St. Michael's
Inspired Care.
Inspiring Science.

Journal Club May 22, 1015

Articles for Review:

1. Kasotakis G, Lakha A, Sarkar B et al. Trainee Participation Is Associated With Adverse Outcomes in Emergency General Surgery: An Analysis of the National Surgical Quality Improvement Program Database. Ann Surg 2014; 260: 483-93
2. Rochon PA, Gurwitz JH, Sykora K et al. Reader's guide to critical appraisal of cohort
studies: 1. Role and design. BMJ 2005;330:895-897
3. Mamdani M, Sykora K, Li P et al. Reader's guide to critical appraisal of cohort studies: 2.
Assessing potential for confounding. BMJ 2005;330:960-2

Please read the above articles and be prepared to discuss the following:

1. What is the clinical question being addressed?
2. What is the study design? Is it appropriate?
3. What are the sources of data?
4. Are the data accurate and valid?
5. What was the intervention and how was it assessed?
6. What outcomes were considered?
7. Are there other clinically important outcomes that should be considered?
8. How large and precise was the difference in the outcome?
9. Are there any factors (confounders) that might explain the differences in the results?
10. State the conclusion. Have the authors addressed the clinical question posed?
11. Does the evidence support the conclusion?
12. Should the results of this study lead to changes in trainee participation in the care of emergency surgical patients?

Reader's guide to critical appraisal of cohort studies: 2. Assessing potential for confounding

Muhammad Mamdani, Kathy Sykora, Ping Li, Sharon-Lise T Normand, David L Streiner, Peter C Austin, Paula A Rochon, Geoffrey M Anderson

Although confounding is an important problem of cohort studies, its effects can be minimised to enable valid comparison

This is the second of three articles on appraising cohort studies

Institute for Clinical Evaluative Sciences, Toronto, ON Canada
Muhammad Mamdani *senior scientist*
Kathy Sykora *senior biostatistician*
Ping Li *analyst*
Peter C Austin *senior scientist*
Department of Health Care Policy, Harvard Medical School, Boston, USA
Sharon-Lise T Normand *professor of health care policy (biostatistics)*
Department of Psychiatry, University of Toronto, ON, Canada
David L Streiner *professor*
Kunin-Lunenfeld Applied Research Unit, Baycrest Centre for Geriatric Care, Toronto, ON, Canada
Paula A Rochon *senior scientist*
Department of Health Policy, Management and Evaluation, Faculty of Medicine, University of Toronto, Toronto, ON Canada
Geoffrey M Anderson *chair in health management strategies*
Correspondence to: G M Anderson, Institute for Clinical Evaluative Sciences, 2075 Bayview Avenue, Toronto, ON M4N 3M5, Canada
geoff.anderson@utoronto.ca

BMJ 2005;330:960–2

In cohort studies, who does or does not receive an intervention is determined by practice patterns, personal choice, or policy decisions. This raises the possibility that the intervention and comparison groups may differ in characteristics that affect the study outcome, a problem called selection bias. If these characteristics have independent effects on the observed outcome in each group, they will create differences in outcomes between the groups apart from those related to the interventions being assessed. This effect is known as confounding.[1] In the first paper in the series we dealt with the design and use of cohort studies and how to identify selection bias.[2] This paper focuses on the definition and assessment of confounders.

What is a confounder?

For a characteristic to be a confounder in a particular study, it must meet two criteria.[1] The first is that it must be related to the outcome in terms of prognosis or susceptibility. For example, in the study of the association between antipsychotic use and hip fracture that we considered in the first paper,[2] age is known to be related to risk of hip fracture and therefore has the potential to be a confounder.

The second criterion that defines a confounder is that the distribution of the characteristic is different in the groups being compared. It can differ in terms of either the mean or the degree of variation or variability in that characteristic. For example, for age to be a confounder in a cohort study, either the average age or the variation in the age in the groups being compared would have to be different. Assessing variation as well as average values is important because groups can have the same average value but very different variation. For example, one group with an average age of 70 could include only people aged 70 and another with the same average age could consist of equal proportions of individuals aged 50 and 90. Nevertheless, even a characteristic that is a strong predictor of outcome will not be a confounder if its distribution is balanced between the comparison groups.

In assessing cohort studies, it is important to identify potential confounders and to examine their distribution in the intervention and comparison groups. Below we describe the three questions that need to be answered.

Has there been a systematic effort to identify and measure potential confounders?

Although currently available evidence helps identify potential confounders, the imperfect state of knowledge means that some characteristics related to the outcome may not have been discovered (unknown confounders). Even if a confounder is known, there may be insufficient data to evaluate it.

In randomised controlled trials, all potential confounders (known or unknown) are expected to be evenly distributed between the groups being compared.[3] Cohort studies, however, have no similar protection against confounding and are especially vulnerable to unknown confounders. This does not mean that all cohort studies are inherently invalid. The unknown potential confounders may not have a large independent effect on the outcome of interest and, therefore, even if unevenly distributed, might not result in much bias. Unknown potential confounders may also be evenly distributed between the groups. Nevertheless, all cohort studies should recognise that unknown confounders could affect the results and, as outlined in the next article in this series,[4] investigators should make an effort to determine how sensitive the results are to unknown confounders.

Although unknown confounders are difficult to deal with in cohort studies, a systematic approach can be used to identify known confounders. This should start with a well designed search of comprehensive databases such as Medline. In the context of the study of the relation between antipsychotic use and the outcome of a hip fracture, a review of the literature suggests that risk factors for hip fracture can be broken down into four categories[5–10]:

- Features of medical history—for example, stroke, osteoporosis
- Exposure to drugs—for example, benzodiazepines, oestrogens
- Demographics—for example, age and sex
- Social and behavioural factors—for example, exercise and diet.

Once the potential confounders have been identified, the next step is to develop ways to measure these in the groups being studied. In many cases, especially when using administrative databases, it may not be possible to measure all known confounders. Even if they are measured, the reliability and validity of the measurement technique may be unclear. In the hip fracture and atypical antipsychotic example (see bmj.com for details of how the cohort was created) we used administrative databases to measure known confounders. These databases are poor sources of information on behavioural and social factors. The failure to include measures of these factors has been identified as a key issue in cohort studies of hip

P+ *Further details on the study cohort and propensity scores are on bmj.com*

fracture,[11] and lack of control for lifestyle factors has been suggested to have a key role in the differences in risk of cardiovascular disease seen in cohort and randomised controlled studies of hormone replacement therapy.[12] Although the administrative databases can provide some information on patient history such as previous falls, they may underestimate their true prevalence. It is important to know which confounders have been measured in the study and how well they have been measured.

Is there information on distribution of potential confounders between groups?

Information on the distribution of potential confounders in the intervention and comparison groups is usually provided in the first table of the paper. Confounding is a problem only if these characteristics are unevenly distributed between the intervention and comparison groups. The table provides information on potential confounders for two comparisons examining the association between atypical antipsychotic use and hip fracture. Tables similar to this should be included in all cohort studies so that the reader can have an overview of the potential for selection bias and confounding.

What methods are used to assess differences in distribution of potential confounders?

Perhaps the most common strategy to identify important imbalances in individual confounders between intervention and comparison groups is to use significance tests such as χ^2 tests (for dichotomous variables) or *t* tests (for continuous variables). A problem with these tests is that the significance levels are sensitive to sample size, and the tests are usually not very meaningful when applied to studies with very large numbers of subjects (as is often the case for cohort studies). Under such circumstances, the differences may be significant but not clinically meaningful. For example, in the comparison restricted to people with dementia in the table, a difference of about three months in mean age between groups is significant ($P<0.001$) but may not be clinically relevant. Alternatively, if the samples are small, differences that are clinically meaningful may not be significant. For these reasons this approach to the assessment of differences is of little value.

An alternative to traditional significance testing is to use standardised differences or effect size to examine between group differences in patient characteristics.

CHRIS STOWERS/PANOS

Cohort characteristics can confound only if they vary between comparison groups

Baseline characteristics of study groups in comparisons of atypical antipsychotic versus no drug in all older people, and atypical versus typical antipsychotic drug in older people with dementia. Values are numbers (percentages) of patients unless stated otherwise

	Comparison 1: All older people				Comparison 2: Older people with dementia			
	Atypical antipsychotic (n=34 960)	No antipsychotic (n=1 251 435)	P value	Standardised difference	Atypical antipsychotic (n=21 427)	Typical antipsychotic (n=33 263)	P value	Standardised difference
Age (years):								
Mean (SD)	80.46 (7.63)	74.50 (6.58)	<0.001	0.90	81.69 (7.11)	81.96 (7.17)	<0.001	0.04
Median (interquartile range)	80 (75-86)	73 (69-79)	<0.001	0.90	82 (77-87)	82 (77-87)	<0.001	0.04
No (%) of women	21 720 (62.1)	714 829 (57.1)	<0.001	0.10	13 406 (62.6)	20 151 (60.6)	<0.001	0.04
Recent drug use								
Oestrogen	1 857 (5.3)	84 364 (6.7)	<0.001	0.06	1 000 (4.7)	983 (3.0)	<0.001	0.09
Bisphosphonates	2 323 (6.6)	48 353 (3.9)	<0.001	0.14	1 417 (6.6)	593 (1.8)	<0.001	0.26
Long acting benzodiazepines	1 177 (3.4)	29 917 (2.4)	<0.001	0.06	532 (2.5)	1 192 (3.6)	<0.001	0.06
Short acting benzodiazpeines	15 722 (45.0)	174 990 (14.0)	<0.001	0.88	9 016 (42.1)	14 267 (42.9)	0.06	0.02
Medical history								
Obesity	1 010 (2.9)	51 306 (4.1)	<0.001	0.06	492 (2.3)	945 (2.8)	<0.001	0.03
Previous falls	3 420 (9.8)	31 712 (2.5)	<0.001	0.45	2 460 (11.5)	3 940 (11.8)	0.196	0.01
Osteoporosis	3 509 (10.0)	84 034 (6.7)	<0.001	0.13	2 119 (9.9)	2 206 (6.6)	<0.001	0.12
Stroke	4 334 (12.4)	44 549 (3.6)	<0.001	0.46	2 779 (13.0)	4 638 (13.9)	0.001	0.03
Parkinsonism	3 613 (10.3)	20 990 (1.7)	<0.001	0.64	2 052 (9.6)	3 154 (9.5)	0.713	0.00
Alcoholism	2 014 (5.8)	18 155 (1.5)	<0.001	0.35	1 355 (6.3)	2 344 (7.0)	0.001	0.03
Hyperthyroidism	148 (0.4)	1 631 (0.1)	<0.001	0.08	83 (0.4)	129 (0.4)	0.993	0.00
Hyperparathyroidism	49 (0.1)	562 (0.04)	<0.001	0.04	31 (0.1)	23 (0.1)	0.006	0.02
Chronic renal failure	2 761 (7.9)	50 478 (4.0)	<0.001	0.19	1 656 (7.7)	2 473 (7.4)	0.204	0.01
Asthma or chronic obstructive pulmonary disease	9 014 (25.8)	240 202 (19.2)	<0.001	0.17	5 155 (24.1)	7 934 (23.9)	0.581	0.00
Rheumatoid arthritis	1 782 (5.1)	57 961 (4.6)	<0.001	0.02	1 014 (4.7)	1 752 (5.3)	0.005	0.02
Visual impairment	978 (2.8)	13 323 (1.1)	<0.001	0.17	623 (2.9)	975 (2.9)	0.873	0.00
Dementia	21 427 (61.3)	58 754 (4.7)	<0.001	2.53				.

