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

Abstract

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

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


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 methodologic 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 safety, or whether particular models are more or less problematic.

Primary Funding Source: Dr. Auerbach received support from the National Heart, Lung, and Blood Institute (K24HL098372).

Editors’ Notes

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 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

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 comparably 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 experience 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 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 January 1, 1989, and July 1, 2010. With the assistance of a reference librarian, we developed a search strategy that used combinations of keywords and Medical Subject Headings 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). See Appendix Table 1 for 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 with 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 time 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 operation 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 $\kappa =$}
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 summarizes the overall characteristics of included studies, and Appendix Table 3 provides detailed information about each individual study. Included studies were generally recent (published since 2000) and set in large, United States–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).

Appendix Table 2. Characteristics of 39 Included Studies*
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 the 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%) 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).

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).

Table. Included Studies Stratified by Outcome and Quality Rating

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. 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 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), 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 appeared to increase with study quality. Most (5 of 6 [83%]) studies showing an association were considered of higher methodologic 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 to the last week of the academic year in National Health Service hospitals in the United Kingdom, Jen and colleagues showed increased mortality in medicine patients but not surgical, neoplasm, or all patients (although, for the latter, the adjusted odds ratio was 1.06; \( P = .05 \)) (59). Likewise, Shuhaiber and colleagues found increased mortality during changeover months for complex cardiac surgeries but not simple coronary artery bypass grafting (52).

Morbidity and Medical Error Outcomes

Twenty-three studies reported morbidity outcomes, such as intraoperative complications (40), anesthesia-related undesirable 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). Only 4 of 23 [17%] studies reported an increase in morbidity; 1 of these was higher quality (8, 27, 47, 58).

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 no important statistical controls (for example, clustering analysis when more than one third of the errors related to 1 clinician) (39). Half 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). 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, methodologic limitations and study heterogeneity make deeper conclusions more difficult to draw. Studies focused on morbidity and medical error typically did not use validated surveillance systems and therefore have 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, auto regressive 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). Studies’ 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 one 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 produced 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 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 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 diagnosis, or clinical decompensation (that is, 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 more likely 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 Headings 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 does frame many reasonable approaches to reducing potential harms. 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.

Developing changeover systems that are informed by human factors 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).

**Article and Author Information**

**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:** 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.

References


15. Lesar TS, Briceland L, Stein DS. Factors related to errors in medication prescribing. JAMA. 1997;277:312-7. [PMID: 9002494]


Related Article

Editorial:
The July Effect: Fertile Ground for Systems Improvement
Paul Barach and Ingrid Philibert

Summary for Patients

Summaries for Patients:
Does the Annual Medical–Trainee Changeover at Teaching Hospitals Harm Patients?