover that occurs when experienced housestaff are replaced with new trainees can adversely affect patient care and outcomes.

–The Editors

year, but to our knowledge there has been no systematic review of available evidence. To summarize existing literature on changeovers and the July effect, we conducted a systematic review of published literature that assessed the impact of trainee switches.

METHODS

Data Sources and Search Strategy

We developed and followed a standard protocol for conducting systematic reviews (11, 12). We searched PubMed, EMBASE, Education Resource Information Center (ERIC), and the Cochrane Library for English-language reports published between 1 January 1989, and 1 July 2010. With the assistance of a reference librarian, we developed a search strategy that used combinations of keywords and Medical Subject Heading terms related to patient care outcomes (*medical errors*; *adverse outcome*; *hospital mortality*; *quality of health care*) and teaching hospitals and clinics (*graduate medical education*; *internship* and *residency*; *academic medical centers*). In addition, we searched for titles that included relevant key phrases (*killing season*; *July effect*; *July phenomenon*). **Appendix Table 1** (available at www.annals.org) provides a detailed listing of search terms.

Study Selection

We identified studies that 1) examined the turnover of physicians-in-training (interns, residents, fellows, or their equivalent) related to the beginning of the academic year; 2) used a control group or time period as a comparator; and 3) reported the effect of the changeover on patient mortality, morbidity, medical errors, or efficiency of care. We chose these criteria to distinguish studies of housestaff cohort turnover from studies that assessed the effect of increasing physician experience on clinical outcomes (13–18).

One of 3 authors independently reviewed titles and abstracts generated by the original search to identify articles for potential inclusion. Another author re-reviewed a 5% random sample of titles to ensure accuracy. Finally, 2 authors independently reviewed the text of the studies deemed potentially eligible to make final determinations about study eligibility.

Data Extraction

Each article that met study eligibility criteria was independently abstracted by 2 authors by using a standardized form. We focused our review on the following key variables: the number of sites and patients studied, the location and type of care system, study period and duration, definition of new academic year and changeover and comparison periods, specialty studied, patient and hospital eligibility criteria, data source, type of control, sample sizes of changeover and comparison groups, statistical tests, control for confounders (such as demographic characteristics, case mix, and time trends), definition of patient care team, resident involvement in patient care, oversight structure, and primary and secondary outcomes and results. If several estimates for study outcomes were reported, the most fully adjusted estimate was abstracted.

After 2 reviewers abstracted each article, we compared the results; differences were reconciled by consensus.

Data Synthesis and Analysis

We organized study outcomes into 4 categories: mortality, morbidity (for example, perioperative complications), medical error (for example, rate of errors in laboratory ordering), and efficiency (for example, length of stay, costs, or operating room time).

We then classified studies according to the degree to which they addressed the major potential biases involved in observational research and analysis of time-series data, specifically whether the investigators 1) guarded against the possibility of differences in patient case mix between comparison periods through adjustment for patient factors, 2) used adequate statistical methods to account for clustering of effects within sites or physicians, 3) used statistical methods to account for within-year (for example, seasonal) variation in outcomes (19, 20) or between-year trends, and 4) incorporated a concurrent control group (such as a nonteaching hospital). Poor-quality studies did not adjust for possible confounding; fair-quality studies adjusted only for such patient characteristics as demographic variables and case mix; good-quality studies adjusted for patient factors and time trends (year-to-year variation or within-year seasonal variation or both); and very-good-quality studies used a concurrent control in addition to adjusting for demographic characteristics, case mix, and time trends (21). We then further combined the studies into 2 broader categories: lower quality (poor plus fair) and higher quality (good plus very good).

Role of the Funding Source

During the study, 1 author received support from the National Heart, Lung, and Blood Institute (K24HL098372). The funding source did not participate in study conception, data collection and analyses, manuscript preparation, the decision to submit the manuscript for publication, or any other part of the study.

RESULTS

Search Results

Our search identified 18 910 citations (**Figure**), of which 53 articles were considered potentially eligible on the basis of our inclusion criteria. Eight articles were identified by manual review of the reference lists of these articles. One additional article published after completion of the search was also included, resulting in a total of 62 articles that underwent full-text abstraction. Of these, 24 articles were excluded because the article contained no original data ($n = 10$), did not address the effect of the academic year–end changeover ($n = 10$), assessed only the effect on patient satisfaction ($n = 1$) (22), did not use a control group or time period as a comparator (23), or contained insufficient data to evaluate ($n = 2$) (24, 25). Agreement between reviewers for study eligibility was high (weighted $\kappa = 0.91$). Thirty-eight articles met all inclusion criteria; 1 article contained 2 separate comparisons with different methods and data sources and was treated as 2 separate studies (26, 27), resulting in a total of 39 studies for analysis.

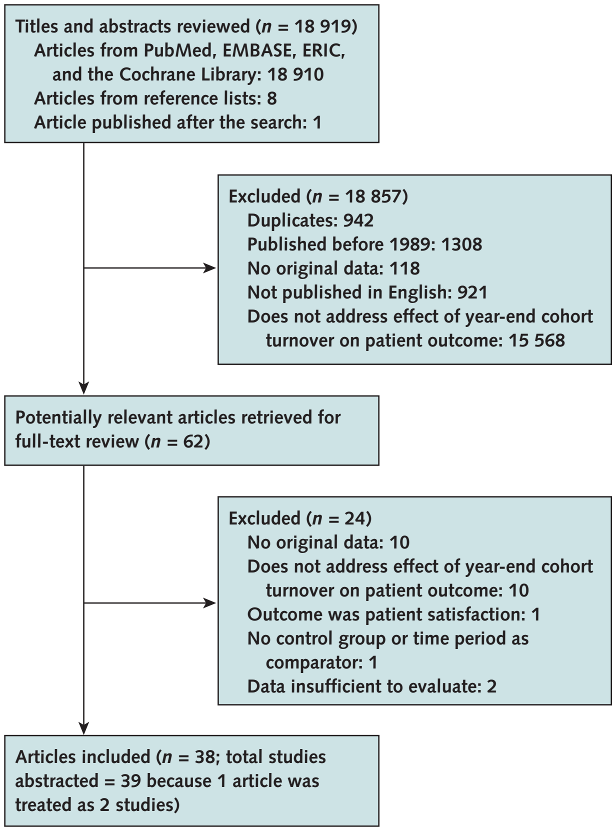
Data were extracted from the 39 studies by using the standardized form. Agreement between reviewers was moderate to high even before disagreements were reconciled through group consensus (weighted $\kappa = 0.65$ to 1.0). Because of heterogeneity of study designs, changeover systems, and outcomes, a meta-analysis was not possible.

Characteristics of Included Studies

Appendix Table 2 (available at www.annals.org) summarizes the overall characteristics of included studies, and **Appendix Table 3** (available at www.annals.org) provides detailed information about each individual study. Included studies were generally recent (published since 2000) and set in large, U.S.-based medical centers. Clinical settings included emergency departments, hospital wards, operating rooms, and intensive care units, and study participants included adult and pediatric patients. The specialties studied and their related patient populations varied considerably, including different combinations of medical (12 studies [31%]) (28–39) or surgical (19 studies [49%]) (8, 26, 27, 40–55) specialties, or both (8 studies [21%]) (9, 56–62). Twenty (53%) were single-site studies (8, 32, 34–39, 42–45, 49–52, 54, 56, 58, 62). For 57% of the studies, the study period ended before or during 2003, when duty-hour restrictions were enacted in the United States (8, 9, 26, 27, 30, 31, 33, 35–37, 39, 40, 42–44, 48, 50, 53, 56, 57, 61, 62). Most used a pre–post design with no concurrent control group (8, 9, 26–29, 32–39, 41, 42, 44, 45, 47–52, 54–56, 58, 59, 62). All studies primarily focused on care delivered in hospitals (or the emergency department) (35, 60). No study analyzed the effect on clinical care occurring in ambulatory settings, although 2 incorporated data from ambulatory settings into their overall analysis (8, 60).

Study quality varied considerably (**Table**). Twenty-four studies (62%) did not describe the patient care team sufficiently to ascertain the supervision structure and differentiated role of trainees in the delivery of patient care (9, 26–28, 30–32, 34, 39–41, 43, 46–49, 53, 54, 57–62). Twenty-eight (72%) did not specify whether the hospital provided 24-hour, onsite supervision by an attending physician (8, 9, 26–28, 30–32, 34, 36, 37, 40–43, 45–48, 52, 53, 55–57, 59–62). Heterogeneity also existed with regard to use of statistical adjustment to control for potential confounding. Many studies (16 [41%]) did not adjust for risk and were therefore considered poor quality (9, 26–28, 32, 35, 38, 39, 42–44, 48, 50, 54, 55, 62). Using the outcome with the highest degree of adjustment, 10 (26%)

Figure. **Summary of evidence search and selection.**



Databases were searched on 1 July 2010. ERIC = Educational Research Information Center.

Table. **Included Studies, Stratified by Outcome and Quality Rating**

Outcome	Quality Rating*	Studies, *n*	Total Sample Size, *n*†	Studies With "July Effect," %
Mortality	Lower	16	343 232	6
	Higher	11	64 785 686	45
Morbidity	Lower	18	468 849	17
	Higher	5	438 283	20
Efficiency	Lower	12	153 705	25
	Higher	7	538 920	57
Medical error	Lower	6	2818	50
	Higher	0	0	0

* Quality ratings: lower = poor and fair studies, as follows; higher = good and very good studies, as follows. Poor = no risk adjustment; fair = adjusted only for demographic variables and case mix; good = adjusted for demographic variables, case mix, and time trends (year-to-year variation or within-year seasonal variation), and sometimes for degree of supervision or resident involvement in patient care and clustering of outcomes; very good = adjusted for "good" rating and used a concurrent control.

† Sample size for most studies refers to number of patients. For 1 study (57), it refers to the number of hospital-months.

adjusted only for demographic variables and case mix and were considered fair quality (29, 33, 34, 37, 41, 45, 47, 49, 51, 58). Five studies (13%) adjusted for patient factors, as well as at least 1 time trend (year-to-year variation), and were designated good quality (8, 36, 52, 56, 59). Eight (21%) adjusted for patient factors and time trends and used a concurrent control (very good quality) (30, 31, 40, 46, 53, 57, 60, 61).

One study did not control for demographic factors and case mix but did control for time trends and used a concurrent control (60). On the strength of the latter, this study was categorized as good (rather than fair) quality. In addition, although Haller and colleagues' study (8) did not control for time (year or seasonal) trends, the study analyzed patient-level data linked to individual clinicians; examined only procedures performed by trainees; and used sophisticated statistical adjustments for case mix, clustering of outcomes, and level of supervision. On the basis of these unique strengths, it was categorized as good (rather than fair) quality. Overall, only 13 of 39 (33%) studies were of higher (good or very good) quality (8, 30, 31, 36, 40, 46, 52, 53, 56, 57, 59–61).

Outcomes of Included Studies

Appendix Table 4 (available at www.annals.org) summarizes the outcomes of each included study.

Mortality

Of 27 studies reporting mortality outcomes, 16 (59%) (9, 26, 28, 33, 34, 38, 41, 44, 45, 47, 49–51, 54, 56, 58) were considered lower quality and 11 (41%) (30, 31, 36, 40, 46, 52, 53, 57, 59–61) were considered higher quality (including 8 studies with concurrent controls [30, 31, 40, 46, 53, 57, 60, 61]) (**Table**). Overall, only 6 (22%) studies showed increased mortality during trainee cohort turnover compared with nonturnover months or nonteaching hospitals (40, 47, 52, 57, 59, 60). However, the proportion of studies with statistically significant worsening of mortality seemed to increase with study quality. Most (5 of 6 [83%]) studies showing an association were considered of higher methodological quality (40, 52, 57, 59, 60), and 45% (5 of 11) of higher-quality studies reported a statistically significant difference in mortality (40, 52, 57, 59, 60). In addition, most (6 of 11 [55%]) higher-quality studies (30, 40, 46, 57, 59, 60) also used a sample size large enough to detect statistically significant differences in mortality (48 000, a sample size adequate to detect a 10% difference in mortality, given a baseline mortality rate of 2.7%). Four of the 6 (67%) higher-quality, large studies (40, 57, 59, 60) reported increased mortality.

For the higher-quality studies showing an association between changeover and mortality, the effect size ranged from a relative risk increase of 4.3% (57) to 12.0% (40) or an adjusted odds ratio of 1.08 (59) to 1.34 (52). Two of the higher-quality studies reported conflicting results for mortality related to hip fracture (40, 46). In addition, of the 5 higher-quality studies that reported increased mortality, 2 reported an increase in some but not all of the mortality outcomes (52, 59). In a study comparing the first with the last week of the academic year in National Health Service hospitals in the United Kingdom, Jen and colleagues (59) showed increased mortality in medicine patients but not surgical, neoplasm, or all patients (although for the latter, the adjusted odds ratio was 1.06; $P = 0.05$). Likewise, Shuhaiber and colleagues (52) found increased mortality during changeover months for complex cardiac surgeries but not simple coronary artery bypass grafting.

Morbidity and Medical Error Outcomes

Twenty-three studies reported morbidity outcomes, such as intraoperative complications (40), undesirable anesthesia-related events (8), nursing home discharge rate (36, 61), or pneumothorax associated with central venous catheter insertion (29). Of these studies, most (18 of 23 [78%]) were of lower quality (26, 27, 29, 34, 35, 38, 41–45, 47–50, 55, 58, 62) (**Table**). Only 4 of 23 [17%] studies reported an increase in morbidity (8, 27, 47, 58); 1 of these was higher quality (8).

Six studies (26, 32, 34, 35, 39, 43) reported medical error outcomes, such as discharge with optimal medical management (34), unscheduled return visits to the emergency department (35), or error rates (26, 32, 39, 43). All were lower quality, with such weaknesses as unclear error detection methods (43) or inadequate statistical controls (for example, clustering analysis when more than one third of the errors related to 1 clinician) (39) (**Table**). Three of the 6 studies (32, 34, 39) suggested that changeovers were associated with worsened safety outcomes.

Efficiency

Length of stay, hospital charges, and such measures as operating room time were commonly reported in the 19 studies examining efficiency measures. Of these 19 studies, 7 (37%) were of higher quality (30, 36, 40, 52, 56, 57, 61) (**Table**). Overall, 7 (37%) of the studies showed decreased efficiency (30, 36, 41, 45, 47, 57, 61). As with the mortality outcomes, the proportion of studies with a statistically significant reduction in efficiency was positively associated with study quality (4 of 7 [57%]) (30, 36, 57, 61) and increasing sample size. Among the higher-quality studies showing increased length of stay, relative worsening of efficiency was between 0.3% (30) and 7.2% (61) compared with nonturnover months or nonteaching hospitals, or both. Two of these studies reported decreased efficiency in some but not all of the outcomes (30, 61).

DISCUSSION

Mortality and efficiency of care tend to worsen at the time of academic year–end changeovers, although the studies do not describe potential contributing causes or, as a result, provide specific guidance for solutions. Few studies addressed morbidity or medical errors with adequate rigor to draw firm conclusions. Of note, none of the included studies examined the effects of year-end switches on ambulatory systems.

Although our review of the literature suggests that the "July effect" exists, methodological limitations and study heterogeneity do not permit firm conclusions about the degree of risk posed and how changeover affects morbidity and rates of medical error. Studies focused on morbidity and medical error typically did not use validated surveillance systems and are therefore vulnerable to ascertainment and detection biases. In addition, most studies did not use the rigorous statistical approaches that are necessary for observational designs. Many studies did not adjust for risk (9, 26–28, 32, 35, 38, 39, 42–44, 48, 50, 54, 55, 62), and few adjusted for variation by season of the year (31, 36, 40, 53, 56, 57, 60, 61), which influences, for example, all-cause mortality (19, 20). Even fewer studies used methods appropriate for hierarchical data (8, 29, 46, 57, 59). A small number of studies used suitable concurrent controls, such as nonteaching hospitals or nonteaching services in a single hospital; this type of approach can effectively control for such factors as seasonal variation and variables that affect both teaching and nonteaching hospitals (30, 31, 40, 46, 53, 57, 60, 61). Future investigations should control for case mix and time trends; use suitable comparison groups; and consider other, more robust analytic approaches for time series data in which successive changeover samples are dependent (for example, autoregressive moving-average methods) (63).