Standardised differences reflect the mean difference as a percentage of the standard deviation. To estimate these, differences between groups are divided by the pooled standard deviation of the two groups. This measure of the distribution is not as sensitive to sample size as traditional tests and provides a sense of the relative magnitude of differences. Standardised differences of greater than 0.1 are typically felt to be meaningful.[13]

In the table, traditional significance testing found that all 19 potential confounders were significantly different ($P < 0.001$) in comparison 1, and that 13 of the 19 characteristics had standardised differences greater than 0.1. Of particular note is the large standardised difference for history of dementia. Restriction of the study to people with dementia eliminates the possibility of confounding from this characteristic. For comparison 2, traditional significance tests showed that 8 of the 18 potential confounders were significantly different ($P < 0.001$) but only two had a standardised difference greater than 0.1. The use of the standardised differences technique shows that comparison 1 has substantial selection bias, particularly for dementia, whereas comparison 2 has much less potential for bias.

Both traditional significance testing and standardised differences focus on one potential confounder at a time and do not provide an overall perspective on how the comparison groups differ. For example, two groups could have the same mean age and proportion of women, but one could contain old men and young women and the other old women and young men. An increasingly common approach to the analysis of cohort studies of health care interventions is to use propensity score methods[14 15]—a technique that involves multivariate assessment of confounders (see bmj.com for a brief discussion and an example).

Selection bias in cohort studies can result in confounding. Here we have defined questions that can help identify potential confounders. In the next article we will examine statistical methods that can be used to reduce the effect of confounding and strategies that can be used to determine if the results of a study are plausible.

Key questions

Has there been a systematic effort to identify and measure potential confounders?

Is there information on how the potential confounders are distributed between the comparison groups?

What methods are used to assess differences in the distribution of potential confounders?

We thank Jennifer Gold and Monica Lee for help in preparing the manuscript.

Contributors and sources: The series is based on discussions that took place at regular meetings of the Canadian Institute for Health Research chronic disease new emerging team. MM is a clinician with extensive research experience in cohort studies of prescription drugs who wrote the first draft of this article and is the guarantor. SLTN, DLS, and PCA are statisticians who commented on drafts of this paper. KS and PL programmed and conducted analyses. PAR and GMA conceived the idea for the series and GMA worked on drafts of this article and coordinated the development of the series.

Funding: This work was supported by a CIHR operating grant (CIHR No MOP 53124) and a CIHR chronic disease new emerging team programme (NET-54010).

Competing interests: None declared.

1 Altman D, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med* 2001;134:663-94.
2 Gurwitz JH, Sykora K, Mamdani M, Streiner DL, Garfinkel S, et al. Reader's guide to critical appraisal of cohort studies: 1. Role and design. *BMJ* 2005;330:895-7.
3 Altman D, Bland MJ. Treatment allocation in controlled trials: why randomise? *BMJ* 1999;318:1209.
4 Normand SLT, Sykora K, Li P, Mamdani M, Rochon PA, Anderson GM. Reader's guide to critical appraisal of cohort studies: 3. Analytical strategies to reduce confounding. *BMJ* (in press).
5 Lawlor DA, Patel R, Ebrahim S. Association between falls in elderly women and chronic diseases and drug use: cross sectional study. *BMJ* 2003;327:1-6.
6 Feskanich D, Willett W, Colditz G. Walking and leisure-time activity and risk of hip fracture in postmenopausal women. *JAMA* 2002;288:2300-6.
7 Haentjens P, Autier PH, Boonen S. Clinical risk factors for hip fracture in elderly women: a case-control study. *J Orthop Trauma* 2002;16:379-85.
8 Cummings SR, Nevitt MC, Browner WS, Stone K, Fox KM, Ensrud KE, et al. Risk factors for hip fracture in white women. *N Engl J Med* 1995;332:767-73.
9 Masud T, Morris RO. Epidemiology of falls. *Age Ageing* 2001;30(suppl 4):3-7.
10 Ensrud KE, Blackwell T, Mangione CM, Bowman PJ, Bauer DC, Schwartz A, et al. Central nervous system active medications and risk for fractures in older women. *Arch Intern Med* 2003;163:949-57.
11 Schneeweiss S, Wang PS. Association between SSRI use and hip fractures and the effect of residual confounding bias in claims database studies. *J Clin Psychopharmacol* 2004;24:632-8.
12 Humphrey LL, Chan BK, Sox HC. Postmenopausal hormone replacement therapy and primary prevention of cardiovascular disease, *Ann Intern Med* 2002;137:273-84.
13 Cohen J. *Statistical power analysis for the behavioural sciences.* Hillsdale, NJ: Academic Press, 1988.
14 Rubin DB, Estimating causal effects from large data sets using propensity scores. *Ann Intern Med* 1997;127:757-63.
15 Joffe MM, Rosenbaum PR. Propensity scores. *Am J Epidemiol* 1999;150:327-33.

(Accepted 18 February 2005)

Reader's guide to critical appraisal of cohort studies: 1. Role and design

Jerry H Gurwitz, Kathy Sykora, Muhammad Mamdani, David L Streiner, Susan Garfinkel, Sharon-Lise T Normand, Geoffrey M Anderson, Paula A Rochon

Cohort studies can provide valuable information unavailable from randomised trials, but readers need to be alert to possible flaws

Valid evidence on the benefits and risks of healthcare interventions is essential to rational decision making. Randomised controlled trials are considered the best method for providing evidence on efficacy. However, they face important ethical and logistical constraints and have been criticised for focusing on highly selected populations and outcomes.[1 2] Some of these problems can be overcome by cohort studies. Cohort studies can be thought of as natural experiments in which outcomes are measured in real world rather than experimental settings. They can evaluate large groups of diverse individuals, follow them for long periods, and provide information on a range of outcomes, including rare adverse events. However, the promise of cohort studies as a useful source of evidence needs to be balanced against concerns about the validity of that evidence.[3 4]

In this three paper series we will provide an approach to the critical appraisal of cohort studies. This article describes the role and design of cohort studies and explains how selection bias can confound the relation between the intervention and the outcome. The second article will outline strategies for identification and assessment of the potential for confounding, and the third article describes statistical techniques that can be used to deal with confounding. Each paper defines a set of questions that, taken together, can provide readers with a systematic approach to critically assessing evidence from cohort studies.

Randomised trial or cohort study?

Cohort studies are similar to randomised controlled trials in that they compare outcomes in groups that did and did not receive an intervention. The main difference is that allocation of individuals is not by chance. Table 1 gives some important similarities and differences between the two types of study. Because they are expensive and recruiting patients can be difficult, randomised controlled trials are generally short term and used to determine efficacy in selected populations under strict conditions. Cohort studies can be used to determine if the efficacy observed in randomised trials translates into effectiveness in broader populations and more realistic settings and to provide information on adverse events and risks.[5]

PHOTOALTO/PHOTONICA

Cohort studies can use diverse populations

Selection bias as a threat to validity

The internal validity of a study is defined as the extent to which the observed difference in outcomes between the two comparison groups can be attributed to the intervention rather than other factors. The biggest advantage of randomised controlled trials compared with cohort studies is that the random allocation process enhances the internal validity of a study by minimising selection bias and confounding.[6] This paper relies on the definitions provided by CONSORT (box 1).[7]

Allocation by chance in a randomised controlled trial should mean that the groups being compared are similar in terms of both measured and unmeasured baseline factors.[8] This is not so in cohort studies, and therefore cohort studies are vulnerable to selection bias. In cohort studies, factors that determined whether a person received the intervention could result in the groups differing in factors related to the outcome, either because people were preferentially selected to receive one treatment or because of choices that they made. These baseline differences in prognosis could confound the assessment of the effect of the intervention.

