Impact of reduction in working hours for doctors in training on postgraduate medical education and patients’ outcomes: systematic review

S R Moonesinghe, consultant and honorary senior lecturer in anaesthesia,1 director,2 J Lowery, specialty trainee, anaesthetics,3 N Shahi, specialty trainee, surgery,4 A Millen, specialty trainee, surgery,5 J D Beard, honorary professor of surgical education and consultant in vascular surgery6

ABSTRACT

Objectives To determine whether a reduction in working hours of doctors in postgraduate medical training has had an effect on objective measures of medical education and clinical outcome.

Design Systematic review.

Data sources Medline, Embase, ISI Web of Science, Google Scholar, ERIC, and SIGLE were searched without language restriction for articles published between 1990 and December 2010. Reference lists and citations of selected articles.

Study selection Studies that assessed the impact of a change in duty hours using any objective measure of outcome related to postgraduate medical training, patient safety, or clinical outcome. Any study design was eligible for inclusion.

Results 72 studies were eligible for inclusion: 38 reporting training outcomes, 31 reporting outcomes in patients, and three reporting both. A reduction in working hours from greater than 80 hours a week to less than 56 or 48 a week has resulted from the New Deal negotiated by the British Medical Association7 and the European Working Time Directive (93/104/EC).8

While the aim of these changes is to improve working conditions and safety, the medical profession has raised concern about the potentially adverse effects on postgraduate training for junior doctors and the provision of high quality care for patients. These concerns are particularly relevant in the European Union but might also be important in the US as the working hours of interns and residents continue to be reviewed.9

When changes to doctors’ working hours are considered and implemented there is a need for an evidence based approach to evaluating their impact on both educational and clinical outcomes. We carried out a systematic review to determine the impact of a reduction in working hours of doctors in postgraduate medical training on objective measures of educational and clinical outcome.

METHODS

We adhered to MOOSE (Meta-Analysis of Observational Studies in Epidemiology) guidelines2 and previously published recommendations for systematic reviews of observational studies3 in the conduct of our review.

Data sources We searched Medline, Embase, Google Scholar, the Educational Resources Information Centre (ERIC), and the System of Information on Grey Literature in Europe (SIGLE). The search was limited to articles published between 1 January 1990 and 20 December 2010.

INTRODUCTION

There has been a progressive reduction in the working hours of doctors in training in Europe and North America over the past 20 years. Legislation reducing junior doctors’ working hours (known as Code 405) was implemented by the New York State Department of Health in 19891 and limited doctors in that state to working an average of 80 hours a week. This was followed by national guidelines recommended by the US Accreditation Council for Graduate Medical Education (ACGME) in 2003.2 Further refinements of these limits were recommended by the US Institute of Medicine in December 2008.3 In the United Kingdom, a progressive reduction in junior doctors’ working hours has resulted from the New Deal negotiated by the British Medical Association7 and the European Working Time Directive (93/104/EC).8

ABSTRACT

Objectives To determine whether a reduction in working hours of doctors in postgraduate medical training has had an effect on objective measures of medical education and clinical outcome.

Design Systematic review.

Data sources Medline, Embase, ISI Web of Science, Google Scholar, ERIC, and SIGLE were searched without language restriction for articles published between 1990 and December 2010. Reference lists and citations of selected articles.

Study selection Studies that assessed the impact of a change in duty hours using any objective measure of outcome related to postgraduate medical training, patient safety, or clinical outcome. Any study design was eligible for inclusion.

Results 72 studies were eligible for inclusion: 38 reporting training outcomes, 31 reporting outcomes in patients, and three reporting both. A reduction in working hours from greater than 80 hours a week to less than 56 or 48 a week has resulted from the New Deal negotiated by the British Medical Association7 and the European Working Time Directive (93/104/EC).8

While the aim of these changes is to improve working conditions and safety, the medical profession has raised concern about the potentially adverse effects on postgraduate training for junior doctors and the provision of high quality care for patients. These concerns are particularly relevant in the European Union but might also be important in the US as the working hours of interns and residents continue to be reviewed.9

When changes to doctors’ working hours are considered and implemented there is a need for an evidence based approach to evaluating their impact on both educational and clinical outcomes. We carried out a systematic review to determine the impact of a reduction in working hours of doctors in postgraduate medical training on objective measures of educational and clinical outcome.

METHODS

We adhered to MOOSE (Meta-Analysis of Observational Studies in Epidemiology) guidelines2 and previously published recommendations for systematic reviews of observational studies3 in the conduct of our review.

Data sources

We searched Medline, Embase, Google Scholar, the Educational Resources Information Centre (ERIC), and the System of Information on Grey Literature in Europe (SIGLE). The search was limited to articles published between 1 January 1990 and 20 December 2010.
2010, with no language restriction. We identified 49,084 articles, which we then screened for inclusion.

**Search strategy**

The Medline search was conducted by exploding the following MeSH terms: “Medical Staff, Hospital” or “Personnel Staffing and Scheduling” or “Workload” or “Time Factors” and combining with the following terms, which we also exploded: “Education, Medical, Continuing” or “Education, Medical, Graduate” or “Specialization” or “Education, Medical” or “Internship and Residency”. We searched the following keywords separately: “medical training” and “European Working Time Directive”. The search found 16,132 articles. After snowballing, we included 39 in the final analysis.

In Embase we searched the following terms: MeSH terms “working time” or “personnel management” or “work schedule” or the keyword “European working time directive” and combined the following terms: MeSH terms: “medical education” or “patient care” or “medical education” or “patient care” or postgraduate education or “surgical training” or “training” or “staff training” or “residency education” or the keyword “postgraduate education”. This search found 8,556 articles. After removing duplicates from the Medline search and snowballing, we included seven in the final analysis.

We searched Google Scholar using the following exact phrases on advanced searches limited only to medicine, pharmacology, and veterinary sciences: “working hours” (11,070 articles), “working time” (5,020); “duty hours” (1,530) and “duty hour” (535). After removing duplicates from the Medline and Embase searches and snowballing, we included 27 studies in the final analysis.