Study heterogeneity also limited our ability to identify which features of a residency program or changeover system are most problematic. In general, studies aggregated data across patient care events or clinical settings in which the resident role in patient care varied markedly. Only a few studies gave specific information about the level of involvement of residents in the specific episode of care (8, 52) or in the clinical setting overall, by adjusting for number of residency programs in the hospital (30) or resident-to-bed or resident-to-discharge ratios (46, 57, 61). Study descriptions of the switch and associated supervisory system also varied. Most did not describe the level of supervision, and for those that did, the degree of supervision varied from 24/7 direct supervision by an attending physician (38, 49–51, 54, 58) to interns initiating supervision as needed (35). Only 1 accounted for level of supervision as a covariate (8). With 1 exception (33), studies did not report whether supervisory systems changed during the time of the changeover itself. Anecdotally, we are aware of training programs that make concerted efforts to have the "best" attending physicians on service in July or alter rounding practices to provide additional oversight for new physicians. Enhanced supervision and deployment of the more effective clinician educators may mitigate the changeover effect by providing a safety net for errors made by new trainees.

It is important to note that the "July effect" entails both a drop in the clinical experience of the physicians in the system and a decrease in the number of physicians who are familiar with the clinical system. One study found that undesirable events occurred as commonly in fifth-year trainees who were new to the hospital as in interns (8), suggesting that unfamiliarity with the working environment indepent of clinical experience may contribute to decreased quality of care. Unfortunately, our review discovered little evidence to discern to what extent worsened mortality and efficiency stem from clinical inexperience, inadequate supervision of trainees in new roles, and loss of "system knowledge" due to cohort turnover.

We found no studies that focused on changeovers in ambulatory care settings. Recent publications have identified features of the year-end outpatient turnover that may amplify risk in ambulatory settings (64, 65) and the types of errors that may occur (23). Studies in ambulatory settings will have several challenges, however. To the extent that baseline event rates are lower, larger sample sizes will be necessary to detect comparable changes. Ascertainment and detection may be more challenging because such settings offer less direct access to patients and less existing infrastructure for safety monitoring. Initial studies might focus on such outcomes as medication errors, delayed or incorrect diagnosis, or clinical decompensation (hospitalization or presentation to the emergency department) and on patient populations more vulnerable to adverse outcomes, such as those with moderate to severe chronic illness or heightened acuity.

Our study has several limitations. Our review may have been influenced by publication bias; unpublished studies may be more likely to have negative results (66). Similarly, published studies may selectively report measured outcomes and not sufficiently correct for multiple testing. Our search strategy was limited to English-language reports and did not include unpublished abstracts from conference proceedings or nonindexed journals. Although a library science expert assisted with the search, variability of terms and Medical Subject Heading terms used in the patient safety literature may have prevented the identification of a few eligible studies.

Changeovers in care teams, particularly those that result from trainee switches, raise critical questions for patients, health care systems, and training programs. The existing evidence base is problematic but frames many reasonable approaches to reducing potential harms. Not all trainees at a given level (for example, interns) pospess the same skills. Increasing emphasis on graded responsibility—in which autonomy increases with competency (67, 68)—may help ensure that individual residents are entrusted with a level of responsibility appropriate for their skill level (69). Principles of graded responsibility linked to competency assessment could be used to frame the format and goals of orientation to new roles (or a new system of care). Optimally, this sort of training would begin before the new role is assumed (potentially by using simulation or extended observation of clinical systems) and continue through the changeover. In addition, changes in the fourth year of medical school may be warranted to better prepare students for internship.

Developing changeover systems that are informed by human factor principles, such as avoiding cognitive overload and fatigue, may also have benefit. For example, hospitals may choose to reduce the initial degree of trainee workload (for example, through lower admission caps or panel sizes and use of physician extenders) and enhance supervision or increase use of multidisciplinary teams (44, 49). An alternate approach would be to take practical strategies to reduce system disruption, such as staggered schedule starts for trainees, so that abrupt changes in clinical and operational experience are avoided. Our review also outlines a rich area for several key research questions. Effective design of interventions, such as those we suggest, will require better information about causes and magnitudes of harms in a variety of clinical settings, particularly outpatient settings; this research agenda presents an opportunity for collaboration among residency programs, health system engineers, and medical center leaders. However, until efficient mitigation strategies are developed, addressing the effects of changeovers will probably require considerable resources (70).

From the University of California, San Francisco, School of Medicine, San Francisco, California.

Acknowledgment: The authors thank Gloria Won for her literature search; Amy Berlin, MD, for data abstraction; and Judith Maselli for statistical analysis.

Grant Support: By grant K24HL098372 from the National Heart, Lung, and Blood Institute (Dr. Auerbach).

Potential Conflicts of Interest: Disclosures can be viewed at www.acponline.org/authors/icmje/ConflictOfInterestForms.do?msNum=M11-1015.

Requests for Single Reprints: John Q. Young, MD, MPP, Department of Psychiatry, University of California, San Francisco, School of Medicine, 401 Parnassus Avenue, Box 0984-APC, San Francisco, CA 94143; e-mail, jqyoung@lppi.ucsf.edu.

Current author addresses and author contributions are available at www.annals.org.

References

1. **Staw BM.** The consequences of turnover. Journal of Occupational Behavior. 1980;1:253-273.
2. **Carley K.** Organizational learning and personnel turnover. Organization Science. 1992;3:20-46.
3. **Price JL.** The Study of Turnover. Ames, IA: Iowa State Univ Pr; 1977.
4. **Abelson MA, Baysinger BD.** Optimal and dysfunctional turnover: toward an organizational level model. Academy of Management Review. 1984;9:331-341.
5. **Krueger AB, Mas A.** Strikes, scabs, and tread separations: labor strife and the production of defective Bridgestone/Firestone tires. Journal of Political Economy. 2004;112:253-289.
6. **Polanyi M.** The Tacit Dimension. Garden City, NY: Doubleday; 1966.
7. **Barzansky B, Etzel SI.** Medical schools in the United States, 2008-2009. JAMA. 2009;302:1349-55. [PMID: 19773574]
8. **Haller G, Myles PS, Taffé P, Perneger TV, Wu CL.** Rate of undesirable events at beginning of academic year: retrospective cohort study. BMJ. 2009;339:b3974. [PMID: 19826176]
9. **Aylin P, Majeed FA.** The killing season—fact or fiction? BMJ. 1994;309:1690. [PMID: 7819988]
10. **Blumberg MS.** It's not OK to get sick in July [Letter]. JAMA. 1990;264:573. [PMID: 2366294]
11. **Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, et al.** Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. JAMA. 2000;283:2008-12. [PMID: 10789670]
12. **Moher D, Liberati A, Tetzlaff J, Altman DG; The PRISMA Group.** Preferred reporting items for systematic reviews and meta-analyses: the PRISMA Statement. Open Med. 2009;3:e123-e130. [PMID: 21603045]
13. **Cunningham EJ, Debar S, Bell BA.** Association between surgeon seniority and outcome in intracranial aneurysm surgery. Br J Neurosurg. 2003;17:124-8; discussion 129. [PMID: 12820753]
14. **Lawton MT, Du R.** Effect of the neurosurgeon's surgical experience on outcomes from intraoperative aneurysmal rupture. Neurosurgery. 2005;57:9-15; discussion 9-15. [PMID: 15987535]
15. **Lesar TS, Briceland L, Stein DS.** Factors related to errors in medication prescribing. JAMA. 1997;277:312-7. [PMID: 9002494]
16. **Prystowsky JB, Bordage G, Feinglass JM.** Patient outcomes for segmental colon resection according to surgeon's training, certification, and experience. Surgery. 2002;132:663-70; discussion 670-2. [PMID: 12407351]
17. **Cosgrove CM, Thornberry DJ, Wilkins DC, Ashley S.** Surgical experience and supervision may influence the quality of lower limb amputation. Ann R Coll Surg Engl. 2002;84:344-7. [PMID: 12398130]
18. **Marshall JB.** Technical proficiency of trainees performing colonoscopy: a learning curve. Gastrointest Endosc. 1995;42:287-91. [PMID: 8536893]
19. **Enquselassie F, Dobson AJ, Alexander HM, Steele PL.** Seasons, temperature and coronary disease. Int J Epidemiol. 1993;22:632-6. [PMID: 8225736]
20. **Ornato JP, Peberdy MA, Chandra NC, Bush DE.** Seasonal pattern of acute myocardial infarction in the National Registry of Myocardial Infarction. J Am Coll Cardiol. 1996;28:1684-8. [PMID: 8962552]
21. **Chan PS, Jain R, Nallmothu BK, Berg RA, Sasson C.** Rapid response teams: a systematic review and meta-analysis. Arch Intern Med. 2010;170:18-26. [PMID: 20065195]
22. **Resnick AS, Disbot M, Wurster A, Mullen JL, Kaiser LR, Morris JB.** Contributions of surgical residents to patient satisfaction: impact of residents beyond clinical care. J Surg Educ. 2008;65:243-52. [PMID: 18571141]
23. **Caines LC, Brockmeyer DM, Tess AV, Kim H, Kriegel G, Bates CK.** The revolving door of resident continuity practice: identifying gaps in transitions of care. J

Gen Intern Med. 2011 May 11. [Epub ahead of print] [PMID: 21559852]

24. **Turner PD.** Cost of care in academic health centers: the July phenomenon revisited [Abstract]. Society of General Internal Medicine 23rd Annual Meeting. Boston, Massachusetts. May 4-6, 2000. J Gen Intern Med. 2000;15:23-233.

25. **Yeaton-Massey A, Cheng Y, Handler S, Granados J, Caughey A.** 604: The July phenomenon: does it apply to obstetrics? [Abstract] Am J Obstet Gynecol. 2009;201(6 Suppl 1):S223-S223.

26. **Kestle JR, Cochrane DD, Drake JM.** Shunt insertion in the summer: is it safe? J Neurosurg. 2006;105:165-8. [PMID: 16970227]

27. **Kestle JR, Cochrane DD, Drake JM.** Shunt insertion in the summer: is it safe? J Neurosurg. 2006;105:165-8. [PMID: 16970227]

28. **Alshekhlee A, Walbert T, DeGeorgia M, Preston DC, Furlan AJ.** The impact of Accreditation Council for Graduate Medical Education duty hours, the July phenomenon, and hospital teaching status on stroke outcomes. J Stroke Cerebrovasc Dis. 2009;18:232-8. [PMID: 19426896]

29. **Ayas NT, Norena M, Wong H, Chittock D, Dodek PM.** Pneumothorax after insertion of central venous catheters in the intensive care unit: association with month of year and week of month. Qual Saf Health Care. 2007;16:252-5. [PMID: 17693670]

30. **Barry WA, Rosenthal GE.** Is there a July phenomenon? The effect of July admission on intensive care mortality and length of stay in teaching hospitals. J Gen Intern Med. 2003;18:639-45. [PMID: 12911646]

31. **Bruckner TA, Carlo WA, Ambalavanan N, Gould JB.** Neonatal mortality among low birth weight infants during the initial months of the academic year. J Perinatol. 2008;28:691-5. [PMID: 18596712]

32. **Chow KM, Szeto CC, Chan MH, Lui SF.** Near-miss errors in laboratory blood test requests by interns. QJM. 2005;98:753-6. [PMID: 16126739]

33. **Finkielman JD, Morales J, Peters SG, Keegan MT, Ensminger SA, Lymp JF, et al.** Mortality rate and length of stay of patients admitted to the intensive care unit in July. Crit Care Med. 2004;32:1161-5. [PMID: 15190967]

34. **Garcia S, Canoniero M, Young L.** The effect of July admission in the process of care of patients with acute cardiovascular conditions. South Med J. 2009;102: 602-7. [PMID: 19434039]

35. **Hartley KL, Gedde SJ, Venkatraman AS, Feuer WJ, Ajuria-Londono L.** Is it safe to have an ophthalmic emergency in July? Ophthalmic Surg Lasers Imaging. 2007;38:358-64. [PMID: 17955839]

36. **Rich EC, Gifford G, Luxenberg M, Dowd B.** The relationship of house staff experience to the cost and quality of inpatient care. JAMA. 1990;263:953-7. [PMID: 2299762]

37. **Smith JP, Mehta RH, Das SK, Tsai T, Karavite DJ, Russman PL, et al.** Effects of end-of-month admission on length of stay and quality of care among inpatients with myocardial infarction. Am J Med. 2002;113:288-93. [PMID: 12361814]

38. **Soltau TD, Carlo WA, Gee J, Gould J, Ambalavanan N.** Mortality and morbidity by month of birth of neonates admitted to an academic neonatal intensive care unit. Pediatrics. 2008;122:e1048-52. [PMID: 18977953]

39. **Walling HW, Veremakis C.** Ordering errors by first-year residents: evidence of learning from mistakes. Mo Med. 2004;101:128-31. [PMID: 15119112]

40. **Anderson KL, Koval KJ, Spratt KF.** Hip fracture outcome: is there a "July effect"? Am J Orthop (Belle Mead NJ). 2009;38:606-11. [PMID: 20145785]

41. **Bakaeen FG, Huh J, Lemaire SA, Coselli JS, Sansgiry S, Atluri PV, et al.** The July effect: impact of the beginning of the academic cycle on cardiac surgical outcomes in a cohort of 70,616 patients. Ann Thorac Surg. 2009;88:70-5. [PMID: 19559195]

42. **Banco SP, Vaccaro AR, Blam O, Eck JC, Cotler JM, Hilibrand AS, et al.** Spine infections: variations in incidence during the academic year. Spine (Phila Pa 1976). 2002;27:962-5. [PMID: 11979171]

43. **Borenstein SH, Choi M, Gerstle JT, Langer JC.** Errors and adverse outcomes on a surgical service: what is the role of residents? J Surg Res. 2004;122: 162-6. [PMID: 15555613]

44. **Claridge JA, Schulman AM, Sawyer RG, Ghezel-Ayagh A, Young JS.** The "July phenomenon" and the care of the severely injured patient: fact or fiction? Surgery. 2001;130:346-53. [PMID: 11490370]

45. **Dhaliwal AS, Chu D, Deswal A, Bozkurt B, Coselli JS, Lemaire SA, et al.** The July effect and cardiac surgery: the effect of the beginning of the academic cycle on outcomes. Am J Surg. 2008;196:720-5. [PMID: 18789415]

46. **Englesbe MJ, Fan Z, Baser O, Birkmeyer JD.** Mortality in medicare patients undergoing surgery in July in teaching hospitals. Ann Surg. 2009;249:871-6. [PMID: 19474693]

47. **Englesbe MJ, Pelletier SJ, Magee JC, Gauger P, Schifftner T, Henderson WG, et al.** Seasonal variation in surgical outcomes as measured by the American College of Surgeons-National Surgical Quality Improvement Program (ACS-NSQIP). Ann Surg. 2007;246:456-62. [PMID: 17717449]

48. **Ford AA, Bateman BT, Simpson LL, Ratan RB.** Nationwide data confirms absence of 'July phenomenon' in obstetrics: it's safe to deliver in July. J Perinatol. 2007;27:73-6. [PMID: 17262037]

49. **Highstead RG, Johnson LS, Johnson LC, Street JH 3rd, Trankiem CT, Kennedy SO, et al.** July—as good a time as any to be injured. J Trauma. 2009; 67:1087-90. [PMID: 19901672]

50. **Myles TD.** Is there an obstetric July phenomenon? Obstet Gynecol. 2003; 102:1080-4. [PMID: 14672490]

51. **Schroeppel TJ, Fischer PE, Magnotti LJ, Croce MA, Fabian TC.** The "July phenomenon": is trauma the exception? J Am Coll Surg. 2009;209:378-84. [PMID: 19717044]

52. **Shuhaiber JH, Goldsmith K, Nashef SA.** Impact of cardiothoracic resident turnover on mortality after cardiac surgery: a dynamic human factor. Ann Thorac Surg. 2008;86:123-30. [PMID: 18573410]

53. **Smith ER, Butler WE, Barker FG 2nd.** Is there a "July phenomenon" in pediatric neurosurgery at teaching hospitals? J Neurosurg. 2006;105:169-76. [PMID: 16970228]

54. **Stout CL, Chapman JR, Scoglietti VC, Long EL, Vaughn DM, Abercrombie CL, et al.** "July Effect": an evaluation of a level I teaching hospital's trauma service seasonal mortality rates [Letter]. Am Surg. 2008;74:878-9. [PMID: 18807684]

55. **Yaghoubian A, de Virgilio C, Chiu V, Lee SL.** "July effect" and appendicitis. J Surg Educ. 2010;67:157-60. [PMID: 20630426]

56. **Buchwald D, Komaroff AL, Cook EF, Epstein AM.** Indirect costs for medical education. Is there a July phenomenon? Arch Intern Med. 1989;149:765-8. [PMID: 2495778]

57. **Huckman RS, Barro J, National Bureau of Economic Research.** Cohort turnover and productivity. The July phenomenon in teaching hospitals. NBER Working Paper Series no. w11182. Cambridge, MA: National Bureau of Economic Research; 2005.