In cohort studies care must be taken to minimise, assess, and deal with selection bias. A comprehensive

This is the first of three articles on appraising cohort studies

Kunin-Lunenfeld Applied Research Unit, Baycrest Centre for Geriatric Care, Toronto, ON, Canada
Paula A Rochon *senior scientist*

Meyers Primary Care Institute, Worcester, MA 01605, USA
Jerry H Gurwitz *executive director*

Institute for Clinical Evaluative Sciences, Toronto, ON, Canada
Kathy Sykora *senior biostatistician*
Muhammad Mamdani *senior scientist*
Susan Garfinkel *research coordinator*

Department of Psychiatry, University of Toronto, Toronto, ON, Canada
David L Streiner *professor*

Department of Health Care Policy, Harvard Medical School, Boston, USA
Sharon-Lise T Normand *professor of health care policy (biostatistics)*

continued over

BMJ 2005;330:895–7

Department of Health Policy, Management, and Evaluation, Faculty of Medicine, University of Toronto, Toronto, ON, Canada

Geoffrey M Anderson *chair in health management strategies*

Correspondence to: G M Anderson, Institute for Clinical Evaluative Sciences, 2075 Bayview Avenue, Toronto, ON M4N 3M5, Canada
geoff.anderson@utoronto.ca

Table 1 Comparison of cohort studies and randomised controlled trials

Item	Cohort studies	Randomised controlled trials
Populations studied	Diverse populations of patients who are observed in a range of settings	Highly selected populations recruited on the basis of detailed criteria and treated at selected sites
Allocation to the intervention	Based on decisions made by providers or patients	Based on chance and controlled by investigators
Outcomes	Can be defined after the intervention and can include rare or unexpected events	Primary outcomes are determined before patients are entered into study and are focused on predicted benefits and risks
Follow-up	Many cohort studies rely on existing experience (retrospective studies) and can provide an opportunity for long follow-up	Prospective studies; often have short follow-up because of costs and pressure to produce timely evidence
Analysis	Sophisticated multivariate techniques may be required to deal with confounding	Analysis is straightforward

approach is needed that includes the selection of appropriate comparison groups, the identification and assessment of the comparability of potential confounders between those comparison groups, and the use of sophisticated statistical techniques in the analysis.

Comparison groups in cohort studies

The essence of any cohort study is the comparison of outcomes between people who received the intervention and those who did not. For example, to answer the question, "Do patients who receive an atypical antipsychotic drug have an increased risk of hip fracture?" a cohort study must ask: "What would have happened to these patients if they had not received the atypical antipsychotic drug?"

Ideally, the comparison group in the cohort study should be identical to the intervention group, apart from the fact that they did not receive the intervention. This ideal comparison group is described by methodologists as providing the "counterfactual" or "potential outcome."[9] In reality, this ideal comparison group does not exist. Part of the art of designing a cohort study is choosing comparison groups that approach this ideal in order to minimise selection bias while maintaining clinically relevance.

The analysis of the association between antipsychotic drugs and hip fracture can be used to define the types of comparisons that could be found in cohort studies. For any specific intervention (such as exposure to atypical antipsychotics) two factors—the exposure experience of the comparison group and the population from which the intervention and comparison groups are selected—define the types of comparisons that are possible (box 2). People taking atypical antipsychotics can be compared with either people taking an alternative antipsychotic or with those prescribed no antipsychotic drugs. These comparisons could be made in a general population (all elderly people) or in a restricted population (elderly people with dementia).

Questions to ask when assessing a cohort study design

What comparison is being made?
Published studies may include more than one type of comparison, but the focus of any appraisal of a cohort study is on an individual comparison between an intervention group and a comparison group in a defined population. A well written study should contain a clear definition of why the two groups were selected and how they were defined. This information is essential for assessment of clinical relevance and potential for selection bias.

Does the comparison make clinical sense?
The clinical relevance of comparisons needs to be assessed for each case. In the analysis of antipsychotic use and hip fracture, for instance, all four types of comparison might be relevant. However, this might not be true in other analyses. For example, although it would be possible for a cohort study to compare HIV positive patients receiving antiretroviral therapy with those receiving no intervention,[10] this comparison would be irrelevant to many clinicians. A more relevant cohort study would compare patients receiving one antiretroviral therapy with patients receiving another intervention.[11] In contrast, a clinically relevant study of the adverse effects of a commonly used treatment such as a non-steroidal anti-inflammatory drug might include a comparison with a no intervention population since no drug treatment could be a realistic option for some people.[12]

Cohort studies should not only describe the populations being compared but also include a discussion of the clinical context for that comparison and provide a justification for the comparison. Readers of these studies should determine if the study makes a comparison that is realistic and relevant to their decision needs.

Box 1: CONSORT definitions of selection bias and confounding[7]

Selection bias—a systematic error in creating intervention groups, causing them to differ with respect to prognosis. The groups differ in measured or unmeasured baseline characteristics because of the way in which participants were selected for the study or assigned to their study groups

Confounding—a situation in which the estimated intervention effect is biased because of some difference between the comparison groups apart from the planned interventions such as baseline characteristics, prognostic factors, or concomitant interventions. For a factor to be a confounder, it must differ between the comparison groups and predict the outcome of interest

Box 2: Possible types of comparisons in cohort study

General population
1 Intervention *v* alternative intervention
2 Intervention *v* no intervention

Restricted population
3 Intervention *v* alternative intervention
4 Intervention *v* no intervention

Table 2 Effect on age distribution and sample size of restricting comparison of atypical antipsychotic with no intervention to individuals with dementia

	All older people		Older people with dementia	
	Atypical antipsychotic (n=34 960)	No Intervention (n=1 251 435)	Atypical antipsychotic (n=21 427)	No intervention (n=58 754)
Mean (SD) age	80.46 (7.63)	74.50 (6.58)	81.69 (7.11)	80.95 (7.64)
No (%) with dementia	21 427 (61.3)	58 754 (4.7)	21 427 (100)	58 754 (100)

Key questions

What comparison is being made?

Does the comparison make clinical sense?

What are the potential selection biases?

What are the potential selection biases?

Selection bias occurs when there is something inherently different between the groups being compared that could explain differences in the observed outcomes. One powerful strategy to minimise selection bias is to restrict inclusion in the study to those with a defined diagnosis or specific characteristics.[3] Restricting the groups to a specific characteristic removes the potential for bias related to that characteristic and can reduce differences in related characteristics. Table 2 presents data from a cohort of older adults given atypical antipsychotics and a no intervention comparison group. Patients taking atypical antipsychotics were over 12 times more likely (63.1% *v* 4.7%) to have dementia. Dementia is related to the risk of hip fracture, and this imbalance may be an important source of confounding. Restricting the study to people with dementia eliminates this source of confounding and reduces selection related to age as the mean age difference between the groups dropped from years to months.

An inevitable consequence of restriction is reduced sample size. In the example, the sample decreased from 1.3 million to about 80 000 when the dementia restriction was applied. When smaller databases are being used, restriction can greatly limit the power of the study. Restriction on the basis of clinical characteristics limits the generalisability of the findings. The more restrictive the population, the less generalisable the results.

It is important to keep in mind the effect the choice of comparison groups will have on potential selection bias when evaluating a cohort study. Some sources of selection bias are clear—for example, if access to atypical antipsychotics was limited to patients of specialists this could result in patients who received these drugs being different from those who did not. Some sources of bias may be more subtle. For example, if doctors thought that atypical antipsychotics had fewer side effects than typical antipsychotics, they might preferentially use the atypical antipsychotics in frailer patients. This form of selection bias, referred to as channelling bias or confounding by indication,[13] occurs when patients are assigned to one intervention or another on the basis of prognostic factors and is key issue in cohort studies.

Readers should recognise the potential for selection bias in all cohort studies and carefully consider possible sources of bias. In the next article we will outline the link between selection bias and confounding and describe a strategy for identifying and assessing the potential for confounding.

We thank Andreas Laupacis for his comments and Jennifer Gold, Michelle Laxer, and Monica Lee for help in preparing the manuscript.

Contributors and sources: The series is based on discussions that took place at regular meetings of the Canadian Institute for Health Research chronic disease new emerging team. PAR is a geriatrician with extensive research experience in cohort studies of prescription drugs who wrote the first draft of this article and is the guarantor. JHG and MM are clinicians and researchers and SLTN and DLS are statisticians who commented on drafts of this paper. KS programmed and conducted analyses and SG conducted literature searches and reviews. PAR and GMA conceived the idea for the series and GMA worked on drafts of this article and coordinated the development of the series.

Funding: This work was supported by a CIHR operating grant (CIHR No. MOP 53124) and a CIHR chronic disease new emerging team programme (NET-54010).

Competing interests: None declared.

1 Gurwitz JH, Col NF, Avorn J. The exclusion of the elderly and women from clinical trials in acute myocardial infarction. *JAMA* 1992;268:1417-22.
2 Murray MD, Callahan CM. Improving medication use for older adults: an integrated research agenda. *Ann Intern Med* 2003;139:425-9.
3 McKee M, Britton A, Black N, McPherson K, Sanderson C, Bain C. Interpreting the evidence: choosing between randomised and non-randomised studies. *BMJ* 1991;1999:312-5.
4 Benson K, Hartz AJ. A comparison of observational studies and randomized, controlled trials. *N Engl J Med* 2000:342:1878-86.
5 Black N. Why we need observational studies to evaluate the effectiveness of health care. *BMJ* 1996;312:1215-8.
6 Grimes DA, Schulz KF. Bias and causal associations in observational research. *Lancet* 2002;359:248-52.
7 Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med* 2001;134:663-94.
8 Altman DG, Bland JM. Treatment allocation in controlled trials: why randomize. *BMJ* 1999;318:1209.
9 Greenland S, Morgenstern H. Confounding in health research. *Annu Rev Public Health* 2001;22:189-212.
10 Badri M, Wilson D, Wood R. Effect of highly active antiretroviral therapy on the incidence of tuberculosis in South Africa: a cohort study. *Lancet* 2002;359:2059-64.
11 Fellay J, Boubaker K, Ledergerber B, Bernasconi E, Furrer H, Battegay M. Prevalence of adverse events associated with potent antiretroviral treatment: Swiss HIV cohort study. *Lancet* 2001;358:1322-7.
12 Mamdani M, Rochon PA, Juurlink DN, Kopp A, Anderson GM, Naglie G, et al. Observational study of upper gastrointestinal haemorrhage in elderly patients given selective cyclo-oxygenase-2 inhibitors or conventional non-steroidal anti-inflammatory drugs. *BMJ* 2002;325:1-6.
13 Psaty BM, Koepsell TD, Lin D, Weiss NS, Siscovick DS, Rosendaal FR, et al. Assessment and control for confounding by indication in observational studies. *J Am Geriatr Soc* 1999;47:749-54.