We searched the System for Information on Grey Literature in Europe (SIGLE) combining the key- words: “European Working Time Directive” and “medical education” (4,831); “European Working Time Directive” and “patient” (479); “European Working Time Directive” and “medical” (807); “European Working Time Directive” and “surgery” (93).

No eligible studies were identified.

We searched the Educational Resources Information Centre (ERIC) using the following keywords in combination: “medical education” and “working hours” or “duty hours”. We found 31 articles; none was eligible for inclusion.

**Study selection**

We selected studies that assessed the impact of a change in duty hours, with details reported of what change had been implemented, and used an objective measure of outcome related to postgraduate medical training, patient safety, or clinical outcome.

We excluded studies reporting subjective measures, such as surveys or questionnaires, unless the results included an objective externally validated measure, such as case numbers or results of assessments. We also excluded studies that assessed the effect of changes in duty hours on medical staff (for example, measures of fatigue, physical or psychological wellbeing) as opposed to patients. Any study design was eligible for inclusion.

Of the 49,084 citations screened, we identified 157 articles to review in detail. We “snowballed” these articles by examining reference lists and searching for citations on Medline, Embase, and ISI Web of Science; this process identified a further 68 articles. Two authors (from SRM, JL, NS, and AM) independently reviewed these 225 papers, and 72 met inclusion criteria: 38 of postgraduate medical training, 31 of outcomes related to patients, and three that reported both (figure).

**Data extraction**

Two reviewers (from SRM, JL, NS, and AM) independently extracted information from each article using standardised data extraction forms, and one author (SRM) reviewed all studies. Using two separate forms (one for studies of educational outcomes, one for studies of patients’ outcomes), we extracted data on study authors, geographical location, year of publication, study cohort characteristics, working hours and pattern before and after the intervention, outcome measures, and main results from all studies.

We also assessed the quality of the studies, again by extracting data onto separate data extraction forms. For studies of training outcomes, we extracted number of participants, study design, institutional setting, source, and method of data collection; whether the study was single or multicentre; whether a description of overall institutional activity was included in studies that used case volume as an outcome measure; and whether statistical analysis was reported. The quality of reporting of cohort characteristics in studies of postgraduate training outcomes was assessed as being “good” if the study reported both the number of hours a week worked and described shift patterns, “moderate” if the study reported only one of these, and “poor” if the study reported neither.

For studies of patients’ outcomes, we extracted data on number of participants, study design, institutional setting, source and method of data collection, whether the study was single or multicentre, whether the study included a control group, whether outcomes were risk adjusted or whether the study presented a comparison...
of patients’ characteristics at baseline, and whether statistical analysis was reported.

RESULTS

Tables 1-3 summarise the key findings from studies evaluating medical training, for improvement in outcomes, no change or variable outcomes, and deterioration in outcomes, respectively. Tables 4-6 summarise the key findings from studies assessing outcomes related to patients, for improvement in outcomes, no change or variable outcomes, and deterioration in outcomes, respectively. Though we did not restrict our search to English language publications, all the studies we identified originated from either the UK or US. Table 7 documents the changes to duty hours recommended in the UK and US. All studies from the US reported the impact of a change from more than 80 duty hours a week to fewer than 80, in accordance with Code 405 legislation or ACGME recommendations. The UK based studies reported the impact of various changes in total duty hours and shift patterns, in accordance with UK legislation and the European Working Time Directive. The standardised data extraction tables described in the methods section are shown in the appendix on bmj.com.

Studies of postgraduate medical training

We analysed 41 studies of postgraduate medical training. All were “before and after” cohort studies. Twenty seven originated from the US, and 14 were conducted in the UK. Twenty eight studied training in surgery or surgical subspecialties (22 in US, six in UK), five studied training in obstetrics or gynaecology, or both (four in US, one in UK), six studied training in anaesthesia (all from the US), one US study was in paediatrics, and one UK study was of medical trainees. Of these 41 studies, two showed an improvement in training outcomes after a reduction in working hours, 12 reported a deterioration, and the 27 remaining showed no change or a combination of positive and negative results. Most studies used multiple outcome measures to evaluate the impact of a reduction in working hours on postgraduate education and training.

Postgraduate training results according to effect of reduced hours

**Improved training outcomes**

Two papers reported an improvement in training outcomes after a reduction in working hours: one UK study of medical trainees and one US study of surgical residents. Both were of low methodological quality as they did not report statistical analyses of the results. Roedling et al reported the change in shift pattern, and Schneider et al reported percentage compliance with duty hours changes, but neither reported the actual number of hours worked by trainees.

**Deterioration in training outcomes**

Twelve studies found a detrimental association between reduced working hours and measures of postgraduate training outcome: six from the UK (three in surgery, two in anaesthesia, and one in gynaecology) and six from the US (all in surgery). Eleven of these reported operative caseload as an outcome measure; one US study used continuity of care by neurosurgical trainees as an outcome. Of the 11 studies that used operative caseload as an outcome measure, three single centre studies did not report institutional operative volumes and therefore we are unable to determine if the reduction in trainees’ caseload could have occurred as a result of a change in institutional activity.

**No change in training outcomes**

Of the 27 studies that found no change or mixed results of training outcomes associated with reducing working hours, 20 originated from the US (four in obstetrics and gynaecology, 16 in surgery, and seven from the UK (four in anaesthesia, three in surgery)). Of the 20 studies from the US, 12 used operative caseload as an outcome measure, two analysed postgraduate examination results, and six used both. Four US papers studying surgical training reported results from large multicentre or multiprogramme cohorts. The 16 remaining papers from the US were studies in single institutions or residency programmes. Only four reported both the actual number of hours worked by trainees before and after rota changes and the shift patterns worked.

**Postgraduate training results according to outcome measures analysed**

**Training opportunities**

Six studies used measures of training opportunities (such as supervised operating lists or teaching sessions). Of these, one UK study of internal medicine physicians found an improvement in attendance at training sessions, and one UK study of anaesthetists found a deterioration in the number of training opportunities in obstetric anaesthesia. The remaining four studies were all conducted in UK anaesthetics departments and found no change in supervision levels overall after various changes in duty schedules.