58. **Inaba K, Recinos G, Teixeira PG, Barmparas G, Talving P, Salim A, et al.** Complications and death at the start of the new academic year: is there a July phenomenon? J Trauma. 2010;68:19-22. [PMID: 20065752]

59. **Jen MH, Bottle A, Majeed A, Bell D, Aylin P.** Early in-hospital mortality following trainee doctors' first day at work. PLoS One. 2009;4:e7103. [PMID: 19774078]

60. **Phillips DP, Barker GE.** A July spike in fatal medication errors: a possible effect of new medical residents. J Gen Intern Med. 2010;25:774-9. [PMID: 20512532]

61. **Rich EC, Hillson SD, Dowd B, Morris N.** Specialty differences in the 'July Phenomenon' for Twin Cities teaching hospitals. Med Care. 1993;31:73-83. [PMID: 8417272]

62. **Shulkin DJ.** The July phenomenon revisited: are hospital complications associated with new house staff? Am J Med Qual. 1995;10:14-7. [PMID: 7727983]

63. **Ramsay CR, Matowe L, Grilli R, Grimshaw JM, Thomas RE.** Interrupted time series designs in health technology assessment: lessons from two systematic reviews of behavior change strategies. Int J Technol Assess Health Care. 2003;19: 613-23. [PMID: 15095767]

64. **Young JQ, Wachter RM.** Academic year-end transfers of outpatients from outgoing to incoming residents: an unaddressed patient safety issue. JAMA. 2009;302:1327-9. [PMID: 19773568]

65. **Young JQ, Niehaus B, Lieu SC, O'Sullivan PS.** Improving resident education and patient safety: a method to balance initial caseloads at academic year-end transfer. Acad Med. 2010;85:1418-24. [PMID: 20736670]

66. **Dickersin K.** The existence of publication bias and risk factors for its occurrence. JAMA. 1990;263:1385-9. [PMID: 2406472]

67. **Sterkenburg A, Barach P, Kalkman C, Gielen M, ten Cate O.** When do supervising physicians decide to entrust residents with unsupervised tasks? Acad Med. 2010;85:1408-17. [PMID: 20736669]

68. **ten Cate O, Snell L, Carraccio C.** Medical competence: the interplay between individual ability and the health care environment. Med Teach. 2010;32: 669-75. [PMID: 20662579]

69. **Green ML, Aagaard EM, Caverzagie KJ, et al.** Charting the road to competence: developmental milestones for internal medicine residency training. Journal of Graduate Medical Education. 2009;1:5-20.

70. **Nuckols TK, Bhattacharya J, Wolman DM, Ulmer C, Escarce JJ.** Cost implications of reduced work hours and workloads for resident physicians. N Engl J Med. 2009;360:2202-15. [PMID: 19458365]

Current Author Addresses: Dr. Young: Department of Psychiatry, University of California, San Francisco, School of Medicine, 401 Parnassus Avenue, Box 0984-APC, San Francisco, CA 94143.
Drs. Ranji and Auerbach: Division of Hospital Medicine, Department of Medicine, University of California, San Francisco, School of Medicine, 533 Parnassus Avenue, Box 0131, San Francisco, CA. 94143-0131.
Dr. Wachter: Division of Hospital Medicine, Department of Medicine, University of California, San Francisco, School of Medicine, 505 Parnassus Avenue, Box 0120, San Francisco, CA 94143-0120.
Drs. Lee and Niehaus: Department of Psychiatry, University of California, San Francisco, School of Medicine, 401 Parnassus Avenue, Box 0984-RTP, San Francisco, CA 94143.

Author Contributions: Conception and design: J.Q. Young, R.M. Wachter, B. Niehaus, A.D. Auerbach.
Analysis and interpretation of the data: J.Q. Young, S.R. Ranji, C.M. Lee, B. Niehaus, A.D. Auerbach.
Drafting of the article: J.Q. Young, S.R. Ranji, A.D. Auerbach.
Critical revision of the article for important intellectual content: J.Q. Young, S.R. Ranji, R.M. Wachter, C.M. Lee, A.D. Auerbach.
Final approval of the article: J.Q. Young, S.R. Ranji, R.M. Wachter, A.D. Auerbach.
Provision of study materials or patients: J.Q. Young.
Statistical expertise: A.D. Auerbach.
Obtaining of funding: A.D. Auerbach.
Administrative, technical, or logistic support: J.Q. Young, B. Niehaus, A.D. Auerbach.
Collection and assembly of data: J.Q. Young, S.R. Ranji, C.M. Lee, B. Niehaus, A.D. Auerbach.

Appendix Table 1. **Search Strings by Database**

Database	Search String
PubMed	((medical errors OR hospital mortality OR quality of health care(mh:noexp) OR quality of health care/og OR quality of health care/standards OR ″quality of health care/statistics and numerical data″ OR quality of healthcare/economics) AND (graduate medical education OR ″internship and residency″ OR academic medical centers)) OR (KILLING(TI) AND SEASON*(TI) OR (JULY(TI) AND PHENOMENON(TI)) OR (JULY(TI) AND EFFECT(TI)))
EMBASE	1. ((′medical error′/exp OR ′mortality′/exp OR ′adverse outcome′/exp OR ′health care quality′/de OR ′clinical indicator′/exp OR ′performance measurement system′/exp OR ′treatment outcome′/exp) AND (′medical education′/mj OR ′clinical education′/exp/mj OR ′residency education′/exp/mj OR ′surgical training′/exp/mj OR ′resident′/mj)) OR 2. ((′medical education′/de OR ′clinical education′/exp OR ′residency education′/exp OR ′surgical training′/exp OR ′resident′/de) AND (′medical error′/exp/mj OR ′mortality′/exp/mj OR ′adverse outcome′/exp/mj OR ′health care quality′/mj OR ′clinicalindicator′/exp/mj OR ′performance measurementsystem′/exp/mj OR ′treatment outcome′/exp/mj)) OR 3. (′teaching hospital′ OR ′teaching hospitals′ AND (′resident′/exp OR residents OR ′residency education′/exp OR physician* NEAR/3 trainee* OR doctor* NEAR/3 trainee*) AND (′mortality′/exp OR ′outcome assessment′/exp)) OR 4. ′killing season′ OR ′july effect′ OR ′july phenomenon′
ERIC	1. KW=(error OR errors OR mortality OR "quality of care″ OR ″quality of health care″ OR ″adverse outcome″ or ″adverse event″ OR morbidity OR ″health care quality″ OR ″clinical indicator″ OR ″performance measurement″ OR ″treatment outcome″) and DE=(medical education OR graduate medical education OR clinical experience) 2. killing season OR july effect OR july phenomenon
Cochrane Library	(medical errors OR hospital mortality OR quality of health care) AND (graduate medical education OR internship and residency OR academic medical centers)

DE = descriptor; ERIC = Educational Research Information Center; KW = keywords; TI = title.

Appendix Table 2. **Characteristics of 39 Included Studies***

Characteristic	Studies, *n* (%)
Publication year	
1989–1994	4 (10)
1995–1999	1 (3)
2000–2004	8 (21)
2005–June 2010	26 (67)
Location	
United States	30 (77)
Canada	3 (8)
United Kingdom	3 (8)
Australia	1 (3)
China	1 (3)
Multinational	1 (3)
Number of sites	
Single site	20 (51)
Multiple sites	19 (49)
Study duration	
1 y	4 (10)
Multiyear	35 (90)
Specialty	
Anesthesia	1 (3)
Critical care	3 (8)
Internal medicine	
General	2 (5)
Cardiology	1 (3)
Multiple medical	2 (5)
Neurology	1 (3)
Multispecialty (medical and surgical)	7 (18)
Pediatrics and neonatology	2 (5)
Surgery	
Cardiothoracic surgery	3 (8)
General	1 (3)
Multiple surgical	2 (5)
Neurosurgery	1 (3)
Obstetrics	2 (5)
Ophthalmology	1 (3)
Orthopedic surgery	2 (5)
Pediatric general and neurosurgery	3 (8)
Trauma surgery	5 (13)
Study design	
Post hoc analysis of RCT data	1 (3)
Prospective	1 (3)
Retrospective, pre–post	29 (74)
Retrospective, pre–post with concurrent control	8 (21)
Adjustment for possible confounders†	
No adjustment	16 (41)
Demographic variables and case mix	10 (26)
Demographic variables and case mix, within-year seasonal variation, or year-to-year time trends (with or without accounting for clustering of outcomes)	11 (28)
Other	2 (5)
Outcome‡	
Mortality	27 (69)
Morbidity	23 (59)
Medical error	6 (15)
Efficiency (e.g., length of stay)	19 (49)

RCT = randomized, controlled trial.
* One article was treated as 2 studies (26).
† Each study was characterized by its most risk-adjusted outcome.
‡ Total number of outcomes greater than number of studies because some studies had outcomes in >1 category.

Appendix Table 3. Characteristics of Included Studies

Study, Year, Location (Reference)	Study Start–Study End	Sites (Hospital Type), *n*	Patient Characteristics (Medical Specialty)	Study Design (Data Source*)	Team Described (24-h Super-vision by Attending Physician)	Changeover Period (Sample Size, *n*)	Comparison Period (Sample Size, *n*)	Variables Adjusted for	Quality Rating†
Alshekhlee et al, 2009, USA (28)	2000–2005	Multiple (THs)	Adult, admitted with acute ischemic stroke (neurology)	Retrospective, pre–post (NIS)	No (NR)	July (NR)	Individual non-July months (NR) Total sample, 162 393	No adjustment	1
Anderson et al, 2009, USA (40)	1998–2003	Multiple (THs and NTHs)	Adult (age ≥65 y), hospitalized for femoral neck or inter-trochanteric fracture (orthopedic surgery)	Retrospective, pre–post, with CC in the NTH (NIS)	No (NR)	July–August (NR)	July–June (TH: 119 792; NTH: 205 196)	D, CM, T, ST, and region	4
Ayas et al, 2007, Canada (29)	1999–2005	2 ICUs in 2 hospitals (THs)	Adults admitted to ICU who received new central venous catheter into subclavian or internal jugular vein; excluded rethreading of existing CVCs and CVCs placed outside of ICU (critical care)	Retrospective, pre–post (ICU APACHE II database)	Yes (No)	July–August (NR)	September–June (NR) Total: 3548 patients who received 5816 CVCs	D, CM, ICU site, CL-P	2
Aylin and Majeed, 1994, UK (9)	1983–1992	Multiple (THs)	All patients who died during hospital stay (multispecialty)	Retrospective, pre–post (NHS)	No (NR)	August 1–7 (number of admissions not known)	July 25–31 (number of admissions not known)	No adjustment	1
Bakaeen et al, 2009, USA (41)	1997–2007	44 (THs)	Adults who underwent open cardiac surgical procedure at VA Cardiac Surgery Program: CABG, aortic or mitral valve surgery, and great vessel surgery (cardiothoracic surgery)	Retrospective, pre–post (VA-CICSP)	No (NR)	July–August (11 975)	September–June (58 641)	MT and MB: D, CM, U. EFF: no adjustment	MT: 2 MB: 2 EFF: 1
Banco et al, 2002, Philadelphia, USA (42)	1994–1997	1 (TH)	Patients who underwent spine surgery from orthopedic surgical service; excluded if neurosurgery service was involved (orthopedic surgery)	Retrospective, pre–post (hospital infection records)	Yes (NR)	July (115 surgeries) and August (108 surgeries)	September–June by month (total of 1101 surgeries)	No adjustment	1
Barry and Rosenthal, 2003, Ohio, USA (30)	1991–1997	38 ICUs in 28 hospitals (5 major THs and 23 minor THs or NTHs)‡	Adult (age ≥16 y), admitted to ICU; excluded burn injuries, death within 1 h, or admission for dialysis or status-post heart surgeries (critical care)	Retrospective, pre–post with CC (minor THs and NTHs were comparison groups, not control) (Regional APACHE III database)	No (NR)	Q1: July–September (12 365)	Entire year (48 853) and Q2 (12 029), Q3 (12 088), and Q4 (12 371)	D, CM, T, RI, site (secondary analysis only)	4
Borenstein et al, 2004, Toronto, Canada (43)	2002	1 (TH)	Pediatric (age <18 y), admitted to either of 2 pediatric surgeons (pediatric surgery)	Prospective, pre–post (daily chart audit by patient care team with error reporting form)	No (NR)	July (108)	June (84)	No adjustment	1
Bruckner et al, 2008, California, USA (31)	1999–2003	38 NICUs (19 THs and 19 NTHs)	Singleton infants with very-low-birthweight (<1500 g) and moderately low birthweight (1500–2499 g), admitted to NICU (pediatric neonatology)	Retrospective, pre–post with CC (NTH NICUs matched by number of deliveries) (California Birth Cohort File)	No (NR)	July–August (872 infants weighing <1500 g and 2748 infants weighing 1500–2499 g)	September–June (4312 infants weighing <1500 g and 12 484 infants weighing 1500–2499 g)	D, CM, T, ST	4
Buchwald et al, 1989, Boston, USA (56)	1982–1984	1 (TH)	Adult, admitted with 1 of 20 DRGs: the study site's most common 10 medical and 10 surgical DRGs (multispecialty)	Retrospective, pre–post (hospital discharge abstract and itemized bills)	Yes (NR)	July–August (1251 total, 543 medical, 708 surgical)	April–May (1338 total, 619 medical, 719 surgical)	MT: no adjustment EFF: D, CM, U, T (charges only), ST	MT: 1 EFF: 3
Chow et al, 2005, Hong Kong (32)	2002–2004	1 (TH)	Adult, admitted to general medicine, medicine subspecialty, or neurology inpatient service; excluded patients admitted to ICU (multiple medical specialties)	Retrospective, pre–post (errors confidentially and voluntarily reported within 24 h of occurrence)	No (NR)	July (13 interns)	August–June (143 intern-months)	No adjustment	1
Claridge et al, 2001, Virginia, USA (44)	1994–1999	1 (TH)	Adult, admitted to trauma service (trauma surgery)	Retrospective, pre–post (trauma registry plus infectious disease laboratory database)	Yes (No)	July–August (506)	April–May (411)	No adjustment	1