(Accepted 18 February 2005)

Endpiece

Good advice

Better to hunt in fields, for health unbought,
Than fee the doctor for a nauseous draught.
The wise, for cure, on exercise depend;
God never made his work for man to mend.

John Dryden (1631-1700) in Epistle to John Driden of Chesterton (1700)

Fred Charatan, retired geriatric physician, Florida

Trainee Participation Is Associated With Adverse Outcomes in Emergency General Surgery

An Analysis of the National Surgical Quality Improvement Program Database

George Kasotakis, MD, MPH, Aliya Lakha, BS,† Beda Sarkar, MD, PhD,* Hiroko Kunitake, MD, MPH,* Nicole Kissane-Lee, MD, EdM,* Tracey Dechert, MD,* David McAneny, MD,* Peter Burke, MD,* and Gerard Doherty, MD**

Objective: To identify whether resident involvement affects clinically relevant outcomes in emergency general surgery.

Background: Previous research has demonstrated a significant impact of trainee participation on outcomes in a broad surgical patient population.

Methods: We identified 141,010 patients who underwent emergency general surgery procedures in the 2005–2010 Surgeons National Surgical Quality Improvement Program database. Because of the nonrandom assignment of complex cases to resident participation, patients were matched (1:1) on known risk factors [age, sex, inpatient status, preexisting comorbidities (obesity, diabetes, smoking, alcohol, steroid use, coronary artery disease, chronic renal failure, pulmonary disease)] and preoperatively calculated probability for morbidity and mortality. Clinically relevant outcomes were compared with a t or χ^2 test. The impact of resident participation on outcomes was assessed with multivariable regression modeling, adjusting for risk factors and operative time.

Results: The most common procedures in the matched cohort (n = 83,790) were appendectomy (39.9%), exploratory laparotomy (8.8%), and adhesiolysis (6.6%). Trainee participation is independently associated with intra- and postoperative events, wound, pulmonary, and venous thromboembolic complications, and urinary tract infections.

Conclusions: Trainee participation is associated with adverse outcomes in emergency general surgery procedures.

Keywords: emergency general surgery, graduate medical education, National Surgical Quality Improvement Program, operative time, outcomes, residents, supervision, trainees

(*Ann Surg* 2014;260:483–493)

Modern graduate medical education is called to balance 2 important needs: provide affordable, high-quality health care to patients, while simultaneously training junior physicians in their specialty of choice. Previous research has demonstrated both beneficial and adverse effects of trainee participation in all levels and fields of modern medicine including, but not limited to, primary care,[1] cardiology,[2,3] radiology,[4] gastrointestinal medicine,[5–7] and surgery.[8–20] Sandhu et al[3] demonstrated that in patients who underwent percutaneous coronary intervention, there were no differences in mortality, inpatient myocardial infarction, or contrast-induced nephropathy between cases performed at major teaching hospitals and those performed in nonteaching ones. This study revealed both the positive and adverse impacts of trainee involvement: percutaneous coronary intervention at teaching hospitals was associated with a decreased risk of emergency coronary artery bypass grafting compared with nonteaching hospitals but at the expense of a higher complication rate.[3] Highlighting the beneficial effect of trainee involvement, Polanczyk and colleagues[2] found lower mortality in patients with myocardial infarction, heart failure, or stroke at teaching institutions, and Buchner et al[6] demonstrated increased colon malignancy detection during endoscopies performed with resident participation.

Contrary to these reports, however, a significant body of evidence suggests that trainee participation may be associated with adverse outcomes. In a study investigating the diagnostic competency of radiology residents, the majority of junior trainees misidentified prominent trauma abnormalities on imaging,[4] and in primary care, adverse outcomes have been linked to more prescribing errors made by junior trainees than their senior counterparts.[1]

Among the most commonly cited risk factors for adverse outcomes in graduate medical education are inexperience or excessive autonomy. These aspects may be even more pronounced in surgical specialty training programs, where trainees' lack of sophisticated technical skills, experience, and familiarity with complex procedures and evolving surgical technologies may play important roles.[17,20] Papandria and associates[18] reported that even common surgical procedures (such as laparoscopic appendectomy, laparoscopic cholecystectomy, and open hernia repair) required longer operative times with trainee involvement,[18] yet the "resident effect" on outcomes has been shown to be minor in a broad surgical patient population undergoing mostly elective procedures.[21]

One particular population in which the effect of surgical resident participation has yet to be clearly defined is the one requiring *emergency* surgical procedures. The unique characteristic of this specific type of surgical intervention is that time is of the essence, and surgical trainees are available around the clock at teaching institutions to expedite operative intervention, until a board-certified surgeon is called to guide and supervise operative management. However, involving residents in emergency surgical cases may unnecessarily prolong operative time, with a possible negative effect on overall outcomes. With the current project, we aim to identify whether trainee participation is associated with clinically relevant outcomes in patients undergoing emergency general surgical procedures.

METHODS

Data Source and Study Population

Data were extracted from the nationwide American College of Surgeons National Surgical Quality Improvement Program (ACS-NSQIP) database, after excluding patients who underwent nongeneral surgery procedures. Out of all adults (16 years of age and older) enrolled at any of the approximately 400 participating institutions

From the Departments of *Surgery; and †Graduate Medical Sciences, Boston University School of Medicine, Boston, MA.

Disclosure: Project funded by the Department of Surgery at Boston University School of Medicine

The authors declare no conflicts of interest.

Reprints: George Kasotakis, MD, MPH, Boston University School of Medicine, Boston Medical Center, Section of Trauma & Acute Care Surgery, 840 Harrison Ave, Dowling 2 South, Ste #2414, Boston, MA 02118. E-mail: gkasot@bu.edu.

ISSN: 0003-4932/14/26003-0483

DOI: 10.1097/SLA.0000000000000889

in North America between 2005 and 2010, those who underwent an emergency general surgical procedure comprised our *Aggregate Cohort*.

Variables

Variables included baseline demographic (age, sex, body mass index) and clinical characteristics (inpatient status, laboratories, and preexisting comorbid conditions that are known risk factors for adverse surgical outcomes),[22–24] such as obesity (a body mass index of >30); diabetes mellitus (requiring daily oral hypoglycemic agents or insulin); smoking (any history of smoking in the year before index procedure); alcohol (defined as >2 drinks per day in the 2 weeks preceding admission) or steroid use (for a chronic condition); coronary artery disease (defined as anyone with a diagnosis of congestive heart failure or history of angina in the 30 days preceding surgery, myocardial infarction in the 6 months before index operation, or any history of percutaneous coronary intervention or cardiac surgery); renal failure (acute or chronic, requiring dialysis or not); and pulmonary disease (chronic obstructive pulmonary disease or asthma). In addition, each patient's expected probability for morbidity and mortality was obtained, as calculated by the ACS-NSQIP surgical risk calculator.[25]

Procedure-related information was also extracted, such as the procedure performed by current procedural terminology code, whether or not a resident was involved (and, if so, his or her postgraduate level), total anesthesia and operative time, ASA class, and whether or not the patient required an intraoperative transfusion, suffered a postoperative bleed requiring transfusion, or required an unplanned return to the operating room (OR). The level of resident supervision was left out from our analysis, as less than 0.1% of total cases were performed without a staff surgeon present in the OR.

Additional outcome variables included clinically relevant global outcomes (mortality and hospital length of stay) and 30-day postoperative complications, pertaining to wound, pulmonary, cardiovascular, renal, and life-threatening infectious complications, per the standardized ACS-NSQIP definitions.[26]

Matching and Data Analysis

Because patient's assignment to undergo an emergency general surgery procedure with or without trainee participation was not random, a coarsened exact matching algorithm[27] was used to match patients with a 1:1 ratio, on the basis of age; sex; inpatient status; morbid obesity; history of diabetes; smoking, alcohol, or steroid use; and preexisting coronary artery disease, renal failure, and pulmonary disease. As it is not possible to match subjects on the basis of exact procedure performed without significantly compromising sample size, and given the inherent greater complexity of cases managed at teaching institutions, subjects were also matched on their expected probability for morbidity and mortality (risk adjustment) per the validated ACS-NSQIP model. The 2 balanced patient groups obtained after application of the coarsened exact matching algorithm comprised the *Matched Cohort*. Absolute and relative frequencies of the various emergency general surgical procedures grouped by type were calculated in both the aggregate and matched cohorts. Depending on data distribution, continuous variables were compared with Pearson *t* test or Wilcoxon test and binary variables with the χ^2 test. Continuous variables are reported as means ± standard deviations if the data were normally distributed, or as medians (25%, 75% percentiles) if otherwise. Binary variables are reported as percentages.

To *quantify* the effect of resident's participation on the statistically significant outcomes, as identified on the aforementioned analysis, multivariable regression analysis models were fitted on the *matched cohort*, adjusting for the said risk factors *and* total operative time. The effect of total operative time is also reported. For the logistic regression models, postestimation Pearson goodness-of-fit tests were obtained and C-statistics were calculated. For the linear regression models, residuals were assessed for outliers and influential values, using leverage and standardized residuals. Observations with outlying and highly influential residuals were left in our sample as true outliers, if data collection was deemed appropriate and correct. Univariate graph plots were obtained to assess for an association between postgraduate level of trainee participation and case complexity/illness severity (as reflected on the preoperatively calculated expected probability for morbidity) and total operative time. Postgraduate years of 8 or more were combined into 1 group. All statistical analyses were performed using Stata 13.1 (StataCorp LP, College Station, TX). Statistical significance was declared at a 2-sided *P* value of 0.05 or less.

RESULTS

Of the 957,813 patients tracked by the ACS-NSQIP participating institutions, a total of 141,010 underwent emergency general surgical procedures between 2005 and 2010 and comprised our *aggregate cohort*. After matching, 2 equal-sized groups of 41,895 patients each formed our *matched cohort.* The baseline characteristics of all cohorts are summarized in Table 1.