**Examination scores**

Nine studies, all originating from the US, measured changes in examination scores for cohorts of trainees before and after duty hour reforms. Two found an improvement in scores, and seven found no difference.
Thirty seven studies included caseload as an outcome measure. Only one study, from the US, showed an increase in caseload after a reduction in working hours.\textsuperscript{10} Eleven studies showed a reduction in operative caseload after reform of duty hours: six US and two UK studies of surgical trainees,\textsuperscript{12,13,48} one UK study of gynaecology trainees,\textsuperscript{16} and two UK studies in anaesthesia.\textsuperscript{14,15} Twenty five studies found no change in operative caseload, whether in surgery (17 studies: 14 in US, three in UK), obstetrics and gynaecology (four studies, all in US), or anaesthesia (four studies, all in UK) after reductions in working hours.

Schneider et al’s 2007 study in US surgical residents was the only paper to report an increase in caseload after a reduction in working hours, and it also found an improvement in postgraduate examination scores.\textsuperscript{10} This was a single centre study of low methodological quality, however, as there was no statistical analysis conducted on the results.

Of the 11 papers that reported a reduction in caseload, eight were in surgery, two in anaesthetics, and one in obstetrics.
Table 3: Studies on effect of reduced working hours that predominantly found deterioration in postgraduate education

<table>
<thead>
<tr>
<th>Study</th>
<th>Details</th>
<th>Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carlin 2007, US17</td>
<td>Surgical residents, single centre, 4 year</td>
<td>Operative caseload—PGY 1, 2, and 4: reduced. Total and chief surgeon experience—PGY 1, 2, and 4: reduced; PGY 3 and 5: NS. Total and chief surgeon experience: overall (PGY 1-5) reduced</td>
</tr>
<tr>
<td>Chung 2004, US18</td>
<td>Surgical residents, 1 year analysis; n=not stated</td>
<td>Inpatient consultations and operative caseload: reduced. Outpatient clinic attendance: NS</td>
</tr>
<tr>
<td>Damadi 2007, US19</td>
<td>Chief residents in surgery, single centre, 4 year analysis; n=17</td>
<td>Chief or assistant operator operative caseload: reduced</td>
</tr>
<tr>
<td>Elbadrawy 2008, UK16</td>
<td>Gynaecology, single centre, 2 year analysis; n=not stated</td>
<td>Operative caseload, (total experience, and major, intermediate, minor, hysterectomy, laparoscopy all analyses separately): all reduced</td>
</tr>
<tr>
<td>Feanny 2005, US20</td>
<td>PGY 4 and 5 surgical residents, single centre, 2 year analysis; n=13</td>
<td>Operative caseload (total, chief operator, first assistant, and emergency operator): reduced</td>
</tr>
<tr>
<td>Fernandez 2009, UK14</td>
<td>Anaesthetic trainees, single centre, 82 month analysis; n=62</td>
<td>Total No of cases, and subgroup analyses of ASA I-III and emergency cases: all reduced. Subgroup analyses of ASA IV/V or subspecialty cases: NS</td>
</tr>
<tr>
<td>Kairys 2008, US21</td>
<td>Surgical residents, multicentre, 5 year analysis; n=not stated</td>
<td>Total major operations: reduced. Chief surgeon: reduced. Assistant surgeon: NS</td>
</tr>
<tr>
<td>Kara 2008, UK11</td>
<td>Surgical trainees, single centre, 46 month analysis; n=not stated</td>
<td>Operative caseload: reduced. Subgroup analyses for PGYs of training: all individually reduced. No statistical tests reported</td>
</tr>
<tr>
<td>Maxwell 2010, UK12</td>
<td>Neurosurgical trainees, single centre, 200 case records reviewed; n= not stated</td>
<td>Continuity of care in elective operations—consent and operate: NS, operate and follow-up: reduced. Continuity of care in emergency operations—admit and operate: reduced, consent, and operate: reduced</td>
</tr>
<tr>
<td>Searle 2008, UK15</td>
<td>Obstetric anaesthetists, single centre, 2 year analysis; n=not stated</td>
<td>General anaesthetic cases: reduced. Training opportunities: reduced. No statistical tests reported</td>
</tr>
<tr>
<td>Stephens 2004, UK13</td>
<td>Surgical SHOs, single centre, 12.5 year analysis; n=95</td>
<td>Elective surgical caseload: reduced. Inguinal herniorrhaphy: reduced. Appendicectomy: NS</td>
</tr>
<tr>
<td>Watson 2010, US22</td>
<td>Surgical residents, single centre, 40 resident years</td>
<td>Major surgery caseload: reduced</td>
</tr>
</tbody>
</table>

NS=not significant; PGY=postgraduate year; ASA=American Society of Anaesthesiologists' physical status score.

One in gynaecology. All except one11 were single centre, and two did not conduct statistical analysis of their results.11,13 Six of the US papers were conducted in the US,17,22 including the only large multicentre study to show a reduction in caseload.22 Of the two UK studies of surgical trainees, one showed a reduction in operative caseload with a reduction in hours from 72 to 48 a week,13 and the other with a reduction from 58 to 54 hours a week.11 Both these studies were of low methodological quality: Kara et al did not conduct a statistical analysis of their results11 and Stephens et al reported a single centre cohort study of only the most junior surgical trainees.13 Both studies of trainees in anaesthetics that reported a reduction in caseload were conducted in the UK. One paper examined the impact of reducing hours from 60 to 56 a week in a specialist paediatric centre14; the other reported the effect of reducing hours from 72 to 58 a week on trainees’ exposure to obstetric general anaesthesia.15 The authors of the latter paper did not statistically analyse their results and commented that the total operative caseload of the department had also fallen, providing an alternative explanation for the reduced number of cases with trainee involvement after reducing working hours.15 The only study of gynaecology trainees that reported a reduction in caseload was conducted in a single UK centre and found a reduction in both the overall number of cases involving trainees and in subgroup analyses of specific procedures.16