Continued on following page

Appendix Table 3—Continued

Study, Year, Location (Reference)	Study Start–Study End	Sites (Hospital Type), *n*	Patient Characteristics (Medical Specialty)	Study Design (Data Source*)	Team Described (24-h Supervision by Attending Physician)	Changeover Period (Sample Size, *n*)	Comparison Period (Sample Size, *n*)	Variables Adjusted for	Quality Rating†
Dhaliwal et al, 2008, Houston, USA (45)	1997–2007	1 (TH, VA)	Adult, underwent open cardiac surgical procedure: CABG, aortic or mitral valve surgery, left ventricular aneurysm repair, or great vessel surgery (cardiothoracic surgery)	Retrospective, pre–post (VA-CICSP)	Yes (NR)	July–August (242)	September–June (1431)	D, CM, U	2
Englesbe et al, 2007, USA (47)	2001–2004	18 (14 THs, 4 NR)	All ages, underwent "major" operation under general, spinal, or epidural anesthesia at academic medical centers or major general or vascular surgery at large, private community hospitals (multiple surgical specialties)	Retrospective, pre–post (ACS-NSQIP)	No (NR)	July–August (9941)	April 15–June 15 (10 313)	MT and MB: D, CM, U EFF: no adjustment	MT: 2 MB: 2 EFF: 1
Englesbe et al, 2009, USA (46)	2003–2006	Multiple (THs and NTHs)§	Adult/geriatric (age 65–99 y), Medicare-insured, underwent 1 of 7 surgical procedures: CABG, elective AAA, carotid endarterectomy, pancreatectomy, esophagectomy, colectomy for colon cancer, or surgical repair of hip fracture (multiple surgical specialties)	Retrospective, pre–post with CC; NTHs were comparison group, not control (CMS Inpatient Files)	No (NR)	July (THs = 29 198; NTHs = NR)	August–June (THs = 294 018; THs and NTHs = NR) Total for NTHs: 1 245 342	D, CM, U, T, CL-H, RI (Secondary analysis only)	4
Finkielman et al, 2004, Rochester, Minnesota, USA (33)	1994–2002	4 ICUs from 2 hospitals at 1 medical center (TH)	Adult, first admission to 1 of 4 ICUs: 1 medical, 2 surgical (general surgery/trauma and thoracic/vascular surgery), and 1 multispecialty (critical care)	Retrospective, pre–post (ICU APACHE III database)	Yes (No)	July (2728)	Individual non-July months (total sample size = 26 356)	D, CM	2
Ford et al, 2007, USA (48)	1998–2002	Multiple (TH, verified to have residents on labor delivery service)	Medicaid-insured pregnant women undergoing delivery of singleton gestations (obstetrics-gynecology)	Retrospective, pre–post (NIS)	No (NR)	July (26 546)	August–June (272 584)	No adjustment	1
Garcia et al, 2009, Miami, USA (34)	ACS: 2000–2002 CHF: 2002–2004	1 (TH)	Adult, admitted with diagnosis of non–ST-segment elevated ACS who underwent coronary angiography or with decompensated CHF (internal medicine and cardiology)	Retrospective, pre–post (NR)	No (NR)	July–September (191 with ACS and 93 with CHF)	October–June (573 with ACS and 516 with CHF)	MT: no adjustment MB: no adjustment EFF: no adjustment ME: D, CM	MT: 1 MB: 1 EFF: 1 ME: 2
Haller et al, 2009, Melbourne, Australia (8)	1995–2000	1 (TH)	Adult, underwent inpatient or ambulatory anesthetic procedure carried out by anesthesia registrar; excluded obstetrics patients (anesthesia)	Retrospective, pre–post (patient data: hospital administrative databases [cross-checked with handwritten medical records]; undesirable events: electronic incident-reporting system, completion mandatory after each procedure)	Yes (NR)	First one twelfth of procedures performed by trainee during first year at hospital (2555)	Last eleven twelfth of procedures performed by trainee during first year at hospital (17 005)	D, CM, U, LS, RI, CL-C, CL-I	3‖ (Did not adjust for T and ST)

Continued on following page

Appendix Table 3—Continued

Study, Year, Location (Reference)	Study Start–Study End	Sites (Hospital Type), *n*	Patient Characteristics (Medical Specialty)	Study Design (Data Source*)	Team Described (24-h Supervision by Attending Physician)	Changeover Period (Sample Size, *n*)	Comparison Period (Sample Size, *n*)	Variables Adjusted for	Quality Rating†
Hartley et al, 2007, Miami, USA (35)	2002–2003 (1 AY)	1 (TH)	Adult, received eye care from intern in emergency department of eye institute (ophthalmology)	Retrospective, pre–post (medical records)	Yes (No)	July 15–August 14 (180 patients, 6 interns)	June 1–June 30 (180 patients, 6 interns)	No adjustment	1
Highstead et al, 2009, Washington, DC, USA (49)	1998–2007	1 (TH)	All ages, admitted to trauma service (trauma surgery)	Retrospective, pre–post (trauma registry)	No (Yes)	July–August (3967)	April–May (3626)	MT: D, CM MB: no adjustment EFF: no adjustment	MT: 2 MB: 1 EFF: 1
Huckman and Barro, 2005, USA (57)	1993–2001	Multiple (NTHs, minor THs, major THs)	All patients admitted to study hospitals (multispecialty)	Retrospective, pre–post with CC (NTHs) (NIS and AHA's Annual Survey of Hospitals)	No (NR)	July–August (NR)	April–May (NR) Total hospital-months = 74 521	D, CM, T, ST, CL-H, RI	4
Inaba et al, 2010, Los Angeles, USA (58)	2001–2006	1 (TH)	All patients admitted to hospital with injury (trauma surgery)	Retrospective, pre–post (hospital trauma registry and morbidity and mortality records)	No (Yes)	July–August (4030)	May–June (4121)	D, CM	2
Jen et al, 2009, UK (59)	2000–2008	Multiple (THs)	Patients admitted with an emergency (multispecialty)	Retrospective, pre–post (NHS)	No (NR)	Week starting first Wednesday in August (147 897)	Week starting last Wednesday in July (151 844)	D, CM, T, CL-H	3
Kestle et al, 2006, Canada (27)	1989–2001	11 (THs)	All patients, underwent ventricular shunt insertion or revision in an English-speaking hospital in Canada (neurosurgery)	Retrospective, pre–post (admission data from hospitals in English-speaking Canada)	No (NR)	July–August (490)	September–June (2578)	No adjustment	1
Kestle et al, 2006, Canada, USA, and Europe (26)	SDT: 1993–1995 ESIT: 1996–1999	26 (THs)	Pediatric (age <18 y), underwent shunt insertion surgery to treat hydrocephalus (pediatric neurosurgery)	Retrospective, pre–post (post hoc analysis of 2 RCTs) (multicenter hydrocephalus clinical trials database)	No (NR)	July–August (138 for outcomes from 2 trials combined, 60 for outcomes from SDT only, 57 for outcomes from ESIT only)	September–June (599 for outcomes from 2 trials combined, 284 for outcomes from SDT only, and 248 for outcomes from ESIT only)	No adjustment	1
Myles, 2003, Chicago, USA (50)	1996–1999	1 (TH)	Admitted to obstetric service with gestational age >20 wk (obstetrics-gynecology)	Retrospective, pre–post (medical records)	Yes (Yes)	July (745) July–September (2139) July–December (4082)	August–June (7069) October–June (5675) January–June (3732)	No adjustment	1
Phillips and Barker, 2010, USA (60)	1979–2006	Multiple (THs and NTHs)	All patients who died inside medical institutions (inpatient, outpatient, ED) because of preventable medication error that was primary cause of death (multispecialty)	Retrospective, pre–post with CC (counties without THs) (U.S. computerized death certificates: data at level of county, not hospital; AHA surveys: proportion of hospitals that are major TH by county; and other secondary datasets)	No (NR)	July (NR)	Individual non-July months (NR) Total deaths = 62 338 584	T, ST	4 (Did not adjust for D or CM)¶
Rich et al, 1990, St. Paul, Minnesota, USA (36)	1980–1986	1 (TH)	Adult, admitted to internal medicine service and length of stay <14 d (internal medicine)	Retrospective, pre–post (hospital financial database)	Yes (NR)	Days since June 24 (NA)	Days since June 24 (NA) Total = 21 679	D, CM, T, ST	3

Continued on following page

Appendix Table 3—Continued

Study, Year, Location (Reference)	Study Start–Study End	Sites (Hospital Type), *n*	Patient Characteristics (Medical Specialty)	Study Design (Data Source*)	Team Described (24-h Supervision by Attending Physician)	Changeover Period (Sample Size, *n*)	Comparison Period (Sample Size, *n*)	Variables Adjusted for	Quality Rating†
Rich et al, 1993, St. Paul and Minneapolis, Minnesota, USA (61)	1983–1987	NR (3 THs and unclear number of NTHs)**	Adult, discharged with 1 of 25 different ICD-9 discharge diagnoses (multispecialty)	Retrospective, pre–post with CC (NTH) (regional trade association)	No (NR)	Days since June 24 (25 000 sample of the 240 476 in dataset)	Days since June 24 (25 000 sample of the 240 476 total in dataset)	D, CM, T, ST, RI (secondary analysis only)	4
Schroeppel et al, 2009, Memphis, Tennessee, USA (51)	2001–2006	1 (TH)	Admitted to trauma service after blunt injury and discharged during same month; excluded patients admitted in cardiac arrest or intubated for <2 d (trauma surgery)	Retrospective, pre–post (trauma registry)	Yes (Yes)	July (1348) Q1: July–September (4140)	Individual non-July months (total = 11 177) Q2 (3662), Q3 (3321), and Q4 (3675)	MT: D, CM EFF: no adjustment	MT: 2 EFF: 1
Shuhaiber et al, 2008, Cambridge, UK (52)	1996–2006	1 (TH)	Adult, undergoing first cardiac operation: complex CABG, thoracic aortic surgery, valve replacement, aortic surgery and valve replacement, and simple CABG; excluded catheter-based interventions and surgical closure of atrial septal defects (cardiothoracic surgery)	Retrospective, pre–post (quality assurance database)	Yes (NR)	Trainee change months: July, August, January, and February (5517 total, 3481 CABG only, 2036 complex cardiac operations) Registrar-performed surgery first month: July and February (735 total cases)	Trainee nonchange months: September–December and March–June (10 773 total, 6782 CABG only, 3991 complex cardiac operations) Registrar-performed surgery last months: June and January (755 total cases)	D, CM, T, RI	3
Shulkin, 1995, Philadelphia, USA (62)	1991	1 (TH)	All patients discharged from study hospital (multispecialty)	Retrospective, pre–post (medical records abstracted by trained analysts blinded to purpose of study)	No (NR)	July–December (14 295)	January–June (14 246)	No adjustment	1
Smith et al, 2006, USA (53)	1998–2000	Multiple (THs and NTHs)	Pediatric (age ≤18 y), underwent CSF shunt surgery or craniotomy for tumor resection; excluded patients transferred from other hospitals and age <90 d for shunt surgery (pediatric neurosurgery)	Retrospective, pre–post with CC in secondary analysis [NTH] (NIS)	No (NR)	July–August (NR)	September–June (NR) Total sample size = 4322 craniotomies (3002 at THs) and 22 056 shunt placements or revisions (14 975 at THs)	D, CM, U, T, ST (secondary analysis only), CL-H, hospital region	4
Smith et al, 2002, Ann Arbor, Michigan, USA (37)	1995–1998	1 (TH)	Adult, admitted with MI to inpatient cardiology service from ED, physician's office, or other hospitals; excluded in-hospital transfers and stays <4 d (cardiology)	Retrospective, pre–post (medical records via trained abstractors)	Yes (NR)	July–September (NR)	April–June (NR) Total = 782	D, CM	2
Soltau et al, 2008, Birmingham, Alabama, USA (38)	1991–2004	1 (TH)	Pediatrics (infants), admitted to NICU after birth at gestational age >24 wk; excluded neonates who died in delivery room (pediatric neonatology)	Retrospective, pre–post (medical record abstracted by trained database specialist)	Yes (Yes, starting in 2001)	July–December (NR) July (NR)	January–June (NR) Individual non-July months (NR) Total = 3445 infants with birthweight 401–1500 g and 7840 infants with birthweight >1500 g	No adjustment	1

Continued on following page

Appendix Table 3—Continued

Study, Year, Location (Reference)	Study Start–Study End	Sites (Hospital Type), *n*	Patient Characteristics (Medical Specialty)	Study Design (Data Source*)	Team Described (24-h Supervision by Attending Physician)	Changeover Period (Sample Size, *n*)	Comparison Period (Sample Size, *n*)	Variables Adjusted for	Quality Rating†
Stout et al, 2008, Macon, Georgia, USA (54)	2003–2006	1 (TH)	All ages, admitted to trauma service (trauma surgery)	Retrospective, pre–post (trauma database)	No (Yes)	July–August (1080)	September–June (5148)	No adjustment	1
Walling and Veremakis, 2004, St. Louis, Missouri, USA (39)	2001	1 (TH)	Adult, admitted or transferred to medicine inpatient service (internal medicine)	Retrospective, pre–post (medical records)	No (No)	July (61 patients, 8 interns first month on service)	December (51 patients, 8 interns second–fourth months on service)	No adjustment	1
Yaghoubian et al, 2010, Los Angeles, USA (55)	1988–2007	2 (THs)	All patients (age ≥5 y), underwent appendectomy at 1 of 2 study hospitals (general surgery)	Retrospective, pre–post (NR)	Yes (NR)	July–Aug (766)	September–June combined (3559)	No adjustment	1

AAA = abdominal aortic aneurysm; ACS = acute coronary syndrome; ACS-NIQIP=American College of Surgeons-National Surgical Quality Improvement Program; AHA = American Hospital Association; APACHE = Acute Physiology and Chronic Health Evaluation; AY = academic year; CABG = coronary artery bypass grafting; CC = concurrent control; CHF = congestive heart failure; CL-C, CL-H, CL-I, CL-P = clustering of outcomes within clinician, hospital, intervention, and patient; CM = case-mix severity; CMS = Centers for Medicare & Medicaid Services; COTH = Council of Teaching Hospitals; CSF = cerebrospinal fluid; CVC = central venous catheter; D = demographic variables; DRG = diagnosis-related group; ED = emergency department; EFF = efficiency; ESIT = Endoscopic Shunt Insertion Trial; ICD-9 = International Classification of Diseases, Ninth Revision; ICU = intensive care unit; LS = level of supervision; MB = morbidity; ME = medical error; MI = myocardial infarction; MT = mortality; NA = not available; NHS = National Health Service, United Kingdom; NICU = neonatal intensive care unit; NIS = National Inpatient Sample; NR = not reported; NTH = nonteaching hospital; RCT = randomized, controlled trial; RI = degree of resident involvement in patient care; SDT = Shunt Design Trial; ST = seasonal (within) trends; T = secular (year-to-year trends); TH = teaching hospital; U = level of urgency; UK = United Kingdom; VA = Veterans Affairs; VA-CICSP = Veterans Affairs Continuous Improvement in Cardiac Surgery Program.

* NIS: nationally representative, hospital discharge database that includes approximately 20% of all inpatient admissions to nonfederal, acute care hospitals in the United States. Hospital designated as "teaching" if it has 1 or more residency program in any specialty; inpatient files from Centers for Medicare & Medicaid Services: includes all fee-for-service, acute care hospitalizations for Medicare patients hospitalized in the United States; VA-CICSP: prospectively collects risk and outcomes data on all patients who undergo cardiac surgery at any of the 44 Veterans Affairs cardiac surgery centers; COTH: membership does not necessarily mean presence of a residency program in the specialty studied.