Patients who underwent procedures with resident involvement were more likely to be inpatient (89% vs 85.45%; $P < 0.001$) and, not surprisingly, had more comorbidities at baseline. This is expected, as sicker patients or those requiring more complex procedures are more likely to be referred to teaching institutions.

Table 2 summarizes the procedures the *aggregate* and *matched cohorts* underwent, and Table 3 summarizes the discharge diagnoses between the 2 cohorts. The most common procedures in *aggregate cohort* (n = 141,010) were appendectomies (29.8%), exploratory laparotomies (12%), and colon resections (7.3%). The most common in the *matched cohort* (n = 83,790) were appendectomies (39.9%), exploratory laparotomies (8.8%), lysis of adhesions (6.6%), and cholecystectomies (6.5%). There is an overall balanced distribution of more complex cases (laparotomies, colectomies, small bowel resections, and abdominal wall hernia repairs), with simpler ones (lysis of adhesions, diagnostic laparoscopies, and inguinal herniorrhaphies) across the 2 groups in the *matched cohort*. This is not surprising, as the 2 groups were matched on their expected probability for mortality and overall morbidity, which takes into account not only baseline comorbid conditions but also complexity of the procedure(s) performed (by current procedural terminology code). Similarly, there is a balanced distribution of discharge diagnoses across the resident and no-resident groups in the *matched cohort* (Table 3): The resident group managed slightly fewer uncomplicated appendicitides (41.52% vs 44.44%), acute cholecystitides (3.06% vs 5.30%), diverticulitides (1.69% vs 2.22%), and gastrointestinal tract malignancies (1.46% vs 1.62%), but more operative bowel obstructions (8.78% vs 7.71%), hernias with (6.27% vs 5.64%) and without bowel obstruction (1.57% vs 0.97%), soft tissue abscesses (3.01% vs 2.85%), and intestinal ischemia (1.17% vs 0.95%).

Clinically relevant 30-day postoperative outcomes are listed in Table 4. As expected, due to the higher incidence of preexisting comorbid conditions and overall more complex cases typically managed at teaching institutions, hospital stay was longer (8.36 ± 14.8 vs 6.50 ± 11.2 days; $P < 0.001$) and 30-day mortality was higher (12.45% vs 8.74%; $P < 0.001$) in the *aggregate cohort*. Also, cases performed with resident participation took almost 20 minutes longer to complete (83.30 ± 60.72 vs 64.52 ± 49.4; $P < 0.001$), and total anesthesia time was longer by almost 27 minutes (134.51 ± 75.15 vs 107.54 ± 61.08; $P < 0.001$), likely due to involvement of anesthesia trainees, who may require longer times to secure the airway

TABLE 1. Baseline Demographic and Clinical Characteristics in the Aggregate and Matched Cohorts

	Aggregate Cohort (n = 141,010)			Matched Cohort (n = 83,790)		
	NO RES N = 52,059	**RES N = 88,951**	***P***	**No RES N = 41,895**	**RES N = 41,895**	***P***
Age, yr	49.84 ± 19.67	49.33 ± 19.44	**<0.001**	46.44 ± 18.87	46.43 ± 18.87	0.999
Sex (female)	51.94	50.89	**<0.001**	52.75	52.75	1.000
Inpatient	85.45	89.00	**<0.001**	84.01	84.01	1.000
BMI ≥30	32.77	32.04	**0.007**	31.34	31.34	1.000
Diabetes	10.91	11.51	**0.001**	6.35	6.35	1.000
Smoking	21.90	23.08	**<0.001**	20.08	20.08	1.000
Alcohol use	3.05	3.44	**<0.001**	1.67	1.67	1.000
Steroid use	3.52	4.57	**<0.001**	0.89	0.89	1.000
Coronary artery disease	7.97	8.49	**0.001**	3.44	3.44	1.000
Renal failure	3.25	4.21	**<0.001**	0.26	0.26	1.000
Pulmonary disease	11.83	13.21	**<0.001**	4.38	4.38	1.000
Expected probability for morbidity	0.18 ± 0.19	0.21 ± 0.21	**<0.001**	0.12 ± 0.13	0.12 ± 0.13	0.725
Expected probability for mortality	0.05 ± 0.13	0.06 ± 0.15	**<0.001**	0.016 ± 0.055	0.017 ± 0.055	0.103

No RES indicates without resident participation; RES, with resident participation. Bold values represent statistically significant.

TABLE 2. Procedure Types Performed in the Aggregate and Matched Cohorts Across the 2 Study Groups: RES and No RES

	Aggregate Cohort		Matched Cohort	
	No RES N = 137,955	**RES N = 77,076**	**No RES N = 58,824**	**RES N = 58,114**
Appendectomy	27.9%	33.2%	38.4%	41.4%
Exploratory laparotomy	13.2%	9.8%	9.5%	8.0%
Colon resection	7.5%	7.1%	4.8%	5.2%
Lysis of adhesions	6.8%	6.8%	6.7%	6.5%
Cholecystectomy	5.4%	5.9%	6.7%	6.4%
Small bowel resection	5.4%	4.3%	4.3%	4.1%
Abdominal wall hernia repair	4.9%	4.5%	5.5%	5.1%
Incision and drainage/debridement	3.4%	2.6%	3.0%	3.4%
Repair of perforated viscus	2.4%	1.8%	1.6%	1.5%
Diagnostic laparoscopy	1.4%	1.8%	1.5%	2.2%
Inguinal hernia repair	1.4%	1.4%	1.5%	1.5%
Other	20.3%	21.0%	16.4%	14.8%

N represents the number of procedures by current procedural terminology code performed in each group.
No RES indicates without resident participation; RES, with resident participation.

and establish and reverse anesthesia. Because of the greater number of preexisting comorbidities and increased procedure complexity in the cases performed with resident participation, it is not surprising that this group had greater complication rates [with the exception of wound dehiscence (1.06% vs 1.14%; $P = 0.164$) and postoperative acute coronary events (0.57% vs 0.50%; $P = 0.088$)].

After matching and eliminating the influence of patients' preexisting comorbidities and case complexity, no mortality difference was noted (3.25% vs 2.96%; $P = 0.085$). Total operative and anesthesia times were still longer if trainees were involved (75.10 ± 54.77 vs 59.17 ± 44.78 minutes; $P < 0.001$ and 122.42 ± 66.22 vs 99.92 ± 55.45 minutes; $P < 0.001$, respectively). Moreover, cases performed by residents required more intraoperative transfusions (3.43% vs 2.55%; $P < 0.001$) and unplanned returns to the OR (4.22% vs 3.80%; $P = 0.002$). Interestingly, cases performed by residents had a lower incidence of postoperative transfusion requirement (1.12% vs 1.28%; $P = 0.031$), possibly related to the fact that they received more blood products intraoperatively.

Regarding wound complications, cases performed with residents had a higher incidence of both superficial (SSSI) (3.50% vs 2.78%; $P < 0.001$) and organ space surgical site infections (OSSI) (2.27% vs 1.77%; $P < 0.001$) after surgery. However, there was no statistically significant difference in wound dehiscence between the 2 groups.

Patients whose procedures were performed with trainee involvement developed more pulmonary complications postoperatively, including pneumonia (1.85% vs 1.67%; $P = 0.043$), unplanned reintubation (1.64% vs 1.15%; $P < 0.001$), and prolonged mechanical ventilation (2.87% vs 2.06%; $P < 0.001$). Resident involvement did not seem to affect major cardiac events [such as myocardial infarction (0.27% vs 0.26%; $P = 0.637$) and cardiopulmonary arrest (0.39% vs 0.32%; $P = 0.071$)], cerebrovascular events (0.16% vs 0.13%; $P = 0.205$), or renal complications [including acute renal failure (0.34% vs 0.31%; $P = 0.427$) and renal failure requiring dialysis (0.37% vs 0.43%; $P = 0.209$)]. However, cases performed with trainees had a higher incidence of venous thromboembolic events [deep

TABLE 3. Discharge Diagnoses by *International Classification of Diseases, Ninth Revision* (*ICD-9*) Code in the Aggregate and Matched Cohorts Across the 2 Study Groups: RES and No RES

	Aggregate Cohort		Matched Cohort	
	No RES N = 52,059	RES N = 88,951	No RES N = 41,895	RES N = 41,895
Acute uncomplicated appendicitis	35.70%	32.49%	44.44%	41.52%
Acute complicated appendicitis	10.89%	8.79%	10.09%	10.19%
Bowel obstruction	8.28%	9.37%	7.71%	8.78%
Hernia w/ obstruction	5.84%	5.88%	5.64%	6.27%
Acute cholecystitis	5.18%	4.94%	5.30%	3.06%
Cellulitis or abscess	3.14%	3.51%	2.85%	3.01%
Gastrointestinal tract perforation	4.21%	5.16%	2.95%	2.91%
Symptomatic cholelithiasis/ chronic cholecystitis	2.25%	2.21%	2.43%	2.69%
Diverticulitis	2.62%	2.28%	2.22%	1.69%
Hernia w/o Obstruction	0.95%	1.32%	0.97%	1.57%
Gastrointestinal tract malignancies	1.98%	2.00%	1.62%	1.46%
Intestinal ischemia	2.11%	3.00%	0.95%	1.17%
Gastrointestinal tract hemorrhage	1.11%	1.16%	0.58%	0.44%
Pancreatic disease	0.47%	0.74%	0.33%	0.38%
Other	15.29%	17.14%	11.92%	14.85%

No RES indicates without resident participation; RES, with resident participation.