Most papers analysed found no difference in caseload associated with a reduction in working hours. Eighteen originated from the US: 14 in surgery,22,27,28,30,37,41,42 and four in obstetrics and gynaecology.23-26 Four of the US papers examining surgical training analysed large multicentre cohorts,29,30,39,40 though the actual working hours of the residents studied in these four papers was not reported and so compliance with duty hour recommendations could not be assessed. Only six of the 14 US papers that used operative caseload as an outcome measure reported institutional case numbers before and after duty hours reform.19,22,31,33,36,37 Four UK studies of anaesthesia training reported no change in either caseload or supervision rates55,56; three of these documented concurrent institutional activity,54,55 but two did not conduct statistical analyses on their results.14,46 Four studies in obstetrics and gynaecology conducted in the US reported no change in caseload with reform of duty hours; one was multicentre,20 three reported institutional volumes.23,25,2621 Only two of these four papers detailed both the actual hours worked and the shift patterns.23,24

Studies of patients’ outcomes

Thirty four papers documented the impact of reducing the working hours of doctors in training on patient outcomes (tables 4-6). Only three studies originated from the UK,50,52 the rest were conducted in the US. Most were “before and after” cohort studies, but one was a randomised controlled trial.53 Most studies reported clinical outcomes such as morbidity and mortality or measures of resource use such as length of stay. Seven, however, reported on patient safety indicators such as rates of adverse events or medical errors,50,53,58 and one used continuity of care in paediatrics as an outcome measure.50 Sixteen studies were in surgery or surgical subspecialties,31,35,38,39,56,58,60-69 seven were in internal medicine,10,52,76-74 two were in critical care,53,75 two
were in pediatrics, and one was in obstetrics. Six studies evaluated multispecialty cohorts. As none of these studies adjusted their statistical analyses appropriately for multiple comparisons, it is possible that some of the significant results in studies that reported multiple outcomes, particularly those where there was a combination of positive, negative, and neutral results, might have occurred by chance. For studies reporting multiple outcome measures, we have categorised them according to whether most outcomes were better, worse, or showed no difference with a reduction in working hours.

**Studies showing predominantly improved outcomes in patients**

Four studies showed improvement in patients’ outcomes with a reduction in working hours; one was conducted in the UK and three in the US. One of these was a randomised controlled trial of high methodological quality that randomised residents in critical and coronary care units to two different working patterns. Patient safety indicators including medical and diagnostic error rates were compared and found to be improved after a change in doctors’ working hours from a traditional shift system, with up to 37 continuous duty hours and 77-81 hours worked a week, to a rota that eliminated extended shifts and reduced the working week to 60-63 hours. Only one of the remaining three cohort studies was a large multicentre study, which found a difference in mortality rates. This paper studied trauma admissions over a four year period and compared outcomes in teaching hospitals before and after duty hour reforms with concurrent outcomes in non-teaching centres (and therefore unaffected by trainee working hours). The authors found improvements in mortality and length of stay in intensive care in teaching departments that were not seen in the control hospitals. One small single centre study from the UK found that rates of adverse events were reduced with a shortening of working hours, though this study could be criticised as different medical specialties were examined in the two cohorts. Finally, a large single centre study of patients undergoing laparoscopic cholecystectomy showed an improvement in complication rates after a reduction in working hours of surgical trainees.

**Studies showing predominantly worse outcomes in patients**

Two large studies in trauma and orthopaedics, both from the US, found that rates of complications, but not mortality, worsened with a reduction in working hours. Browne et al’s multicentre analysis of patients undergoing surgery for hip fracture found that the occurrence of nine postoperative complications was worse in teaching hospitals (but not in control non-teaching hospitals); the incidence of eight other complications was similar and only one was improved. The length of stay in hospital and routine discharge rates were also worse in teaching hospitals after reforms in duty hours. As the statistical analyses were not adjusted for multiple comparisons, the significant findings could still have occurred by chance. Salim et al’s study of trauma patients showed an increase in complication rates in a large single centre cohort over four years (two before and two after duty hour reform). Baseline characteristics of the two cohorts were presented and were similar, though the outcomes were not adjusted for patients’ risk; it is possible therefore that the difference in complications could be attributable to unknown differences in risk factors in patients. Furthermore, though the differences in complication rates were significant, they were small (total complication rate 5.64% vs 7.28%; preventable complication rate 0.89% vs 1.28%; non-preventable complication rate 4.75% vs 5.81%), and therefore the clinical relevance of these changes could be questioned. Nevertheless, the authors highlighted that these results were worrying and could at least in part be attributable to the change in duty hours.

**Studies showing predominantly no difference in outcomes in patients**

Most studies showed that duty hour reform did not affect standards of care of patients. Twenty eight studies reported no significant difference in patients’ outcomes after reductions in the working hours of doctors in training or reported a combination of positive, negative, and similar outcomes. Only two of these originated from the UK. The study by Collum et al...
### Table 5: Studies on effect of reduced working hours that found no change or variable changes in different patients’ outcomes