† Quality ratings: 1 = poor, no adjustment; 2 = fair, adjusted only for demographic variables and case mix; 3 = good (criteria for 2 plus adjustment for year-to-year time trends or within-year time trends); 4 = very good (criteria for 3 plus inclusion of a concurrent control).

‡ Major THs were members of COTH; minor THs were not COTH members but had ≥1 residency program.

§ THs were COTH members.

|| Although this study did not control for time (year or seasonal) trends, the study used an adverse event detection method that had been validated in a prior study; analyzed patient-level data linked to individual clinicians; examined only procedures performed by trainees; and used sophisticated statistical adjustments for case mix, clustering of outcomes, and level of supervision. On the basis of these unique strengths, it was categorized as good (rather than fair) quality.

¶ Although this study did not control for demographic variables or case mix, it did control for time trends and used a concurrent control. On the strength of the latter, this study was categorized as good quality.

** THs were hospitals where interns provide the bulk of care.

Appendix Table 4. Results of Included Studies

Study, Year, Location (Reference)	Quality Rating*	Mortality	Morbidity	Efficiency	Other Outcomes and Comments
Alshekhlee et al, 2009, USA (28)	1	In-hospital mortality, unadjusted: No difference in trend in mortality by month for all years: 2000 ($P = 0.59$), 2001 ($P = 0.29$), 2002 ($P = 0.25$), 2003 ($P = 0.28$), 2004 ($P = 0.93$), 2005 ($P = 0.70$)			
Anderson et al, 2009, USA (40)	4	Hospital mortality, adjusted relative risk: 1.12 ($P < 0.05$)	Overall intraoperative complication rate, adjusted: P = NS (specific results not given) Perioperative complication rates, adjusted: P = NS (specific results not given)	Hospital LOS, adjustment: P = NS (specific outcomes not provided) Hospital charges, adjustment: P = NS (specific data not reported)	
Ayas et al, 2007, Canada (29)	2		Rate of pneumothorax within 48 h of CVC insertion: AOR, 1.24 (95% CI, 0.79–1.97; $P = 0.35$)		Data entered by trained ICU nurses with automated checks for errors. Type of CVC measured but not adjusted for.
Aylin and Majeed, 1994, UK (9)	1	Proportional mortality ratio (unadjusted): Ischemic heart disease: MI, 1.03 (CI, 1–1.06); other, 0.99 (CI, 0.95–1.03) Stroke: 1.00 (CI, 0.97–1.03) Asthma: 1.03 (CI, 0.81–1.31) All malignant neoplasms: 1.00 (CI, 0.98–1.02) Gastric and duodenal ulcers: 1.04 (CI, 0.9–1.21) Accidents: car, 1.04 (CI, 0.93–1.17); other, 1.00 (CI, 0.91–1.11) All other causes of death: 0.99 (CI, 0.97–1.01)			Primary statistical test was proportional mortality ratio. Did not have data on number of admissions, so mortality rates could not be calculated. Patient population includes elective and nonelective.
Bakaeen et al, 2009, USA (41)	MT: 2 MB: 2 EFF: 1	30-d operative mortality, AOR: 0.99 (CI, 0.89–1.11; $P = 0.90$)	Perioperative morbidity, AOR: 1.01 (CI, 0.96–1.07; $P = 0.67$)	Cross-clamp time, unadjusted: 84 vs. 83 min ($P = 0.009$)† Cardiopulmonary bypass time, unadjusted: 126 vs. 124 min ($P = 0.0412$)† Surgery duration, unadjusted: 4.9 vs. 4.8 h ($P < 0.0001$)† Total operating room time, unadjusted: 6.4 vs. 6.3 h ($P < 0.0001$)†	Secondary analysis stratified primary outcomes by isolated CABG vs. other procedures. Results similar. Of 44 study sites, 40 confirmed to have trainees involved in care of cardiac surgical patients.
Banco et al, 2002, Philadelphia, USA (42)	1		Surgical site infection rate, unadjusted: P = NS for July and August compared with every other month except January (which had higher rate, $P < 0.0227$)		Specific results not provided. Surveillance method itself not described.
Barry and Rosenthal, 2003, Ohio, USA (30)	4	Hospital mortality, AOR Q1 vs mean for all months: Major THs: 0.96 (CI, 0.91–1.02; $P = 0.18$) Minor THs: 1.07 (CI, 0.98–1.16; $P = 0.11$) NTHs: 0.96 (CI, 0.91–1.01; $P = 0.11$) AOR for Q1 similar to AOR for Q2, Q3, and Q4; AOR by month showed no difference		ICU LOS, adjusted: Major THs: 0.3 (CI, −0.7 to 1.2) Minor THs: 0.2 (CI, −0.9 to 1.4) NTHs: −0.8 (CI, −1.4 to −0.1)† Hospital LOS, adjusted: Major THs: −0.8 (CI, −2.7 to 1.0) Minor THs: 0.1 (CI, −2.1 to 2.3) NTHs: −1.1 (CI, −2.3 to 0.2) (Coefficients = adjusted percentages difference relative to overall mean for entire year)	Secondary analyses performed by month (rather than quarter) and for each of the 5 academic years, each hospital, and each ICU. No results showed a difference in mortality in Q1 or July compared with mean for the year.

Continued on following page

Appendix Table 4—Continued

Study, Year, Location (Reference)	Quality Rating*	Mortality	Morbidity	Efficiency	Other Outcomes and Comments
Borenstein et al, 2004, Toronto, Canada (43)	1		Adverse outcome rate, unadjusted (defined as unintended harm to patient resulting from medical treatment or as result of natural history of disease divided by total number of patient-days): 6.7% vs. 5.0% (P = 0.21)		Error rate, unadjusted (defined as number of errors that had the potential to cause substantial harm divided by total number of patient-days): All errors: 7.5% vs. 7.2% (P = 0.9); only errors attributable to residents: 3.0% vs. 3.7% (P = 0.5). Unclear method for identifying errors via record review.
Bruckner et al, 2008, California, USA (31)	4	Neonatal mortality (≤ 27 d of birth), AOR: Very-low-weight: 0.98 (CI, 0.78–1.23) Moderately low weight: 0.79 (CI, 0.55–1.14)			
Buchwald et al, 1989, Boston, USA (56)	MT: 1 EFF: 3	Hospital mortality, unadjusted: Medical DRGs: 3.5% vs. 2.3% (P = NS) Surgical DRGs: 1.5% vs. 2.1% (P = NS)		LOS, adjusted: All 20 DRGs: 9.6 vs. 9.5 d (P = NS) 10 medical DRGs: 7.6 vs. 7.6 d (P = NS) 10 surgical DRGs: 11.4 vs. 11.3 d (P = NS) Total charges, adjusted: All 20 DRGs: $9282 vs. $9008 ($P$ = NS) 10 medical DRGs: $5344 vs. $5285 ($P$ = NS) 10 surgical DRGs: $14 806 vs. $14 452 ($P$ = NS)	Specific P values not provided.
Chow et al, 2005, Hong Kong (32)	1				Near-miss error rate in laboratory requests, unadjusted (number of errors divided by 100 intern-days at risk): OR: 2.64 (CI, 1.29–5.38; P = 0.004).† Most common near-miss errors: mislabeled laboratory specimens and blood sample taken from wrong patient.
Claridge et al, 2001, Virginia, USA (44)	1	Hospital mortality, unadjusted: Overall: 3.8% vs. 5.6% (P = NS) Low-injury group: 1.0% vs. 2.8% (P = NS) Moderate-injury group: 1.8% vs. 0% (P = NS) Severe-injury group: 16.3% vs. 23.5% (P = NS)	Infection rate, unadjusted: 12.5% vs. 14.2% (P = NS) Units of pRBC, unadjusted: 1.7 vs. 1.5 (P = NS) Morbidity results stratified by severity of injury not provided	Overall, unadjusted: Hospital LOS: 22.0 vs. 19.3 d (P = NS) ICU LOS: 15.1 vs. 13.3 d (P = NS) ED time: 271 vs. 267 min (P = NS) Correction of acidosis: 7.4 vs. 9.1 h (P = NS) Total hospital costs (in thousands): $15.2 vs. $12.4 ($P$ = NS) P = NS for all outcomes when stratified into low injury, moderate injury, and severe injury	Specific P values not provided. Additional analysis stratified outcomes into low, moderate, and severe degree of injury (as defined by injury severity scale). Results similar.
Dhaliwal et al, 2008, Houston, USA (45)	2	30-d operative mortality, AOR: 0.28 (CI, 0.07–1.19; P = 0.09)	Perioperative morbidity, AOR: 0.83 (CI, 0.54–1.28; P = 0.41)	Cross-clamp time, unadjusted: 78 vs. 73 min (P = 0.084) Cardiopulmonary bypass time, unadjusted: 133 vs. 124 min (P = 0.031)† Surgery duration, unadjusted: 354 vs. 338 min (P = 0.016)† Total operating room time, unadjusted: 464 vs. 440 min (P = 0.0003)†	

Continued on following page

Appendix Table 4—**Continued**

Study, Year, Location (Reference)	Quality Rating*	Mortality	Morbidity	Efficiency	Other Outcomes and Comments
Englesbe et al, 2007, USA (47)	MT: 2 MB: 2 EFF: 1	30-d mortality rate, AOR: 1.41 (CI, 1.11–1.80; $P = 0.005$)	30-d morbidity rate, AOR: 1.18 (CI, 1.07–1.29; $P = 0.0005$)†	Operating room time, unadjusted: 2.36 vs. 2.22 h ($P < 0.001$)† Time to incision, unadjusted: 0.62 vs. 0.59 h ($P < 0.001$)† Time till out of operating room: 0.44 vs. 0.32 h ($P < 0.001$)†	Study period began in 2001, but 80% of cases in database were from 2003 and 2004. The 4 community hospitals have unclear teaching status and contribute a small proportion of the patients to this database. Heterogeneous mix of surgical procedures.
Englesbe et al, 2009, USA (46)	4	30-d operative mortality: AOR, THs and NTHs: CABG: 0.91 (CI, 0.82–1.02) and 0.95 (CI, 0.89–1.02) CEA: 0.83 (CI, 0.61–1.14) and 0.98 (CI, 0.86–1.13) AAA repair: 1.06 (CI, 0.86–1.30) and 1.03 (0.9–1.17) Esophagectomy: 1.15 (CI, 0.83–1.61) and 1.29 (CI, 0.94–1.75) Colectomy: 1.05 (CI, 0.85–1.28) and 1.00 (CI, 0.92–1.08) Pancreatectomy: 1.05 (CI, 0.72–1.54) and 1.00 (CI, 0.66–1.49) Hip fracture repair: 0.99 (CI, 0.89–1.09) and 0.93 (CI, 0.90–0.96)			No difference in mortality between periods when additional analysis stratified TH outcomes by elective vs. nonelective procedures, >300 beds, and resident:bed ratio (0, 0–0.1, 0.1–0.5, >0.5). No difference when July/August data were compared with May/June data.
Finkielman et al, 2004, Rochester, Minnesota, USA (33)	2	Hospital mortality rate, standardized mortality ratio: July: 0.95 (CI, 0.82–1.08) AOR July vs. individual non-July months: P = NS except for February, which had higher rate		Mean ICU LOS, adjusted: 10.67 d (July) vs. range of 10.73–11.70 d (non-July months); P = NS Observed/predicted ratio: 0.98 (July) vs. range of 0.96–1.03 (non-July months); P = NS	Secondary analyses stratified by ICU site showed similar results. Specific P values for LOS outcomes not provided.
Ford et al, 2007, USA (48)	1		P = NS for all 12 outcomes: cesarean delivery, urethral/bladder injury, third-degree laceration, fourth-degree laceration, perineal wound complication, postpartum hemorrhage, transfusion, shoulder dystocia, infection in amniotic cavity, anesthesia-related complication, brachial plexus injury, and birth asphyxia.		THs were designated per NIS and then verified independently to have residents working on the labor floor.
Garcia et al, 2009, Miami, USA (34)	MT: 1 MB: 1 EFF: 1 ME: 2	Hospital mortality, unadjusted (July–September vs. October–June): ACS: 1.0% vs. 1.4% ($P = 0.71$) CHF: 3.3% vs. 2.5% ($P = 0.44$)	ACS: Periprocedural nonfatal complication rate, unadjusted: 2.1% vs. 2.8% ($P = 0.60$)	LOS, unadjusted, for patients with CHF: 6.2 vs. 6.7 d ($P = 0.55$)	ACS: Discharge rate on optimal medical management, AOR: 0.93 (CI, 0.67–1.3; $P = 0.70$) CHF: Discharge rate on optimal medical management, AOR: 0.45 (CI, 0.25–0.81; $P = 0.007$)†, Entirely explained by assessment of LVF. LVF assessment counted only if it took place during admission. In other studies, if assessment occurred soon after discharge, it was considered within range of optimal.

Continued on following page

Appendix Table 4—Continued

Study, Year, Location (Reference)	Quality Rating*	Mortality	Morbidity	Efficiency	Other Outcomes and Comments
Haller et al, 2009, Melbourne, Australia (8)	3		Undesirable event rate, AOR: All training years: 1.40 (CI, 1.24–1.58; $P < 0.001$)† Training year 1: 1.31 (CI, 1.11–1.55)† Training year 2: 1.69 (CI, 0.94–3.04) Training year 3: 1.19 (CI, 0.53–2.71) Training year 4: 1.25 (CI, 0.92–1.68) Training year 5: 1.78 (CI, 1.39–2.29)† Outcomes were near-miss plus adverse events (i.e., events that could have reduced or did reduce the safety margin for the patient)		Only procedures performed by trainees were included in the study. Comprehensive adjustment for confounders, including type and duration of procedure, time of day, mode of supervision, experience level of resident, and clustering. Reporting system used in this study has 80% sensitivity and 91% specificity for intraoperative errors identified by peer reviewers in medical charts. System weights adverse events and near-misses equally. Degree of supervision measured at level of event/outcome: direct 1:1 in operating room, 1:≥2 in operating room, 1:≥2 outside operating room, and attending at home.
Hartley et al, 2007, Miami, USA (35)	1		Final visual acuity, unadjusted: 0.40 vs. 0.44 ($P = 0.72$)		Unscheduled returns visits to ED, unadjusted: 6% vs. 5% ($P = 0.82$) Agreement between initial and final diagnosis, unadjusted: 96% vs. 98% ($P = 0.70$) Analysis restricted to patient care delivered by interns. Urgency not measured or controlled for.
Highstead et al, 2009, Washington, DC, USA (49)	MT: 2 MB: 1 EFF: 1	Hospital mortality, unadjusted: 7% vs. 6% ($P = 0.11$); both study periods outperformed mortality predicted by TRISS model	Total complication rate, unadjusted (total complications divided by total number of patients): 12% vs. 13% ($P = 0.07$) Patient complication rate, unadjusted (number of patients who experienced ≥1 complication divided by total number of patients): 6% vs. 6% ($P = 0.80$)	ICU LOS, unadjusted: 5.6 vs. 5.3 d ($P = 0.96$) Hospital LOS, unadjusted: 4.6 vs. 4.5 d ($P = 0.92$)	
Huckman and Barro, 2005, USA (57)	4	Hospital mortality, adjusted, compared with NTHs: Major THs: $\beta = 0.122$ ($P < 0.01$); 4.3% relative increase in July mortality rate compared with NTHs Minor THs: $\beta = 0.014$ ($P =$ NS)		LOS, adjusted: Major THs: $\beta = 0.111$ ($P < 0.01$)†; 1.9% relative increase in LOS (mean, 5.80 d) compared with NTHs relative to reference period of April/May; LOS remains 1%–2% higher through December Minor THs: $\beta = 0.049$ ($P < 0.01$)†; LOS overall, 5.3 d; this represents a 0.9% increase compared with NTHs relative to the reference period of April/May	Full-time resident per inpatient hospital bed ratio used to differentiate NTHs (0), minor THs (<0.25), and major THs (≥0.25). Separate analyses for major and minor THs compared with NTHs. Relative mortality rate remained 2%–4% higher through December. Effect of changeover associated with 1500–2750 accelerated deaths per year in major THs in USA.