TABLE 4. Clinically Relevant 30-Day Postoperative Outcomes

	Aggregate Cohort (n = 141,010)			Matched Cohort (n = 83,790)		
	No RES N = 47,178	RES N = 61,353	*P*	No RES N = 41,895	RES N = 41,895	*P*
Mortality	8.74	12.45	**<0.001**	2.96	3.25	0.082
Length of stay	6.50 ± 11.2	8.36 ± 14.8	**<0.001**	4.59 ± 7.98	4.97 ± 9.63	**0.019**
Operative technique						
Operative time	64.52 ± 49.41	83.30 ± 60.72	**<0.001**	59.17 ± 44.78	75.10 ± 54.77	**<0.001**
Anesthesia time	107.54 ± 61.08	134.51 ± 75.15	**<0.001**	99.92 ± 55.45	122.42 ± 66.22	**<0.001**
Intraoperative transfusion	5.64	9.33	**<0.001**	2.55	3.43	**<0.001**
Postoperative bleed requiring transfusion	2.45	3.16	**<0.001**	1.28	1.12	**0.031**
Unplanned return to the operating room	5.66	8.54	**<0.001**	3.80	4.22	**0.002**
Wound complications						
Superficial wound infection	3.13	4.09	**<0.001**	2.78	3.50	**<0.001**
Deep wound infection	1.14	1.06	0.164	0.89	0.73	**0.011**
Organ space infection	2.20	3.27	**<0.001**	1.77	2.27	**<0.001**
Wound dehiscence	1.08	1.27	**0.002**	0.69	0.63	0.266
Pulmonary complications						
Postoperative pneumonia	3.20	4.23	**<0.001**	1.67	1.85	**0.043**
Unplanned reintubation	2.33	3.64	**<0.001**	1.15	1.64	**<0.001**
Prolonged ventilation (>48 hr)	5.82	9.47	**<0.001**	2.06	2.87	**<0.001**
Cardiovascular complications						
Myocardial infarction	0.50	0.57	0.088	0.26	0.27	0.637
Cardiopulmonary arrest	0.80	1.21	**<0.001**	0.32	0.39	0.071
Cerebrovascular Accident	0.28	0.37	**0.006**	0.13	0.16	0.205
Deep venous thrombosis	1.06	1.56	**<0.001**	0.62	0.80	**0.002**
Pulmonary embolism	0.37	0.56	**<0.001**	0.28	0.43	**<0.001**
Renal complications						
Acute renal failure	0.61	0.71	**0.029**	0.31	0.34	0.427
Renal failure requiring dialysis	1.17	1.46	**<0.001**	0.43	0.37	0.209
Urinary tract infection	1.78	2.36	**<0.001**	1.14	1.45	**<0.001**
Life-threatening infectious complications						
Sepsis	2.98	4.13	**<0.001**	2.13	2.42	**0.005**
Septic shock	2.95	3.88	**<0.001**	1.41	1.51	0.205

No RES indicates without resident participation; RES, with resident participation. Bold values represent statistically significant.

TABLE 5. *Matched Cohort* Analysis: Effect of Trainee Participation and Operative Time on Outcomes (Adjusting for Risk Factors and Operative Time)

	Effect of Trainee Participation on Outcomes		Effect of Operative Time on Outcomes (per 30 min)	
	Effect Estimate/Odds Ratio (95% CI)	***P***	**Effect Estimate/Odds Ratio (95% CI)**	***P***
Length of stay	0.07 (−0.04 to 0.17)	0.242	0.62 (0.58–0.65)	**<0.001**
Superficial wound infection	1.23 (1.13–1.34)	**<0.001**	1.10 (1.09–1.12)	**<0.001**
Organ space infection	1.21 (1.09–1.34)	**<0.001**	1.10 (1.08–1.13)	**<0.001**
Intraoperative transfusion	1.20 (1.07–1.34)	**0.001**	1.30 (1.27–1.32)	**<0.001**
Postoperative bleeding requiring transfusion	0.78 (0.69–0.90)	**<0.001**	1.15 (1.12–1.17)	**<0.001**
Unplanned return to the operating room	1.08 (1.00–1.16)	**0.041**	1.06 (1.04–1.08)	**<0.001**
Postoperative pneumonia	1.06 (0.94–1.18)	**0.348**	1.08 (1.06–1.11)	**<0.001**
Unplanned reintubation	1.38 (1.21–1.57)	**<0.001**	1.07 (1.04–1.09)	**<0.001**
Prolonged mechanical ventilation (>48 hr)	1.43 (1.29–1.59)	**<0.001**	1.13 (1.11–1.15)	**<0.001**
Deep venous thrombosis	1.25 (1.05–1.49)	**0.011**	1.07 (1.04–1.11)	**<0.001**
Pulmonary embolism	1.42 (1.11–1.81)	**0.005**	1.09 (1.05–1.13)	**<0.001**
Urinary tract infections	1.23 (1.08–1.40)	**0.001**	1.08 (1.06–1.11)	**<0.001**
Sepsis	1.07 (0.97–1.18)	0.155	1.1 (1.08–1.12)	**<0.001**

Bold values represent statistically significant.

venous thrombosis (0.80% vs 0.62%; $P = 0.002$), pulmonary embolism (0.43% vs 0.28%; $P < 0.001$)], and urinary tract infections (1.45% vs 1.14%; $P < 0.001$), possibly related to the longer total operative time. A greater rate of septic complications was noted in the resident group also (2.42% vs 2.13%; $P = 0.005$), likely due to the greater number of infectious events.

The quantified effect of trainee participation on clinically relevant outcomes (adjusting for aforementioned risk factors, expected probability for morbidity and mortality, and operative time) is summarized in Table 5. The same table demonstrates the independent effect of every 30 minutes added to the total operative time, controlling for risk factors, expected probability for morbidity and mortality, and resident participation. Model fit was highly satisfactory for all logistic regression models, and C-statistics ranged from 0.80 to 0.93. No error in data entry was obvious in the influential or outlying observations in the linear regression model for the length of stay.

When adjusting for the aforementioned risk factors *and* operative time, the resident's involvement was not independently associated with hospital length of stay (β coefficient: 0.07; $P = 0.242$), but it was found to be an independent predictor for most adverse postoperative outcomes, except for postoperative transfusion requirements (possibly due to greater blood product transfusion intraoperatively), in which trainee participation was protective [odds ratio (OR): 0.78 (0.69–090), $P < 0.001$], and pneumonia and sepsis events, for which trainee participation exerted no effect [O.R: 1.06 (0.94–1.18); $P = 0.348$ and 1.07 (0.97–1.18); $P = 0.155$, respectively). Resident-assisted cases were 23% more likely to develop an SSSI ($P < 0.001$), 21% more likely to develop an OSSI ($P < 0.001$), and 20% more likely to require an intraoperative transfusion ($P = 0.001$), independent of total operative time. Cases performed by residents were 8% more likely to require unplanned reoperation ($P = 0.041$), 38% ($P < 0.001$) more likely to necessitate unplanned reintubation, 43% more likely to experience failure to wean off mechanical ventilation ($P < 0.001$), 25% more likely to develop deep venous thrombosis ($P = 0.011$), 42% more likely to develop a pulmonary embolism ($P = 0.005$), and 23% more likely to develop postoperative urinary tract infections ($P = 0.001$).

Similarly, longer operative times seemed to exert a statistically significant effect on outcomes in our secondary analysis, independent of trainee participation or known risk factors: For every 30 minutes a case lasted longer, hospital stay was prolonged by 0.62 days [β coefficient: 0.62 (0.58–0.65), $P < 0.001$). Longer operative times were also associated with a greater incidence of SSSI and OSSI: For every additional half-hour in total operative time, incidence of SSSI and OSSI increased by 10% each ($P < 0.001$), and so did intra- (by 30%, $P < 0.001$) and postoperative (15%, $P < 0.001$) transfusion requirements and unplanned take backs to the operating theater (6%, $P < 0.001$). Pulmonary and thromboembolic complications also increased with added operative time because of the increased length of immobility and the likely inadequacy of extremity compression devices. Finally, urinary tract infections and sepsis incidence were positively influenced by longer operative times [OR: 1.08 (1.06–1.11); $P < 0.001$, and OR: 1.10 (1.08–1.12); $P < 0.001$, respectively, per 30 minutes of operative time).

Evaluating the effect of postgraduate year of training to the total operative time, it seems that senior residents and fellows tend to require more time in the OR (Fig. 1), but this effect may be due to self-selection for participation in more complex surgical cases (Figs. 2 and 3). Attendings are also more likely to perform the majority of the operation, when junior residents scrub in complex cases. The 3 graphs also demonstrate that staff surgeons can perform fairly complex surgical procedures in less time than typically required when trainees participate.

DISCUSSION

Our findings suggest that trainee participation in emergency general surgery procedures is associated with longer operative time and more intraoperative transfusions and is independently associated with adverse postoperative outcomes, including wound, pulmonary and venous thromboembolic complications, and urinary tract infections. Although some residual, unmeasured confounding cannot be completely excluded, as no information on level of supervision is available, it seems that this effect is independent of total operative time, case complexity, and preexisting comorbidities. Operative time

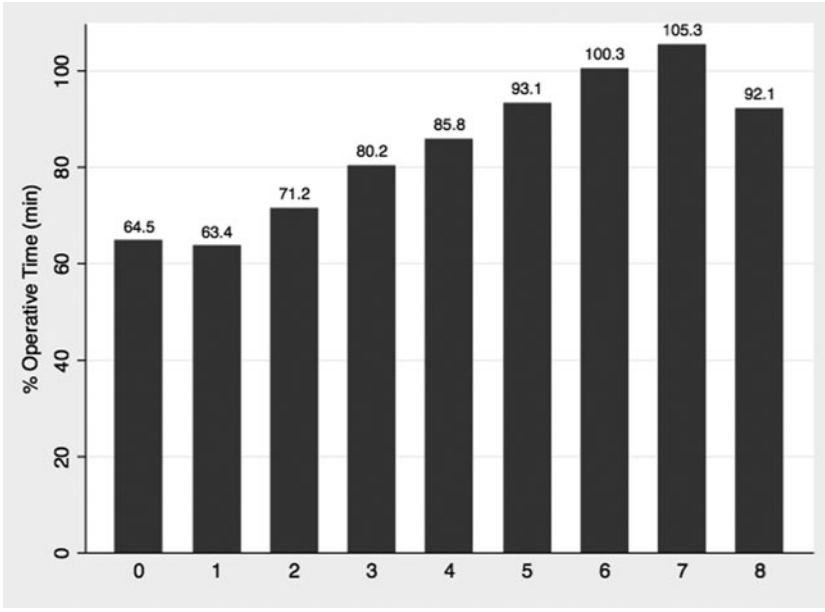

FIGURE 1. Operative time (minutes) versus postgraduate year (year 0 corresponds to cases performed by attending surgeons alone).