<table>
<thead>
<tr>
<th>Study</th>
<th>Details</th>
<th>Outcomes (change after reduction in hours)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Afessa 2005, US$^{75}$</td>
<td>Medical ICU patients, single centre, 21 week analysis; n=626</td>
<td>Intensive care mortality: NS. Length of stay in intensive care: NS. Hospital mortality: NS. Hospital length of stay: NS</td>
</tr>
<tr>
<td>Bailit 2004, US$^{54}$</td>
<td>Obstetrics patients 12 month analysis; n=not stated</td>
<td>3rd and 4th degree lacerations, umbilical arterial pH &lt;7.0, fever in labour, primary LSCS rate, GA LSCS rate, incident reporting: all NS. Postpartum haemorrhage: decreased. Neonatal resuscitations: decreased. Reported medication errors associated with resident performance: too infrequent for comparison across time periods</td>
</tr>
<tr>
<td>Bell 2010, US$^{76}$</td>
<td>Very low birthweight infants, multicentre, 5 year analysis; n=11 137</td>
<td>Death within 7 days: NS. Death within 28 days: NS. Short term morbidity: NS</td>
</tr>
<tr>
<td>Callum 2010, UK$^{11}$</td>
<td>All hospital inpatients, multicentre, 3 year analysis, n= not stated</td>
<td>Hospital SMR: No change</td>
</tr>
<tr>
<td>De Virgilio 2007, US$^{51}$</td>
<td>Trauma patients, single centre, 8 year analysis; n=3491</td>
<td>Overall complications: NS. ARDS, renal failure, wound complications, wound infection: all NS. Intra-abdominal abscess, pneumonia, DIC: all improved. Pulmonary embolus, septicemia: both worse</td>
</tr>
<tr>
<td>Gopaldas 2010, US$^{64}$</td>
<td>CardiocThoracic patients, multicentre, 10 year analysis, n=374 941 (teaching hospitals); n=239 222 (control group, non-teaching hospitals)</td>
<td>Mortality: improvement in both teaching and non-teaching hospitals. No of complications per patient: increased in teaching hospitals but reduced in non-teaching hospitals. Length of stay: reduced in both teaching and non-teaching hospitals</td>
</tr>
<tr>
<td>Gottlieb 1991, US$^{70}$</td>
<td>Internal medicine patients, single centre, 8 week analysis; n=1103</td>
<td>In-hospital mortality: NS. Hospital readmission: NS. Serious medication errors: reduced. Nosocomial fever: reduced. Hospital length of stay: reduced</td>
</tr>
<tr>
<td>Helling 2010, US$^{53}$</td>
<td>Trauma patients, multicentre, 5 year analysis, n=99 407 (level 1 centres with residents); n=26 969 (level 2 centres without residents)</td>
<td>Mortality: significant reduction in both level I and level II centres; no difference (non-inferiority test) in change in outcome between level I and level II centres. HLOS: NS in both level I or level II centres</td>
</tr>
<tr>
<td>Horwitz 2007, US$^{57}$</td>
<td>Internal medicine patients, single centre, 2 year analysis; n=14 260 (teaching wards); n=6664 (control group, non-teaching wards)</td>
<td>Intensive care unit admission, rate of discharge to home or rehabilitation facility elsewhere, pharmacist interventions to prevent error: all improved in teaching hospitals compared with non-teaching hospitals after reform. Adverse drug reactions, mortality, length of stay: all NS</td>
</tr>
<tr>
<td>Howard 2004, US$^{52}$</td>
<td>Adult patients with primary diagnosis of congestive cardiac failure, acute myocardial infarction or pneumonia, multicentre, 3 year analysis; n=170 214 (teaching hospitals); n=143 455 (control group; non-teaching hospitals)</td>
<td>Mortality: beneficial trend towards lower mortality over time nearly identical between teaching and non-teaching hospitals. Improvement in outcome not necessarily attributable to change in working pattern</td>
</tr>
<tr>
<td>Hutter 2006, US$^{75}$</td>
<td>Surgical inpatients, single centre, 2 year analysis; n=3976</td>
<td>Morbidity: NS. Mortality: NS</td>
</tr>
<tr>
<td>Kaafarani 2005, US$^{64}$</td>
<td>Surgical inpatients, single centre, 2 year analysis; n=1197</td>
<td>Morbidity: NS. Mortality: NS</td>
</tr>
<tr>
<td>Laine 1993, US$^{51}$</td>
<td>Internal medicine patients, single centre, 2 month analysis; n=526</td>
<td>No of patients suffering at least one in-hospital complication: increased. No of patients having at least one diagnostic delay: increased. Major morbidity: NS. Mortality: NS</td>
</tr>
<tr>
<td>Lofgren 1990, US$^{74}$</td>
<td>Internal medicine patients, single centre, 8 month analysis; n=146</td>
<td>Inpatient mortality, complications, hospital length of stay, and No of consultations, No of procedures, No of radiographs: all NS. No of laboratory tests ordered: increased</td>
</tr>
<tr>
<td>McBurey 2008, US$^{51}$</td>
<td>Paediatric visits, single centre, 1 year analysis; n=6431</td>
<td>Continuity of care: NS</td>
</tr>
<tr>
<td>McIntyre 2010, UK$^{12}$</td>
<td>Emergency medical patients, single centre, 2 year analysis; n=16 974</td>
<td>In-hospital mortality: NS. Hospital LOS: NS. Readmissions within 30 days of discharge: NS</td>
</tr>
<tr>
<td>Myck 2005, US$^{51}$</td>
<td>All hospital inpatients, single centre, 1 year analysis; n=not stated</td>
<td>Adverse drug event (ADE) incidence rates; No of confirmed ADEs; number of ADEs per 100 patient days; No of preventable ADEs: all outcomes: NS</td>
</tr>
<tr>
<td>Naylor 2005, US$^{51}$</td>
<td>Emergency cholecystectomy patients, single centre, n=275</td>
<td>Complication rates: NS</td>
</tr>
<tr>
<td>Rogers 2005, US$^{57}$</td>
<td>Trauma patients, single centre, 1 year analysis; n=1092</td>
<td>Delayed diagnosis and missed injury: NS. Complication rate: NS</td>
</tr>
<tr>
<td>Rosen 2009, US$^{54}$</td>
<td>Surgical patients, multicentre, 5 year analysis (3 years pre-reform; 2 years post); VA hospitals: n=826 047, Medicare hospitals: n=13 367 273</td>
<td>Continuity of care: NS. Differences in either VA or Medicare hospitals. Technical care: NS in Medicare hospitals, VA hospitals: increase in odds ratio of technical care PSI event in more less teaching-intensive hospitals; NS relative changes in year 2 after reform. Other composite rates: increased in VA in year 2 after reform in more less teaching-intensive hospitals but not in Medicare in either year after reform</td>
</tr>
<tr>
<td>Shenans 2005, US$^{54}$</td>
<td>Trauma patients, single centre, 22 months; n=2826</td>
<td>Length of hospital stay: NS. Length of ICU stay: NS. Ventilator days: NS. Mortality: NS. Complications: NS</td>
</tr>
<tr>
<td>Shetty 2007, US$^{57}$</td>
<td>Medical and surgical patients with 35 diagnoses associated with high mortality rates, multicentre: n=1 268 738 (medical), n=243 207 (surgical)</td>
<td>Mortality: medical patients: improved, surgical patients: NS</td>
</tr>
<tr>
<td>Shonka 2009, US$^{38}$</td>
<td>Otolaryngology patients, single centre, 7 year analysis; n=not stated</td>
<td>Mortality, 30 day readmission rates; length of stay: NS difference in any outcome across years</td>
</tr>
<tr>
<td>Silber 2009, US$^{53}$</td>
<td>Medical (acute myocardial infarction, heart failure, stroke, or gastrointestinal bleeding) and surgical (general, orthopaedic, and vascular) patients in Medicare and VA hospitals, 5 year analysis; Medicare: n=6 059 015; Veterans Affairs (VA): n=210 276</td>
<td>Hospital length of stay: NS difference between cohorts in Medicare or VA hospitals for either medical or surgical patients</td>
</tr>
<tr>
<td>Volpp 2009, US$^{75}$</td>
<td>VA hospital admissions: medical (principal diagnoses of acute myocardial infarction, congestive heart failure, gastrointestinal bleeding or stroke) and surgical (general, orthopaedic, or vascular surgical patients), multicentre, 6 year analysis; n=318 636</td>
<td>Mortality: medical patients: similar improvement in unadjusted mortality rates in hospitals in all fourths of resident:bed ratios; surgical patients: no apparent difference observed in hospitals with different resident:bed ratios</td>
</tr>
<tr>
<td>Volpp 2009, US$^{75}$</td>
<td>Medicare hospital admissions: medical (principal diagnoses of acute myocardial infarction, congestive heart failure, gastrointestinal bleeding, or stroke) and surgical (general, orthopaedic, or vascular surgical patients), multicentre, 6 year analysis; n=8 529 595</td>
<td>Mortality: comparison of the most teaching intensive hospitals with non-teaching hospitals: NS relative change in mortality in either medical or surgical patients</td>
</tr>
</tbody>
</table>