Continued on following page

Appendix Table 4—Continued

Study, Year, Location (Reference)	Quality Rating*	Mortality	Morbidity	Efficiency	Other Outcomes and Comments
Inaba et al, 2010, Los Angeles, USA (58)	2	Hospital mortality, AOR: 1.1 (CI, 0.8–1.5; *P* = 0.52)	Preventable and potentially preventable complication rate, AOR: 1.9 (CI, 1.1–3.2; *P* = 0.013)†	Hospital LOS, adjusted: −0.3 d (CI, −0.7 to 0.2 d; *P* = 0.239) ICU LOS, adjusted: −1.0 (CI, −2.1 to 0.1 d; *P* = 0.090)	Complications related to intravascular catheters had by far the largest difference between the 2 time periods (18% vs. 2%). Unclear whether result positive for other types of complications.
Jen et al, 2009, UK (59)	3	Hospital mortality, AOR: All patients: 1.06 (CI, 1.00–1.13; *P* = 0.05) Surgical patients: 1.01 (CI, 0.83–1.22; *P* = 0.86) Patients with neoplasm: 0.99 (CI, 0.85–1.16; *P* = 0.89) Medical patients: 1.08 (CI, 1.01–1.16; *P* = 0.03)†			Because only 2 contiguous wk were studied, controlling for seasonal variation is not relevant. Last wk of July may not be best comparison because effect of changeover may reach into July (e.g., terminal leave).
Kestle et al, 2006, Canada (27)	1		Shunt infection rate, unadjusted: 9% vs. 7.6% (*P* = 0.30) Shunt survival after initial insertion, unadjusted: *P* = NS (point estimate not given) Shunt survival for all shunt insertions, unadjusted: July/August worse (*P* = 0.008)† but effect size not given		
Kestle et al 2006, Canada, USA, and Europe (26)	1	Hospital mortality, unadjusted: 0.7 vs. 0.2 (*P* = 0.22); data combined from 2 trials	Data combined from 2 trials, unadjusted: Mean shunt survival duration: 1.7 vs. 2.4 y (*P* = 0.10) Shunt survival rate at 1 y: 0.55 vs. 0.63 (*P* = 0.10) Incidence shunt infection: 13.8% vs. 8.8% (*P* = 0.08) Wound dehiscence: 2.9% vs. 0.7% (*P* = 0.05) Ventricles cannulated on first attempt: 85% vs. 86% (*P* = 0.85) Incidence CSF leakage: 0 vs. 0.7 (*P* = 0.35) Other complications: 4.3 vs. 4.8 (*P* = 0.81)	SDT: Mean duration of surgery: 39.4 vs. 39.9 min (*P* = 0.81)	SDT: Technical errors: 10% vs. 6.3% (*P* = 0.55) ESIT: Correct placement of catheter: 68% vs. 60% (*P* = 0.22) 10 sites for SDT and 16 sites for ESIT. Both trials include interventions done by trainees and nontrainees and neither controlled for level of resident involvement. In ESIT, proportion of residents performing procedures similar in the 2 study periods. Homogeneous mix of procedures.
Myles, 2003, Chicago, USA (50)	1	Mortality rates, unadjusted, July vs. August–June: Stillbirth rate: *P* = NS Neonatal mortality rate: *P* = NS Perinatal mortality rate: *P* = NS Results similar for other time period comparisons	17 outcome measures For July vs. August–June: *P* = NS for 13 of the outcomes; positive findings for 4 measures		Effect size for mortality data not provided for July vs. August–June. Absence of detail on how delivery records reviewed (e.g., 2 abstractors? Blinded to study question?). Minimal details on study hospital. Multiple "changeover periods" tested with July-only data not presented fully.

Continued on following page

Appendix Table 4—Continued

Study, Year, Location (Reference)	Quality Rating*	Mortality	Morbidity	Efficiency	Other Outcomes and Comments
Phillips and Barker, 2010, USA (60)	4	Fatal medication error rate (defined as medication errors recorded as the primary cause of death) Observed: expected mortality ratio for July: Inpatients: 1.062 (CI, 1.023–1.100)† Outpatient/ED settings: 1.060 (CI, 1.025–1.095)† All patients: 1.061 (CI, 1.035–1.087)† Counties with THs: 1.10 (CI, 1.06–1.14)† Counties without THs: 1.03 (CI, 1.00–1.07) Other causes of death not different in July: adverse effects, surgical errors, all causes (specific results not given)			Did not adjust for diagnosis and case-mix severity. Controlled for "teaching status" by comparing mortality rates for counties with and without major THs.
Rich et al, 1990, St. Paul, Minnesota, USA (36)	3	Hospital deaths per admission, adjusted: P = NS (point estimate not provided; CI, −0.035 to 0.056)	Nursing home placement, per admission per month of year, adjusted rate: P = NS (point estimate not provided; CI, −0.047 to 0.021) Readmission within 30 d per admission per month of year, adjusted rate: P = NS (point estimate not provided; CI, −0.052 to 0.003)	LOS, adjusted: −0.036 d (CI, −0.006 to −0.66 d; $P < 0.05$)† (i.e., for each additional month of experience, LOS declined 0.036 d) Hospital charges, adjusted: −0.0094 (P = 0.02) (i.e., for each additional month of housestaff experience, hospital charges declined 0.94%)	Mortality and morbidity outcomes: CI provided, but not point estimates
Rich et al, 1993, St. Paul and Minneapolis, Minnesota, USA (61)	4	Hospital mortality, AOR: All patients: 0.99997 (P = 0.41) Medical patients: P = 0.675 (point estimate not provided) Surgical patients: P = 0.646 (point estimate not provided)	Rate of discharge to nursing home facility, adjusted: All patients: P = 0.3 Medical patients: P = 0.3 Surgical patients: P = 0.86 Operative complications rate, adjusted: Medical patients: P = 0.61 Surgical patients: P = 0.29	LOS, adjusted (percentage change in THs relative to NTHs over course of the year): All patients: 3.2% increase (P = 0.022)† Medical patients: P = NS (point estimate not provided) Surgical patients: 7.2% increase (P = 0.011)† Total charges, adjusted (percentage change in THs relative to NTHs over course of year): All patients: 3% increase (P = 0.059) Medical patients: P = 0.904 Surgical patients: 7.5% increase (P = 0.012)†	Point estimates and CIs not provided for mortality outcome stratified by medicine and surgery. THs defined as hospitals where "interns provide bulk of care." Teaching status (yes/no) controlled for in regression model (via interaction term). All sites in Minneapolis area/Minnesota. "Non-teaching" hospitals may have some graduate medical education programs. Secondary analysis used an alternative definition of teaching status: number of interns per patient at each hospital each year. Results similar.
Schroeppel et al, 2009, Memphis, Tennessee, USA (51)	MT: 2 EFF: 1	Hospital mortality, AOR: July vs. each other month: P = NS; point estimates and CIs not reported here Q1 vs. each other quarter: P = NS for each of the 3 comparisons		ICU LOS, unadjusted: P = NS for July compared with each other month, except February (3.2 vs. 3.9 d; $P < 0.05$) Ventilator-days, unadjusted: P = NS for July compared with each other month Time in the resuscitation room, unadjusted: P = NS for July compared with each other month, except October, February, and June, where $P < 0.05$	

Continued on following page

Appendix Table 4—Continued

Study, Year, Location (Reference)	Quality Rating*	Mortality	Morbidity	Efficiency	Other Outcomes and Comments
Shuhaiber et al, 2008, Cambridge, UK (52)	3	Hospital mortality, AOR: Surgeries during trainee change months vs. non-change months: CABG only: 1.08 (CI, 0.81–1.42; *P* = 0.61) Complex cardiac operations: 1.34 (CI, 1.29–1.37; *P* = 0.02)† Surgeries performed entirely by registrar, first vs. last months: 0.89 (CI, 0.45–1.78; *P* = 0.75)		Cases during change vs. nonchange months: CABG only: ICU LOS, adjusted: *P* = 0.90 Hospital LOS, adjusted: *P* = 0.13 Time in surgery for resident cases: 2.2 min (CI, 0.3–4.0 min; *P* = 0.02)† Complex cardiac operations: ICU LOS, adjusted: *P* = 0.51 Hospital LOS, adjusted: *P* = 0.77 Time in surgery for resident cases: 0.9 min (CI, −2.7 to 4.5 min; *P* = 0.63) Surgery performed by registrar in first vs. last month: ICU LOS, adjusted: CABG, *P* = 0.54; complex cardiac operations, *P* = 0.57 Hospital LOS, adjusted: CABG, *P* = 0.54; complex cardiac operations, *P* = 0.40 Time in surgery for resident cases: −1.5 min (CI, −13.9 to 10.9 min; *P* = 0.82)	For cases performed during trainee change months, study adjusted for whether resident performed the case and did separate analysis for CABG only and complex cardiac surgery. For cases performed entirely by registrar during first or last month, did 1 analysis for all types of surgeries combined because of limited sample size. No adjustment for level of urgency.
Shulkin, 1995, Philadelphia, USA (62)	1		Adverse event rate, unadjusted (number of events, weighted by severity, divided by total number of patient discharges): Hospital overall: *P* = NS (point estimates not given) Department level: *P* = NS for 24 individual departments except for neurosurgery: 7% vs. 3.9% (*P* = 0.004)		Point estimates and CI not given for primary outcomes. *P* value for primary outcome given only after removal of 2 documentation measures from outcome; author's justification for this decision was that it had limited impact on clinical quality. Weighting of adverse events derived from nonvalidated scale.
Smith et al, 2006, USA (53)	4	Hospital mortality, AOR: Craniotomy for tumor resection: 0.43 (CI, 0.14–1.32; *P* = 0.14) Shunt surgery: 0.96 (CI, 0.58–1.60; *P* = 0.88)	Craniotomy for tumor resection, AOR: Adverse discharge disposition: 1.03 (CI, 0.71–1.51; *P* = 0.87) Neurologic complications: 1.00 (CI, 0.63–1.59; *P* = 0.99) Operative hematoma: 0.80 (CI, 0.40–1.58; *P* = 0.52) RBC transfusion: 0.70 (CI, 0.41–1.19; *P* = 0.19) Shunt surgery, AOR: Adverse discharge disposition: 0.85 (CI, 0.66–1.11; *P* = 0.24) Neurologic complications: 1.27 (CI, 0.75–2.16; *P* = 0.37) Operative hematoma: 1.44 (CI, 0.77–2.71; *P* = 0.25) RBC transfusion: 0.81 (CI, 0.48–1.37; *P* = 0.43)		In secondary analysis, model included interaction term between TH status and study outcome. Teaching status designation in NIS indicates presence of ≥1 residency program in any specialty, but not necessarily pediatric neurosurgery. Study estimates that two thirds to four fifths of hospitals in study had neurosurgical residents.
Smith et al, 2002, Ann Arbor, Michigan, USA (37)	2			LOS, adjusted: April–June compared with July–September: −1.0 d (CI, −1.9 to −0.1 d; *P* = 0.05); average LOS 1 d shorter in April–June than in July–September	Study primarily designed to assess the effect of the end-of-month (not year-end) changeover in residents and attendings on LOS.

Continued on following page

Appendix Table 4—Continued

Study, Year, Location (Reference)	Quality Rating*	Mortality	Morbidity	Efficiency	Other Outcomes and Comments
Soltau et al, 2008, Birmingham, Alabama, USA (38)	1	Hospital mortality, unadjusted: Infants weighing 401–1500 g: $P = 0.49$ Infants weighing > 1500 g: P = NS	Infants weighing 401–1500 g: Bronchopulmonary dysplasia, unadjusted: $P = 0.98$ Necrotizing enterocolitis, unadjusted: $P = 0.63$ Severe intraventricular hemorrhage or cystic periventricular leukomalacia, unadjusted: $P = 0.20$		P values provided but not point estimates or CIs. Outcomes presented as run-chart.
Stout et al, 2008, Macon, Georgia, USA (54)	1	Hospital mortality, unadjusted: 4.72% vs. 5.23% ($P > 0.25$)			
Walling and Veremakis, 2004, St. Louis, Missouri, USA (39)	1				Ordering error rate (defined as requiring subsequent written/verbal clarification, being discontinued/reversed/canceled within 12 h, or having not ordered an intervention that was recommended by an attending physician): Percentage of orders with errors, unadjusted: 3.1% vs. 0.84% ($P < 0.001$)† Charts with errors, unadjusted: 34.8% vs. 15.7% ($P < 0.001$)† Did not account for clustering of errors within a given intern—study reported that 55% of the charts and 35% of the errors were associated with 1 intern.
Yaghoubian et al, 2010, Los Angeles, USA (55)	1		Wound infection rate, unadjusted: 4.8% vs. 4.3% ($P = 0.6$) Postoperative abscess drainage rate, unadjusted: 1.2% vs. 1.5% ($P = 0.6$)	Hospital LOS, unadjusted: 2.5 vs. 2.5 d ($P = 1.0$)	Secondary analysis compared July/August with April/May ($n = 691$). Wound infection rate, unadjusted: 4.8% vs. 2.5% ($P = 0.02$)†. Other results not different.

AAA = aortic abdominal aneurysm; ACS = acute coronary syndrome; AOR = adjusted odds ratio; CABG = coronary artery bypass grafting; CEA =carotid endarterectomy; CHF = congestive heart failure; CI = confidence interval; CSF = cerebrospinal fluid; CVC = central venous catheter; DRG = diagnosis-related group; ED = emergency department; EFF = efficiency; ESIT = Endoscopic Shunt Insertion Trial; ICU = intensive care unit; LOS = length of stay; LVF = left ventricular function; MB = morbidity; ME = medical error; MI = myocardial infarction; MT = mortality; NIS = National Inpatient Sample; NS = not significant; NTH = nonteaching hospital; OR = odds ratio; pRBC = packed red blood cells; RBC = red blood cell; RCT = randomized, controlled trial; SDT = Shunt Design Trial; TH = teaching hospital; TRISS = Trauma and Injury Severity Score; UK = United Kingdom.

* Quality ratings: 1 = poor, no adjustment; 2 = fair, adjusted only for demographic variables and case mix; 3 = good (criteria for 2 plus adjustment for year-to-year time trends and/or within-year time trends); 4 = very good (criteria for 3 plus inclusion of a concurrent control).

† Statistically significant finding.

A July Spike in Fatal Medication Errors: A Possible Effect of New Medical Residents

David P. Phillips, PhD[1] *and Gwendolyn E. C. Barker, BA*[2]

[1]Department of Sociology, University of California at San Diego, La Jolla, CA, USA; [2]School of Public Health, University of California at Los Angeles, Los Angeles, CA, USA.

BACKGROUND: Each July thousands begin medical residencies and acquire increased responsibility for patient care. Many have suggested that these new medical residents may produce errors and worsen patient outcomes—the so-called "July Effect;" however, we have found no U.S. evidence documenting this effect.

OBJECTIVE: Determine whether fatal medication errors spike in July.

DESIGN: We examined all U.S. death certificates, 1979–2006 (n=62,338,584), focusing on medication errors (n=244,388). We compared the observed number of deaths in July with the number expected, determined by least-squares regression techniques. We compared the July Effect inside versus outside medical institutions. We also compared the July Effect in counties with versus without teaching hospitals.

OUTCOME MEASURE: JR = Observed number of July deaths / Expected number of July deaths.