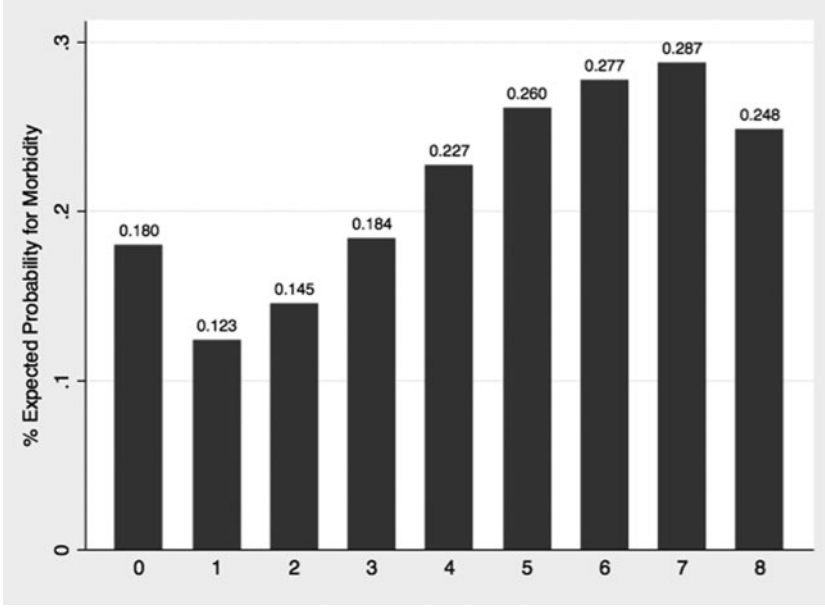

FIGURE 2. Expected probability of morbidity versus postgraduate year (year 0 corresponds to cases performed by attending surgeons alone).

is identified as an another important factor independently associated with intra- and postoperative transfusions; unplanned reoperations; wound, pulmonary, infectious, and venous thromboembolic complications; and longer hospital stays, adjusting for baseline comorbidities, case complexity, and resident participation.

Our findings are similar to several other studies that have shown a similar effect of trainee participation in general surgery: Krell and associates[28] demonstrated that resident involvement was independently associated with postoperative wound complications and venous thromboembolism in patients undergoing laparoscopic gastric bypass, after adjusting for risk factors and surgeon and hospital case volume. They posit that the higher incidence of these complications is likely due to longer operative times. Our findings are similar, although trainee participation was independently associated with postoperative morbidity, even when operative time was controlled for. We additionally demonstrate that operative time itself was independently associated with numerous complications, even when adjusting for trainee participation.

Scarborough et al[19] similarly showed an increased incidence of serious and overall postoperative complications when appendectomies were performed with resident participation, adjusting for baseline comorbidities and total operative time. However, they demonstrated that operative times and risk for postoperative morbidity increased with postgraduate year of training and attributed this effect to the increased level of autonomy senior residents enjoy in the operating theater. However, they did not risk-adjust these 2 outcomes in their analysis for such a conclusion to be definitively drawn. This is in contrast to our findings, which demonstrate that there is indeed a greater incidence of postoperative morbidity in cases performed with senior resident/fellow involvement. However,

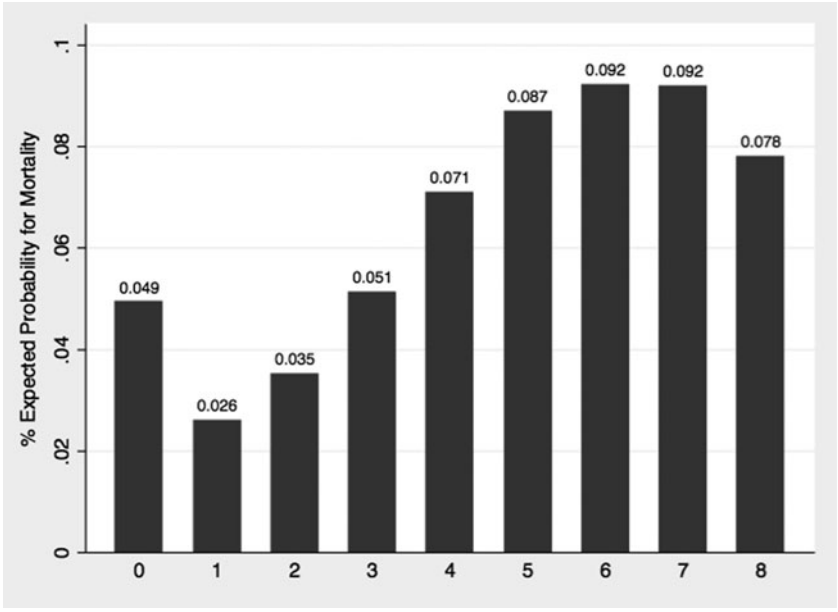

FIGURE 3. Expected probability of mortality versus postgraduate year (year 0 corresponds to cases performed by attending surgeons alone).

this is likely due to the fact that more complex cases are assisted by more senior trainees. Similarly, Wilkiemeyer and colleagues[29] examined the relationship between various levels of resident level and postoperative outcomes in hernia repairs, reporting an increased recurrence rate in the cases performed by junior residents, compared with those performed by their senior counterparts.

Contrasting our findings is the work of Kiran et al,[21] who, following similar methodology as ours, concluded that resident's involvement in surgical procedures is safe, citing that a modest increase in SSSI was likely due to longer operative times. However, that study had a smaller matched sample size and included both elective and emergency cases, rendering comparisons difficult. It is possible that time spent in the OR in an elective setting may not be as significant determinant of postoperative outcomes as in emergency procedures, in which host physiology may be appreciably deranged.

Similarly, Kazaure et al[30] reported no trainee effect in risk of overall complications or hospital length of stay and only a minimal increase in total OR time. However, they grouped all complications in a single binary variable, assessed the effect of trainees who operated alone on only 3 procedures (appendectomies, cholecystectomies, and herniorrhaphies), and included both elective and emergency cases. Finally, their article did not state what variables they controlled for in their multivariable regression model (except for case complexity, using relative value units, and not the more robust calculated risk for morbidity and mortality), making comparisons difficult, if not impossible. Itani and colleagues[31] concluded that the degree of resident's supervision did not affect outcomes; however, in their analysis, they compared cases performed by residents and an attending immediately available against cases performed by staff surgeons alone or with residents participating, rendering it difficult to infer the "trainee effect."

The findings of the present work should be interpreted with caution and in the context of the study design. Major strengths of our analysis lie in our large sample size, matching of the 2 groups on numerous comorbidities known to adversely affect postoperative outcomes, and inclusion of the expected probability for morbidity and mortality, which balances the effect of host risk factors *and* case complexity. However, this is a secondary analysis of prospectively collected data, not originally intended to test our specific hypothesis. In addition, the data, while highly reliable, are derived from ACS-NSQIP–participating institutions and thus are skewed toward larger, tertiary medical centers that perform more complex emergency general surgery procedures, whereas smaller hospitals may be underrepresented.[32] It is also important to note that data are collected for a maximum of 30 days postoperatively, which may underrepresent true surgical-related mortality.[33] Additional considerations include resident's autonomy, which may be greater in fairly simple emergency procedures (in an appendectomy or abscess drainage, for example), as opposed to elective cases that require more advanced technical skills (such as a Whipple or a gastric bypass), and the fact that the ACS-NSQIP database does not report the extent of trainee participation or attending supervision. Such information could potentially improve our understanding of the relationship between resident's involvement and postoperative outcomes and account for residual confounding, although previous research has not demonstrated differences with varying degrees of trainee supervision.[30,31] Finally, the degree of resident's autonomy in postoperative patient care—not tracked by the ACS-NSQIP—likely also influences outcomes.

Our findings underline several important aspects of trainee education and its impact on patient outcomes in emergency general surgical procedures. Resident training and involvement is crucial to the sustainability of any health care system, as the future of health care is highly dependent on our ability to provide high-quality postgraduate medical education. Nonetheless, it is the responsibility of training programs to ensure that patient's safety is maintained with appropriate trainee supervision. The small increase in morbidity associated with resident involvement indicates that appropriate supervision should be a priority, and greater emphasis should be placed on surgical emergency simulation training to improve efficiency in and out of the OR.[34,35]

Improving the channels of communication between trainees and staff surgeons should also be prioritized during surgical emergencies. Residents should be well prepared for emergency procedures through simulation training and aim to maximize their efficiency in the OR. In addition, time objectives for the resident-assigned operative tasks should be established early in operative management, with planned "takeovers" by the attending physician, if these time objectives are not met.[28] Such a strategy could not only allow

effective hands-on training time for residents but also minimize inefficient use of operative time.

CONCLUSIONS

Resident involvement in emergency general surgical procedures seems to be independently associated with longer operative times and both intraoperative and postoperative complications. This association is independent of total operative time, case complexity, and preexisting comorbid conditions. Operative time also seems to be an independent risk factor for postoperative morbidity. Surgeons practicing at teaching institutions should be aware of this association, provide appropriate supervision as needed, and minimize unnecessary operative time. Enhancing residents' preparedness for emergency operations, perhaps through simulation, may also aid in limiting operative times and thus minimizing adverse postoperative outcomes.

ACKNOWLEDGMENTS

The authors thank the staff of the Clinical Quality and Patient Safety Office at Boston University, and especially Pam Rosenkranz, for their tireless support. All authors read and approved the article.