NS=not significant; NY=New York; VA=Veterans’ Affairs.
examined changes in hospital standardised mortality ratios (HSMRs) in one region of the UK over a three year period and found that they improved over time in line with improvement in such ratios in the rest of the UK.\(^5\) The authors reported no specific details about changes in duty hours, either locally or nationally; it is therefore possible that similar changes in working patterns were implemented outside the geographical area of interest. This raises the question of whether all these improvements were associated with reducing working hours or coincidental with it. McIntyre et al examined outcomes of emergency medical admissions over a two year period and found similar mortality, length of stay in hospital, and 30 day readmission rates.\(^5\) This was a small single centre study, however, and might have been underpowered to detect differences in outcome.

Twenty six studies from the US showed similar or mixed patients’ outcomes after a reduction in duty hours. Several of these analysed relatively small or single centre cohorts of patients and might have been underpowered to detect changes in some of the outcomes reported. These included studies in critical care,\(^6\) obstetrics,\(^7\) neonatology,\(^8\) paediatrics,\(^9\) trauma,\(^10\) surgery,\(^11\) and internal medicine.\(^12\) Five US papers included concurrent analyses of “control” hospitals or wards that did not employ doctors in training and were therefore unaffected by changes in duty hour regulations.\(^13,14,15\) Four of these five papers compared mortality rates: three found similar improvements over time in both teaching and non-teaching hospitals in cardio thoracic,\(^16\) trauma,\(^17\) and medical patients\(^18\); one study of medical inpatients was underpowered to detect a difference.\(^19\)

Three other US studies compared outcomes in hospitals of different “teaching intensity”\(^20\) (as defined by the ratio of residents to beds).\(^21\) Rosen and co-workers studied patient safety indicators in surgical patients and found a combination of positive, negative, and neutral results for different outcomes.\(^22\) The two other studies were both conducted by Volpp et al, with identical methods and over the same time period in two different US healthcare systems (Veterans Affairs and Medicare) and looked at mortality rates in medical and surgical patients separately.\(^23\) In Veterans Affairs hospitals, outcomes for medical patients improved with duty hour reforms in hospitals of higher “teaching intensity”; for surgical patients, however, there was no significant change in outcome associated with the reduction in working time.\(^24\) These data are similar to those from one other study that found an improvement in mortality in medical patients but not surgical patients.\(^25\) In the study of Medicare hospitals, however, there was no apparent change in outcome in either medical or surgical patients after a reduction in duty hours.\(^26\) Similarly, Silber et al found no difference in length of stay in hospital either in medical or surgical patients admitted to either Medicare or Veterans Affairs hospitals.\(^27\)

**DISCUSSION**

In this systematic review of the impact of a reduction in working hours for junior doctors, using objective measures of postgraduate medical training, patient safety, and clinical outcome, we found that most studies found either a beneficial or neutral impact on patient safety and clinical outcome and limited or no effect on postgraduate training as measured by procedural volume per trainee or examination results. When duty hours are reduced to below 56 or 48 hours a week, in accordance with European legislation, we could not draw conclusions on the impact on patients outcomes or medical training because of conflicting results from different institutions and specialities and the poor quality of some of the studies evaluated.