RESULTS: Inside medical institutions, in counties containing teaching hospitals, fatal medication errors spiked by 10% in July and in no other month [JR=1.10 (1.06–1.14)]. In contrast, there was no July spike in counties without teaching hospitals. The greater the concentration of teaching hospitals in a region, the greater the July spike (r=.80; P=.005). These findings held only for medication errors, not for other causes of death.

CONCLUSIONS: We found a significant July spike in fatal medication errors inside medical institutions. After assessing competing explanations, we concluded that the July mortality spike results at least partly from changes associated with the arrival of new medical residents.

KEY WORDS: medication error; mortality; July Effect; teaching hospitals; medical residents.
J Gen Intern Med 25(8):774–9
DOI: 10.1007/s11606-010-1356-3

INTRODUCTION

Inexperienced medical staff are often considered a possible source of medical errors.[1–6] One way to examine the relationship between inexperience and medical error is to study changes in the number of medical errors in July, when thousands begin medical residencies and fellowships.[1,7–11] This approach allows one to test the hypothesis that inexperienced residents are associated with increased medical errors[1,8,9,11–15]—the so-called "July Effect."

Previous attempts to detect the July Effect have mostly failed,[1,8–17] perhaps because these studies examined small,[8,10–13,15–17] non-geographically representative samples,[8–17] spanning a limited period,[11–16] although a study of anaesthesia trainees at one Australian hospital over a 5-year period did demonstrate an increase in the rate of undesirable events in February—the first month of their academic year.[1] In contrast, our study examines a large, nationwide mortality dataset spanning 28 years. Unlike many other studies,[18] we focus on *fatal* medication errors—an indicator of important medical mistakes. We use these errors to test the "New Resident Hypothesis"—the arrival of new medical residents in July is associated with increased fatal medication errors.

Received January 21, 2009
Revised September 18, 2009
Accepted March 24, 2010

METHODS

Primary Dataset

We examined all official U.S. computerized death certificates (n=62,338,584).[19] Our dataset begins with 1979, when hospital status (e.g., inpatient) was first recorded, and ends with 2006, the latest data year available.

We assumed that, inside medical settings, fatal medication errors are more likely to be influenced by inexperienced residents than by patients. In contrast, *outside* medical settings, we assumed that inexperienced residents play a relatively smaller role, while the patient plays a correspondingly larger role. For example, Phillips, Barker, and Eguchi [20] showed that a significantly larger fraction of medication errors outside medical institutions involved alcohol—indicating the reduced importance of medical residents and the increased importance of the patient.

Therefore, we focused on persons dying inside medical settings: inpatients, outpatients, and those dying in the emergency department. Outpatients are included in our analysis because, in our dataset, 'outpatient' officially refers to persons receiving medical care inside medical institutions, without being admitted to the hospital.[21] Because outpatient and ED settings are not distinguished in our dataset, we cannot analyze these settings separately. We compare persons dying inside medical institutions (inpatients, outpatients/ED) with persons dying before reaching medical institutions (those dead on arrival "DOA").

Geographic detail and exact date of death are unavailable after 2004; consequently, all analyses requiring this information omit later data. In some analyses, we examined both primary and secondary causes of death. For these analyses, our study period begins with 1983, when secondary causes of death were first coded on computerized certificates.

Definitions

We define fatal medication errors as deaths in which medication errors are recorded as the primary cause of death. All other causes of death analyzed are also defined according to the primary cause. Officially acknowledged medication errors (n=244,388) are coded E850-E858 in the International Classification of Diseases, 9th Revision (ICD9) [22] and X40-X44 in the 10th revision (ICD10).[23] Medication error involves "accidental overdose of drug, wrong drug given or taken in error, and drug taken inadvertently [and] accidents in the use of drugs and biologicals in medical and surgical procedures."[22,23] This category is equivalent to "fatal *preventable* adverse drug event" used elsewhere.[24]

The ICD category "medication errors" is distinct from the ICD category "adverse effects," which we also examined. Adverse effects signify cases where "correct drug [was] properly administered in therapeutic or prophylactic dosage, as the cause of adverse effect" [22,23] (E930-E949 (ICD9); Y40-59 (ICD10)). This category includes unexpected allergic reactions resulting from proper drug administration and is equivalent to "fatal *non-preventable* adverse drug events," used elsewhere.[24] In addition to these categories, we examined surgical errors (E870-E876 (ICD9); Y60-Y69 (ICD10)), external causes (e.g., accidents, homicides, and suicides), and all deaths combined.

In contrast to many other studies[2,25–27] we analyze: (1) only *preventable* adverse effects;[25] (2) only *medication* errors (rather than combining several types of *medical* errors like medicinal and surgical);[2,26] (3) only *fatal* medication errors;[27] (4) only those medication errors coded as the *primary* cause of death (rather than medication errors coded as primary, secondary, and/or tertiary).[2,26] In addition, we examine a nationwide dataset, whereas most other studies extrapolate to nationwide figures from small non-geographically representative samples.[2,26] For these reasons, the number of medication errors in our study differs from the number in other studies.

Secondary Datasets

Computerized death certificates do not record whether the patient died in a teaching hospital, but they do record the county of death (1979-2004). Starting in 1980, American Hospital Association (AHA) surveys [28–30] recorded hospital types in each county. We used these surveys to identify counties containing major teaching hospitals [9] near the beginning (1980), middle (1992), and end (2004) of our study period. For each county, we calculated the proportion of hospitals that are major teaching hospitals; we assumed that this proportion is a good indicator of the influence of teaching hospitals and of medical residents in a county. A related indicator, the proportion of *patients* treated in major teaching hospitals, cannot be accurately measured with AHA datasets.

In addition to computerized death certificates and AHA surveys, we examined monthly data from three other datasets:

(1) Hospital admissions, recorded by the National Hospital Discharge Survey (1979-1997);[31] monthly admissions were not coded after 1997.
(2) Visits to the ED, recorded by the National Hospital Ambulatory Medical Care Survey, Emergency Department (1992-2005).[32]
(3) Visits to outpatient departments, recorded by the National Hospital Ambulatory Medical Care Survey, Outpatient Department (1992-2005).[33]

The latter two datasets provide complete information from January through November but omit a varying number of days in December. Consequently, we did not analyze December data for these datasets.

Statistical Analysis

We used two procedures to estimate significance levels, depending on the dataset investigated. For the death certificate data, we used standard procedures.[34–45] These procedures cannot be easily employed for the other datasets examined because these datasets use very complex multi-stage cluster sampling techniques.[46] For these datasets, we estimated significance levels with bootstrap procedures.[46]

For each of the 28 years under analysis, we determined a least-squares regression equation [34] for the monthly data; this procedure allowed us to estimate the expected number of events in a given month of a given year. In this regression procedure, we used two independent variables: (1) number of days in the month (28–31), and (2) number of the month (1–12). We then summed the 28 expected values for a given month to determine the total expected value for that month during the entire 28-year study period.[34]

We generated a regression equation for each year separately, rather than a regression equation for all years combined, because the first procedure corrects for possible changes in the monthly distribution from one year to another. The second procedure, using combined data, generates nearly identical expected values for each of the 12 months. For example, there is a correlation of 0.999 between the expected number of monthly deaths generated by the two procedures. (All correlations reported in this paper are the standard Pearson correlations.)

When analyzing mortality from each cause separately, we used linear regression because there is a linear pattern in monthly deaths from each cause under study—for the 28 years combined, the quadratic regression coefficients were insignificant ($b2=-0.29$; $t=-0.14$).[34] Inspection of regression results for each year separately also reveals no significant departure from linearity.

Linear regressions were also appropriate for nearly all other analyses because the quadratic regression coefficients were insignificant: for inpatient mortality ($b2=-0.29$; $t=-0.14$), outpatient and ED mortality ($b2=-4.74$; $t=-2.01$), and DOA (b2=

-2.29; t=-1.63). However, for ED admissions and for mortality from all causes combined, cubic regressions were appropriate.

Two-tailed significance tests are customary but sometimes inappropriate.[47–49] For some of our analyses to be meaningful, one-tailed tests are required. For example:

(1) We examine the difference:

D1 = July Effect *inside* teaching hospital counties—July Effect *outside* teaching hospital counties

We expect D1 to be both statistically significant *and* to have a positive value, thus requiring a one-tailed test.

(2) We also examine the difference:

D2 = July Effect *inside* medical institutions—July Effect *outside* medical institutions

Here too, we expect the difference to be both statistically significant *and* to have a positive value, thus requiring a one-tailed test.

(3) We also examine the correlation between the July Effect in a region and the concentration of teaching hospitals in that region. Here, we expect the correlation to be both statistically significant *and* to have a positive value, thus requiring a one-tailed test.

Unless otherwise stated, all our significance tests are two-tailed.

Following official recommendations [35] and our earlier practice, [36–43] we calculated standard errors [44,45] and significance levels, even though we examined complete counts, not samples.

RESULTS

Figure 1 displays for each month the ratio:

$$R = \text{Observed number of deaths/Expected number of deaths}$$

for inpatient deaths from medication errors. When R exceeds 1.00, observed mortality exceeds the number expected. In July, observed mortality significantly exceeded the expected level [1.062 (1.023-1.100)]. In all other months, mortality levels did not deviate significantly from expected. Henceforth, we use “JR” to indicate the value of R for July.

Figure 1 reveals a July Effect for data aggregated for 28 years. The July Effect was also evident when each year was examined separately. For inpatient deaths from medication errors, JR exceeded 1.00 for 21 of the 28 years (P=0.006; one-tailed binomial test). During the study period, JR displayed no trend (b=0.0003; t=0.104; P=N.S.). In particular, JR did not decline after July 1, 2003, when resident hours were reduced.[27] In the three years before this reduction (2000-2002), the average JR was 1.03; in the three years after this reduction (2004-2006), the average JR was 1.05.

Figure 2 displays JR for medication errors occurring in three settings: inpatient, outpatient/ED, and DOA. As expected, JR was not elevated [0.998 (0.945-1.052)] for DOA but was elevated for inpatients [1.062 (1.023-1.100)] and for those dying in outpatient/ED settings [1.060 (1.025-1.095)]. Henceforth, we combine these “intra-institutional” deaths. As in Figure 1, mortality from intra-institutional medication errors spiked only in July. For medication errors, JR inside medical institutions [1.061 (1.035-1.087)] was significantly larger than JR for DOA [0.998 (0.945-1.052); P=0.02; one-tailed ratio of ratios Z-test].[45]

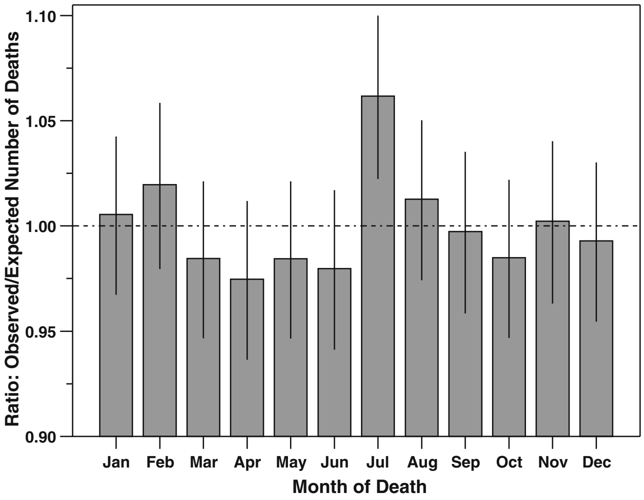


Figure 1. Ratio of observed to expected deaths for inpatient medication errors by month, United States, 1979-2006 (with 95% confidence intervals). Unless otherwise noted, error bars in Figure 1 and in subsequent figures were determined using a poisson approximation.[46]

The July spike for intra-institutional medication errors does not appear to have resulted from a rise in admissions to medical institutions, because inpatient admissions *decreased* in July [-3% (-5% to -2%)]; and neither increased nor decreased significantly for outpatient admissions [0% (-2% to 2%)] or ED admissions [1% (0% to 2%)].

Figure 3 compares JR for medication errors with JR for other causes of death. Except for medication errors, no cause of death displayed a significant July Effect inside medical institutions. In particular, JR was not elevated for adverse effects, i.e., for medication deaths not considered to result from

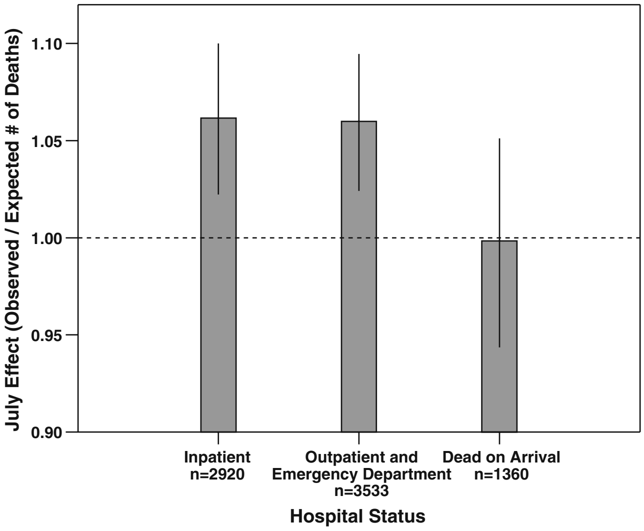


Figure 2. July effect for fatal medication errors by hospital setting, United States, 1979-2006 (with 95% confidence intervals). Error bars were calculated using Daly and Bourke.[44]

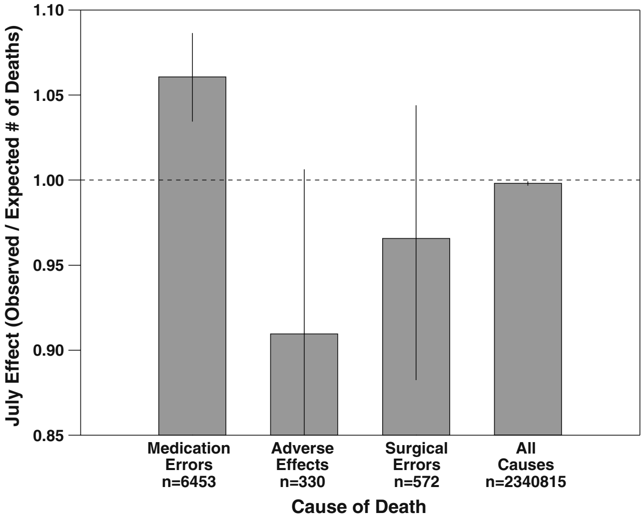


Figure 3. July effect for fatal medication errors and for comparison causes of death inside medical institutions, United States, 1979-2006 (with 95% confidence intervals).

error. Similarly, there was no July Effect for deaths inside medical institutions from all causes combined.

Given the "New Resident Hypothesis," JR should be largest in geographic regions with the largest concentrations of teaching hospitals. To test this prediction, we calculated the proportion:

Number of major teaching hospitals/Total number of hospitals

for each of the nine officially defined U.S. regions.[35] There was a strong regional correlation between this proportion and JR (r=0.80; t=3.54; n=9; P=0.005, one-tailed test). Thus, the greater the concentration of teaching hospitals in a region, the greater the July Effect for intra-institutional medication errors in that region.

In contrast, the comparison causes of death in Figure 3 did not display regional correlations of this sort (for intra-institutional mortality from adverse effects: r=-0.34; t=-0.95; P=N.S.; for surgical errors: r=-0.13; t=-0.36; P=N.S.; for all causes: r=0.46; t=1.36; P=N.S.).