**Study limitations**

Our study is limited by several factors. Firstly, the heterogeneity of the included studies does not permit meta-analysis of the results, and their predominantly retrospective observational methods means that the individual studies, and therefore our review, could be subject to bias. Secondly, the included studies varied considerably in quality, though we did include an evaluation of their quality in our results (see tables B and D in the appendix on bmj.com). Some of the studies that evaluated patient safety did not include risk adjustment or present baseline characteristics of patients, which would allow true comparison of clinical outcomes. Several studies that evaluated training included no statistical analysis or had poorly defined cohort characteristics. As several postgraduate education studies did not report denominator data on overall hospital activity, a change in hospital workload could have accounted for differences in trainee caseload numbers after reducing working hours. Thirdly, the outcome measures used in the individual studies varied

**Table 6** | Studies on effect of reduced working hours that predominantly found deterioration in patients’ outcomes

<table>
<thead>
<tr>
<th>Study</th>
<th>Details</th>
<th>Outcomes (change after reduction in hours)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Browne 2009, US(^6)</td>
<td>Hip fracture patients, multicentre, 4 year analysis; n=48 430 in teaching and non-teaching (control) hospitals</td>
<td>Death: NS. Complications: 8 categories: NS, 9 categories: worse in teaching hospitals, 1 category: improved in teaching hospitals. Length of stay in hospital and routine discharge: both worse in teaching hospitals</td>
</tr>
<tr>
<td>Salim 2007, US(^7)</td>
<td>Trauma patients, single centre, 4 year analysis; n=116 854</td>
<td>Mortality: NS in total or preventable death rates. Increased. Preventable complication rate: increased. Non-preventable complication rate: increased</td>
</tr>
</tbody>
</table>

NS=not significant.
Table 7 Summary of changes and recommendations in duty hours

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum duty hours/week</td>
<td>80 hours, averaged over 4 weeks</td>
<td>80 hours, averaged over 4 weeks</td>
<td>80 hours, averaged over 4 weeks</td>
<td>56 hours, averaged over 26 weeks</td>
<td>56 hours averaged over 26 weeks</td>
<td>48 hours averaged over 26 weeks</td>
</tr>
<tr>
<td>Maximum shift length</td>
<td>24 hours with 3 hour transition period</td>
<td>30 hours (admitting patients up to 24 hours, then 6 additional hours for transitional and educational activities)</td>
<td>30 hours (admitting patients for up to 16 hours, plus 5 hour protected sleep period between 10 pm and 8 am, with remaining hours for transitional and educational activities)</td>
<td>No restriction</td>
<td>13 hours</td>
<td>13 hours</td>
</tr>
<tr>
<td>Minimum rest period between shifts</td>
<td>8 hours. At least one 24 hour period off duty/week</td>
<td>10 hours after day shift</td>
<td>10 hours after day shift; 12 hours after night shift; 14 hours after any extended duty period of 30 hours, not returning until 6 am next day</td>
<td>8 hours between shifts, 24 hours every 7 days or 48 hours every 14 days</td>
<td>11 hours between shifts</td>
<td>11 hours between shifts</td>
</tr>
</tbody>
</table>

considerably, and, particularly in the studies evaluating postgraduate medical training, the reported measures might not have been appropriate for the intended purpose. Finally, publication bias could have affected our results, though we attempted to avoid this by conducting an extensive review of the “grey literature.”

Scope and limitations of published literature

We can make several observations regarding the nature and origin of the published studies in this topic. Firstly, the concern over the implications of duty hour regulations has led to a much larger and higher quality body of work from the US than Europe. This could in part be because the most significant changes in the US occurred in 1989 and 2003, as opposed to European changes, which are more recent; there might be European studies yet to be completed or published. Secondly, it is notable that despite the lack of language restrictions and an extensive review of the “grey literature,” we were unable to identify any European studies that fulfilled our inclusion criteria and were conducted outside the UK. This could reflect that the concerns of the US and UK medical communities are not reciprocated elsewhere or perhaps that compliance with working time legislation is not as rigorous in mainland Europe as in the UK. Thirdly, most studies of postgraduate education analysed cohorts of doctors in training in “craft specialties” that have an emphasis on technical competence, such as surgery, anaesthesia, and obstetrics and gynaecology. It is therefore possible that internal medicine physicians do not believe that restrictions on working hours will have negative implications for training. Alternatively, it could be that there are fewer easily interpretable measures of training outcome for physicians than there are in more technically focused specialties. Finally, there is a much larger body of literature examining the implications of reducing working hours on patients’ outcomes originating from the US compared with the UK. Possible reasons for this include a belief by the UK healthcare community that legislative changes are unlikely to affect quality of care of patients, or perhaps, more probably, the lack of robust systems for collecting and analysing large datasets of risk adjusted outcomes in patients in the UK and Europe.

Impact on patients’ outcomes

The literature examining the association between reducing working hours and objective measures of outcomes in patients shows that there is no clear signal to indicate either benefit or harm. While it might seem intuitive that doctors working fewer hours will be less tired, make fewer errors, and that therefore patients’ outcomes should improve, both clinical experience and the results of studies conducted with identical methods but in different healthcare systems indicate that other influences might be at least as important.79 80 These could include the number and quality of clinical handovers,81 the level of supervision of doctors in training,82 and factors unrelated to doctors’ duty hours and schedules such as the continuity of care provided by the entire multidisciplinary team,83 the standard of nursing care,84 and many other differences within and between institutions in delivery of healthcare. The heterogeneity of results in our review might be due to such differences related to structure and process, and explaining these variations is the focus of a growing body of work.85-87 A greater understanding of such issues would be valuable in ascertaining if there are standards of care that might help to improve outcomes in patients when doctors’ working patterns are changed so that such standards could be more widely implemented.

Impact on educational outcomes

The studies focusing on training outcomes indicate that the Code 405 or ACGME regulations in the US have had an inconsistent effect and that there are insufficient data from studies of high methodological quality to be able to draw firm conclusions on the impact of the European Working Time Directive or New Deal in the UK. When interpreting these findings, one should consider that, alongside the reduction in working hours, other changes in postgraduate medical education and provision of healthcare might influence the quality of training. In the UK, training has moved away from an apprenticeship model to a time limited programme, with greater emphasis on clinical supervision.
Concurrently, and consequently, there has been a change in the role of the junior doctor in healthcare service delivery. Use of ancillary staff to provide services previously provided by junior doctors, such as phlebotomy, cannulation, and basic administrative duties, allows more time for useful training activity in a working week limited by duty hour regulations. It is possible that cohorts of doctors whose training outcomes did not change with a reduction in working hours have also seen an alteration in their clinical and administrative responsibilities. The requirement for this sort of change was a key recommendation of the recent inquiry into the implications of the European Working Time Directive on medical training in England: “make every moment count.”

One of the difficulties in evaluating the effects of changes in working hours is the lack of validated measures for assessing the outcome of training. We chose to focus on studies that reported objective measures, such as operative case numbers or clinic attendances, as these are less likely to be biased than subjective surveys of the profession’s opinion. Even these objective measures, however, do not fully evaluate the quality of training. For example, case numbers might be a helpful guide to experience but not of training quality: performing 100 simple cases might not be as useful as undertaking 10 complex procedures. Studies in “craft specialties” focused predominantly on the number of operative cases undertaken. In all these specialties, however, time spent outside the operating theatre, such as in the outpatient clinic, emergency room, and on inpatient wards, also has value and is important for the development of a rounded independent practitioner. Furthermore, quantitative measures such as procedural volume must be differentiated from measures of knowledge, skill, attitude, and behaviour, all of which are essential attributes for the independent medical practitioner. In the UK, workplace based assessments, such as direct observation of procedural skills, case based discussion, and multi-source feedback, are now being used to evaluate competence and aid progression of junior doctors. Use of other assessments of procedural skill, such as cumulative sum (CUSUM) analysis, and evaluating the association of the outcomes of these assessments with patients’ outcomes also requires further evaluation.

Ultimately, the goal of postgraduate medical training is to produce doctors who are safe competent independent practitioners. Both in the UK and US, super-specialisation, particularly in “craft specialties,” means that a consultant or attending physician today might not be required to carry out the same range of independent duties as their predecessors. Additionally, the medical profession is seeking alternate methods of training doctors outside their working hours, including the use of medical simulation and clinical skills laboratories for training in human factors and technical procedures. It might therefore be possible for a time limited training programme to produce doctors who are clinically safe and able to work independently in the modern healthcare system but who might have previously been considered to have insufficient experience to fulfil a wider range of clinical responsibilities.

Conclusions and policy implications
The most important test of success of postgraduate training is the professional performance of doctors who reach the end of it. In this review, studies reporting patient safety and clinical outcome measures examined only the immediate effect of reductions in duty hours on care of patients; however, the potential impact of such changes can be fully evaluated only some years after duty hours reforms, when cohorts of doctors trained in a limited number of hours a week are finally practising independently. To that end, we have the following suggestions. Firstly, a consensus should be reached by the medical profession on appropriate measures to assess the quality of postgraduate medical training. These should be both quantitative and qualitative and could include both process measures, such as procedural volume, supervision levels, and time spent in “alternative” training environments such as simulators, and outcome measures such as the results of formative and summative assessments. Secondly, once such measures have been agreed, we recommend that they are confidentially reported to the organisations responsible for the quality assurance and regulation of training, such as the ACGME in the US, and the Medical Royal Colleges and General Medical Council in the UK. The reporting process should be conducted both on an individual doctor basis for trainees, interns, and residents, and on an institutional level for teaching departments and programmes. This would enable comparison between different regions and specialties and identification of practices that lead to high quality training. Finally, we recommend the conduct of longitudinal studies, evaluating the relation between the agreed process and outcome measures for postgraduate training and objective measures of outcomes related to patients for clinicians in the first few years of independent practice, such as rates of clinical errors or complications or medical negligence or malpractice claims. Such studies would enable the evaluation of the validity and reliability of measures of training quality for the prediction of the future standard of care of patients provided by independent clinicians and would permit the refinement and re-evaluation of these measures.

It has been stated that “training is patient safety for the next 30 years.” We wholeheartedly support this view. Our study, which summarises a diverse and sometimes methodologically flawed body of literature, has been unable to reach firm conclusions, but we consider that this in itself is an important observation. We have highlighted the need for a more systematic approach to evaluating the impact of legislative changes of duty hours and the challenges of conducting high quality audit and research in this area. In the future, it will only be through the conduct of large, collaborative, multicentre evaluations of training and outcome that both the public and the profession can be reassured that the standard of medical training, and
Reducing doctors’ working hours from over 80 hours a week has had limited impact on postgraduate training and minimal effect on patients’ outcomes in the US.

In the UK, there is insufficient evidence from high quality studies to draw a conclusion on the effect of reducing working hours to less than 56 or 48 hours a week on objective measures of postgraduate training or clinical outcomes.

therefore of future care of patients, is of the highest possible quality and will be maintained or improved over time.

**Contributors:** SRM and JDB were responsible for the conception and design of the study. SRM had principal responsibility for analysing and interpreting the data and for drafting the article, revisions, and final approval. JL, NS, and AM contributed to analysis and interpretation of the data. JDB contributed to drafting, revision, and final approval of the article.

**Funding:** SRM works within the UCL/UCLH Joint Comprehensive Biomedical Research Centre, which received a proportion of funding from the National Institute for Health Research’s funding scheme. Part of this work was conducted while SRM was a National Institute for Academic Anaesthesia (NIAA) Research Fellow, supported by a grant awarded to the NIAA’s Health Services Research Centre by the Frances and Augustus Newman Foundation. This study was not commissioned and no project-specific funding was received. The funders had no role in the study design, data collection, analysis, and interpretation; writing the manuscript; or the decision to submit the research for publication.

**Competing interests:** All authors have completed the Unified Competing Interest form at www.icmje.org/coi_disclosure.pdf (available on request from the corresponding author) and declare: no support from any organisation that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work.

**Ethical approval:** Not required.

**Data sharing:** No additional data available.


Ghaferi AA, Osborne NH, Birkmeyer JD, Dimick JB. Hospital characteristics associated with failure to rescue from complications after pancreatectomy. *J Am Coll Surg* 2010;211:325-30.


Accepted: 24 January 2011