Given the "New Resident Hypothesis," the July Effect should be concentrated in counties with teaching hospitals. To test this prediction, we examined counties with teaching hospitals near the beginning, middle, and end of the study period (Fig. 4). Henceforth, we term these "teaching hospital counties" and compare them with all other counties. As expected, for teaching hospital counties, JR was elevated (by 10%) for intra-institutional medication error deaths [JR=1.10 (1.06-1.14)]. In contrast, JR was not elevated for all remaining counties [JR= 1.03 (1.00-1.07)]. JR for teaching hospital counties was significantly larger than JR for all remaining counties (P= 0.03; one-tailed ratio of ratios Z-test).[45] The comparison causes of death displayed no July Effect either for teaching hospital counties or for other counties (Fig. 4).

In a further test of the "New Resident Hypothesis," we compared the following proportion for two groups:

Number of major teaching hospitals/Total number of hospitals

Group 1 consists of 102 counties for which the proportion of teaching hospitals increased over time. Group 2 consists of all remaining 2,324 counties. For Group 2 (the overwhelming majority of all counties) the annual JR (1979-2004) decreased over time (b= -0.009). In contrast, for Group 1, the annual JR (1979-2004) increased over time (b=0+.0008). Consistent with the "New Resident Hypothesis," the slope for Group 1 significantly exceeds the slope for Group 2 (t=2.12; d.f.=48; P= 0.02, one-tailed test).[34]

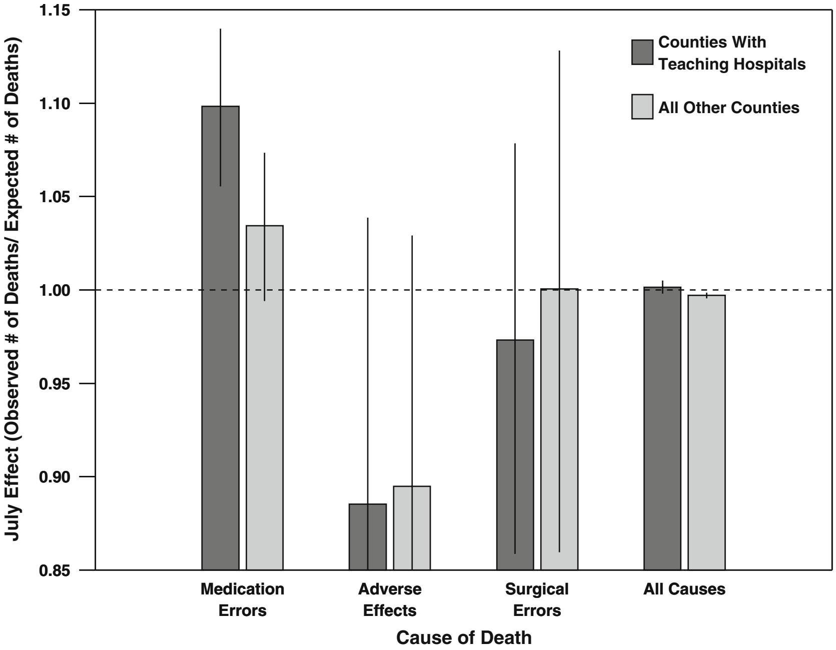


Figure 4. July effect by cause of death, for teaching hospital counties and for all other counties, United States, 1979-2004 (with 95% confidence intervals). The mortality dataset identifies only counties with at least 100,000 people; thus "all other counties" may include sparsely populated counties that contain teaching hospitals.

The above analyses examined medication errors coded as *primary* cause of death. If, in July, death registrars are unusually likely to code medication errors as a primary rather than as a secondary cause, then there should be a compensatory *drop* in medication errors coded as a secondary cause. However, July medication errors coded as a secondary cause decreased by only 9 [JR=0.99 (0.95 to 1.04)]. In contrast, these errors increased by 379 when coded as a primary cause.

If, in July, death registrars are unusually likely to ascribe a death to medication error rather than to adverse effects or suicides by medication, then these latter causes should *decrease* in July. However, neither adverse effects nor medication suicides decreased significantly in July. In July, adverse effects decreased by only 33 [JR=0.91 (0.81 to 1.01)]; medication suicides decreased by only 51 [JR=0.97 (0.93 to 1.02)].

DISCUSSION

Inside medical institutions, fatal medication errors spiked in July and in no other month. This July spike appeared only in counties containing teaching hospitals; in these counties, July mortality from medication errors was 10% above the expected level. These findings were evident only for medication errors and not for other causes of death or for deaths outside medical institutions.

Alternative Hypotheses

Although our findings are consistent with the "New Resident Hypothesis," other hypotheses are conceivable; these are assessed below.

1) *The July Effect may result from various behavioral changes occurring during the summer.* For example: A) a possible spike in summer alcohol consumption, combined with harmful alcohol-medication interactions; B) a summer spike in injuries from accidents and other "external causes," combined with increased medical efforts (e.g., prescriptions) to treat these injuries; C) an increase in summer tourism (tourists may receive worse health care). In addition, the July Effect may appear only in teaching hospital counties, because these counties might have an elevated proportion of summer tourists.

If the July Effect in fact resulted from these summertime behavioral changes, then there should be a general summertime increase in medication errors—not only in July but also in August. No such August spike is found.

2) *The July Effect may result from the July 4th holiday.* However, while July 4th is celebrated nationwide, the July Effect is evident only in teaching hospital counties. Moreover, medication errors do not spike in other months containing national holidays.
3) *The July Effect may result from coding changes in July.* Our findings above undermine this hypothesis (e.g., our analysis of adverse effects and medication suicides). The misclassification of some other cause of death might contribute to the July Effect, though we have seen no studies which show that in July there is a spike in the misclassification of any cause of death as medication error. Finally, it is difficult to understand how these putative types of misclassification could occur only in July and only in teaching hospital counties.

The analyses above suggest that, at present, the New Resident Hypothesis is the best available explanation for our findings.

Advantages and Limitations

Our use of official, computerized death certificates offers significant advantages: this dataset enabled us to examine a large, nationwide, multi-decadal sample and thereby detect a statistically significant July spike not found in earlier studies. However, our dataset is limited to the most severe type of medication errors (those resulting in death) and provides little detail about each medication error.

In part, because of these limitations, several questions remain: Is there a July Effect for *non*-fatal medication errors? What are the detailed mechanisms contributing to the July Effect (e.g., miscommunication, inadequate oversight)? Why is there a July spike in fatal medication errors but not in fatal surgical errors? These important questions require further study, perhaps with different kinds of datasets that provide more detail per case.

Implications

Despite these gaps in research, our findings have several implications for medical policy—they provide fresh evidence for: 1) re-evaluating responsibilities assigned to new residents; 2) increasing supervision of new residents; 3) increasing education concerned with medication safety. Incorporating these changes might reduce both fatal and non-fatal medication errors and thereby reduce the substantial costs[1] associated with medication errors.

CONCLUSION

Our nationwide, multi-decadal study enabled us to discover previously unknown evidence for a July spike in fatal medication errors. This spike seems to result at least partly from changes associated with new medical residents. The July Effect seems to be a significant public health problem and warrants further investigation.

Acknowledgments: *Funding: Marian E. Smith Foundation. We thank Kimberly M. Brewer, Robert D. Kleinberg, and Miranda Phillips for valuable comments and suggestions.*

Conflict of Interest: *None disclosed.*

Corresponding Author: *David P. Phillips, PhD; Department of Sociology, University of California at San Diego, 0533, 9500 Gilman Drive, La Jolla, CA 92093-0533, USA (e-mail: dphillips@ucsd.edu).*

REFERENCES

1. **Haller G, Myles PS, Taffe P, et al.** Rate of undesirable events at beginning of academic year: retrospective cohort study. BMJ. 2009;339: b3974.
2. **Kohn LT, ed, Corrigan J, ed, Donaldson MS, ed.** To Err Is Human: Building a Safer Health System. Washington, DC: National Academy Press; 2000.
3. **Altman DE, Clancy C, Blendon RJ.** Improving patient safety—five years after the IOM report. N Engl J Med. 2004;351:2041–3.
4. **Phillips J, Beam S, Brinker A, et al.** Retrospective analysis of mortalities associated with medication errors. Am J Health Syst Pharm. 2001;58:1835–41.
5. **Weingart SN, Wilson RM, Gibberd RW, Harrison B.** Epidemiology of medical error. BMJ. 2000;320:774–7.
6. **Lesar TS, Briceland LL, Delcoure K, Parmalee JC, Masta-Gornic V, Pohl H.** Medication prescribing errors in a teaching hospital. JAMA. 1990;263:2329–34.
7. U.S. medical school seniors enjoy most successful "match day" in 30 Years: 2008 residency program match is the largest in history. Washington, D.C.: Association of American Medical Colleges; 2008. http://www.aamc.org/newsroom/pressrel/2008/080320.htm. Accessed March 6, 2010.
8. **Finkielman JD, Morales IJ, Peters SG, et al.** Mortality rate and length of stay of patients admitted to the intensive care unit in July. Crit Care Med. 2004;32:1161–5.
9. **Barry WA, Rosenthal GE.** The effect of July admission on intensive care mortality and length of stay in teaching hospitals. J Gen Intern Med. 2003;18:639–45.
10. **Ford AA, Bateman BT, Simpson LL, Ratan RB.** Nationwide data confirms absence of 'July phenomenon' in obstetrics: it's safe to deliver in July. J Perinatol. 2007;27:73–6.
11. **Englesbe MJ, Pelletier SJ, Magee JC, et al.** Seasonal variation in surgical outcomes as measured by the American College of Surgeons-National Surgical Quality Improvement Program (ACS-NSQIP). Ann Surg. 2007;246:456–62.
12. **Banco SP, Vaccaro AR, Blam O, et al.** Spine infections: variations in incidence during the academic year. Spine. 2002;27:962–5.
13. **Rich EC, Gifford G, Luxenberg M, Dowd B.** The relationship of house staff experience to the cost and quality of inpatient care. JAMA. 1990;263:953–7.
14. **Rich EC, Hillson SD, Dowd B, Morris N.** Specialty differences in the 'July Phenomenon' for Twin Cities teaching hospitals. Med Care. 1993;31:73–83.
15. **Shulkin DJ.** The July phenomenon revisited: are hospital complications associated with new house staff? Am J Med Qual. 1995;10:14–7.
16. **Buchwald D, Komaroff AL, Cook EF, Epstein AM.** Indirect costs for medical education. Is there a July phenomenon? Arch Intern Med. 1989;149:765–8.
17. **Claridge JA, Schulman AM, Sawyer RG, Ghezel-Ayagh A, Young JS.** The 'July phenomenon' and the care of the severely injured patient: fact or fiction? Surgery. 2001;130:346–53.
18. **Barker KN, Flynn EA, Pepper GA, Bates DW, Mikeal RL.** Medication errors observed in 36 health care facilities. Arch Intern Med. 2002;162:1897–903.
19. Mortality detail file, 1979-2004. Hyattsville, Md.: National Center for Health Statistics (Computer Data File).
20. **Phillips DP, Barker GEC, Eguchi MM.** A steep increase in domestic fatal medication errors with use of alcohol and/or street. Arch Intern Med. 2008;168:1561–6.
21. Documentation of the Mortality Tape File for 1992 Data. Hyattsville, MD: National Center for Health Statistics; 1992.
22. ICD9.chrisendres.com Free Online Searchable 2009 ICD-9-cm. http://icd9cm.chrisendres.com/. Accessed March 6, 2010.
23. International Statistical Classification of Disease-Related Health Problems, 10th revision. Geneva, Switzerland: World Health Organization; 2006. http://www.who.int/classifications/icd/en/. Accessed March 6, 2010.
24. **Gurwitz JH, Field TS, Harrold LR, et al.** Incidence and preventability of adverse drug events among older persons in the ambulatory setting. JAMA. 2003;289:1107–16.
25. **Lazarou J, Pomeranz BH, Corey PN.** Incidence of adverse drug reactions in hospitalized patients: a meta-analysis of prospective studies. JAMA. 1998;279:1200–5.
26. **Sox HC Jr, Woloshin S.** How many deaths are due to medical error? Getting the number right. Eff Clin Pract. 2000;3:277–83.
27. **Yoon HH.** Adapting to duty-hour limits—four years on. N Engl J Med. 2007;356:2668–70.
28. American Hospital Association. AHA Annual Survey of Hospitals: 1980. Chicago: American Hospital Association; 1982.
29. American Hospital Association. AHA Annual Survey of Hospitals: 1992. Chicago: American Hospital Association; 1994.
30. American Hospital Association. AHA Annual Survey of Hospitals: 2004. Chicago: American Hospital Association; 2006.
31. National Center for Health Statistics. National Hospital Discharge Survey: Multi-Year Data File, 1979-97. Hyattsville, Md: National Center for Health Statistics; 1999. http://www.cdc.gov/nchs/nhds.htm. Accessed March 6, 2010.
32. National Center for Health Statistics. National Hospital Ambulatory Medical Care Survey: 1992-2005, Emergency Department Computer Data File. Hyattsville, Md: Centers for Disease Control and Prevention: 2007. http://www.cdc.gov/nchs/ahcd.htm. Accessed March 6, 2010.
33. National Center for Health Statistics. National Hospital Ambulatory Medical Care Survey: 1992-2005, Outpatient Department Computer Data File. Hyattsville, Md: Center for Disease Control and Prevention; 2007. http://www.cdc.gov/nchs/ahcd.htm. Accessed March 6, 2010.
34. **Kleinbaum DG, Kupper LL, Muller KE, Nizam A.** Applied Regression Analysis and Other Multivariable Methods. 2nd ed. Belmont: Wadsworth Publishing Co; 1988.
35. National Center for Health Statistics. Vital Statistics of the United States, Yearly Volumes: Volume 2, Mortality, Part A. Section 7. Washington, D.C.: Government Printing Office.
36. **Phillips DP, Christenfeld N, Ryan NM.** An increase in the number of deaths in the United States in the first week of the month—an association with substance abuse and other causes of death. N Engl J Med. 1999;341:93–8.
37. **Phillips DP, Paight DJ.** The impact of televised movies about suicide: a replicative study. N Engl J Med. 1987;317:809–11.
38. **Phillips DP, Carstensen LL.** Clustering of teenage suicides after television news stories about suicide. N Engl J Med. 1986;315:685–9.
39. **Phillips DP, Bredder CC.** Morbidity and mortality from medical errors: an increasingly serious public health problem. Annu Rev Public Health. 2002;23:135–50.
40. **Phillips DP, Christenfeld N, Glynn LM.** Increase in U.S. medication-error deaths between 1983 and 1993. Lancet. 1998;351:643–4.
41. **Phillips DP, Jarvinen JR, Phillips RR.** A spike in fatal medication errors at the beginning of each month. Pharmacotherapy. 2005;25:1–9.
42. **Phillips DP, Ruth TE, Wagner LM.** Psychology and survival. Lancet. 1993;342:1142–5.
43. **Phillips DP, Liu GC, Kwok K, Jarvinen JR, Zhang W, Abramson IS.** The Hound of the Baskervilles effect: natural experiment on the influence of psychological stress on timing of death. BMJ. 2001;323:1443–6.
44. **Daly LE, Bourke GJ.** Interpretation and Uses of Medical Statistics. 5th ed. Oxford: Blackwell Science Ltd; 2000:544.
45. **Altman DG, Machin D, Bryant TN, Gardner MJ.** Statistics With Confidence. 2nd ed. Bristol, England: BMJ Books; 2000:46–7.
46. **Chernick MR.** Bootstrap Methods: A Guide for Practitioners and Researchers. 2nd ed. Hoboken, NJ: Wiley-Interscience; 2008.
47. **Daly LE, Bourke GJ.** Interpretation and Uses of Medical Statistics. 5th ed. Oxford: Blackwell Science Ltd.; 2000:123–5.
48. **Altman DG, Machin D, Bryant TN, Gardner MJ.** Statistics With Confidence. 1st ed. Bristol, England: BMJ Books; 1989:60–1.
49. **Bland M.** An Introduction to Medical Statistics. 2nd ed. Oxford: Oxford University Press; 1995:138.