REFERENCES

1. Ryan C, Ross S, Davey P, et al. Prevalence and causes of prescribing errors: the PRescribing Outcomes for Trainee Doctors Engaged in Clinical Training (PROTECT) study. *PLoS One*. 2014;9:e79802.
2. Polanczyk CA, Lane A, Coburn M, Philbin EF, Dec GW, DiSalvo TG. Hospital outcomes in major teaching, minor teaching, and nonteaching hospitals in New York state. *Am J Med*. 2002;112:255–261.
3. Sandhu A, Moscucci M, Dixon S, et al. Differences in the outcome of patients undergoing percutaneous coronary interventions at teaching versus nonteaching hospitals. *Am Heart J*. 2013;166:401–408.
4. McLauchlan CA, Jones K, Guly HR. Interpretation of trauma radiographs by junior doctors in accident and emergency departments: a cause for concern? *J Accident Emerg Med*. 1997;14:295–298.
5. Hope WW, Hooks WB, III, Kilbourne SN, et al. Assessing resident performance and training of colonoscopy in a general surgery training program. *Surg Endosc*. 2013;27:1706–1710.
6. Buchner AM, Shahid MW, Heckman MG, et al. Trainee participation is associated with increased small adenoma detection. *Gastrointest Endosc*. 2011;73:1223–1231.
7. Eckardt AJ, Swales C, Bhattacharya K, et al. Does trainee participation during colonoscopy affect adenoma detection rates? *Dis Colon Rectum*. 2009;52:1337–1344.
8. Fahrner R, Turina M, Neuhaus V, et al. Laparoscopic cholecystectomy as a teaching operation: comparison of outcome between residents and attending surgeons in 1,747 patients. *Langenbeck Arch Surg*. 2012;397:103–110.
9. Iannuzzi JC, Rickles AS, Deeb AP, et al. Outcomes associated with resident involvement in partial colectomy. *Dis Colon Rectum*. 2013;56:212–218.
10. Kern SQ, Lustik MB, McMann LP, et al. Comparison of outcomes after minimally invasive versus open partial nephrectomy with respect to trainee involvement utilizing the American College of Surgeons National Surgical Quality Improvement Program. *J Endourol*. 2014;28:40–47.
11. Iannuzzi JC, Chandra A, Rickles AS, et al. Resident involvement is associated with worse outcomes after major lower extremity amputation. *J Vasc Surg*. 2013;58:827–831. e1.
12. Naiditch JA, Lautz TB, Raval MV, et al. Effect of resident postgraduate year on outcomes after laparoscopic appendectomy for appendicitis in children. *J Laparoendosc Adv Surg Tech A*. 2012;22:715–719.
13. Offner PJ, Hawkes A, Madayag R, et al. General surgery residents improve efficiency but not outcome of trauma care. *J Trauma*. 2003;55:14–19.
14. Rovito PF, Kreitz K, Harrison TD, et al. Laparoscopic Roux-en-Y gastric bypass and the role of the surgical resident. *Am J Surg*. 2005;189:33–37.
15. Schoenfeld AJ, Serrano JA, Waterman BR, et al. The impact of resident involvement on post-operative morbidity and mortality following orthopaedic procedures: a study of 43,343 cases. *Arch Orthop Trauma Surg*. 2013;133:1483–1491.
16. Castleberry AW, Clary BM, Migaly J, et al. Resident education in the era of patient safety: a nationwide analysis of outcomes and complications in resident-assisted oncologic surgery. *Ann Surg Oncol*. 2013;20:3715–3724.
17. Davis SS, Jr, Husain FA, Lin E, et al. Resident participation in index laparoscopic general surgical cases: impact of the learning environment on surgical outcomes. *J Am Coll Surg*. 2013;216:96–104.
18. Papandria D, Rhee D, Ortega G, et al. Assessing trainee impact on operative time for common general surgical procedures in ACS-NSQIP. *J Surg Educ*. 2012;69:149–155.
19. Scarborough JE, Bennett KM, Pappas TN. Defining the impact of resident participation on outcomes after appendectomy. *Ann Surg*. 2012;255:577–582.
20. Tseng WH, Jin L, Canter RJ, et al. Surgical resident involvement is safe for common elective general surgery procedures. *J Am Coll Surg*. 2011;213:19–26; discussion 28.
21. Kiran RP, Ahmed Ali U, Coffey JC, et al. Impact of resident participation in surgical operations on postoperative outcomes: National Surgical Quality Improvement Program. *Ann Surg*. 2012;256:469–475.
22. Shah R, Velanovich V, Syed Z, et al. Limitations of patient-associated comorbidity model in predicting postoperative morbidity and mortality in pancreatic operations. *J Gastrointest Surg*. 2012;16:986–992.
23. Turner PL, Saager L, Dalton J, et al. A nomogram for predicting surgical complications in bariatric surgery patients. *Obes Surg*. 2011;21:655–662.
24. Margenthaler JA, Longo WE, Virgo KS, et al. Risk factors for adverse outcomes following surgery for small bowel obstruction. *Ann Surg*. 2006;243:456–464.
25. American College of Surgeons Surgical Risk Calculator. Available at: http://riskcalculator.facs.org. Accessed March 27, 2014.
26. American College of Surgeons National Surgical Quality Improvement Program. User Guide for the 2010 participant use data file. Available at: http://site.acsnsqip.org/wp-content/uploads/2012/03/2010-User-Guide_FINAL.pdf. Accessed August 3, 2013.
27. Blackwell M, Iacus SM, King G, et al. Cem: coarsened exact matching in Stata. *Stata J*. 2009;9:524–546.
28. Krell RW, Birkmeyer NJ, Reames BN, et al. Effects of resident involvement on complication rates after laparoscopic gastric bypass. *J Am Coll Surg*. 2014;218:253–260.
29. Wilkiemeyer M, Pappas TN, Giobbie-Hurder A, et al. Does resident post graduate year influence the outcomes of inguinal hernia repair? *Ann Surg*. 2005;241:879–882; discussion 882-884.
30. Kazaure HS, Roman SA, Sosa JA. The resident as surgeon: an analysis of ACS-NSQIP. *J Surg Res*. 2012;178:126–132.
31. Itani KM, DePalma RG, Schifftner T, et al. Surgical resident supervision in the operating room and outcomes of care in Veterans Affairs hospitals. *Am J Surg*. 2005;190:725–731.
32. Shiloach M, Frencher SK, Jr, Steeger JE, et al. Toward robust information: data quality and inter-rater reliability in the American College of Surgeons National Surgical Quality Improvement Program. *J Am Coll Surg*. 2010;210:6–16.
33. Visser BC, Keegan H, Martin M, et al. Death after colectomy: it's later than we think. *Arch Surg*. 2009;144:1021–1027.
34. De Ponti R, Marazzi R, Doni LA, et al. Simulator training reduces radiation exposure and improves trainees' performance in placing electrophysiologic catheters during patient-based procedures. *Heart Rhythm*. 2012;9:1280–1285.
35. Nestel D, Groom J, Eikeland-Husebo S, et al. Simulation for learning and teaching procedural skills: the state of the science. *Simul Healthc*. 2011;6:S10–S13.

DISCUSSANTS

J.A. Sosa (Durham, NC):

Thank you for an interesting study that highlights the importance of graduate surgical education and its implications for patient safety and policy. The findings are important and notable, as they contradict several other studies using ACS/NSQIP and VASQIP that have demonstrated little to no clinically significant difference in patient outcomes based on resident involvement. If not interpreted with care, policy makers, payers, and the public could construe that surgical care at academic health centers is compromised by trainees, which could have unfortunate ramifications for everyone in the room, and trainees and patients. The authors carefully frame their discussion and limitations but will need to ensure going forward that others' discussions of implications remain evidence based. Statistically significant findings in this study may not represent clinically significant differences, given the large size of the data set.

My worry is about drawing causal inferences from an observational study in the setting of possible selection bias. The authors nicely

acknowledge concern that more complex cases go to teaching institutions and residents, as confirmed by differences in the unmatched cohorts. The authors do a good job trying to match but are constrained by the limitations of the data set itself. Although matching can reduce differences in populations, it cannot adjust for unobserved factors. They attempt to match for case complexity using CPT codes, which does not necessarily account for, say, the difference between a routine appendectomy that takes 15 to 30 minutes or a complex one that, say, takes 3 hours for a perforation. The increasing OR time and morbidity associated with increased postgraduate training here supports the notion that these findings might be swayed by some degree of selection bias. To feel confident in the observed findings, it would be nice to have a control variable.

I have 3 questions.

First, could a comparison of final diagnoses be performed that might be informative if the more interesting or challenging cases are being directed to residents, as evidenced by the diagnosis code? There should be consideration of severity of diagnosis in the model.

Second, would it also lend credibility if the authors could show that increasing year of residency was associated with improved outcomes in their final model, since residents should be more proficient over time as they progress through training? So, was year of training included in the multivariable model?

And, third, could the authors use the NSQIP variable for the availability of the attending surgeon in the OR or surgical suite in their model to supplement their analysis?

Response From G. Kasotakis:

I would agree that we have to be very careful in interpreting our data. We cannot really infer causation when it comes to looking at outcomes whenever residents are involved. And, of course, we cannot do away with residents if we want to have a future in any surgical field.

To answer your questions, we did not use diagnosis codes to not excessively constrain sample size. However, we can easily look at this information in the final article.

With regard to the effect of trainee level, year of training was not included in our multivariable model; however, we noted a univariate trend with longer operative times in cases performed by senior trainees. We found that this was not due to the increased autonomy senior residents enjoy in the operating room, but a self-selection to more complex cases, as we discussed.

Finally, we had hoped to be able to use the NSQIP variable that documents attending availability or presence in the OR; however, more than 99.9% of the cases we analyzed had the attending as present in the operating room, essentially rendering this parameter unusable. Also, when we looked at the type of procedures and diagnoses of patients who went to surgery without a staff surgeon present in the OR, we noted that the vast majority of them were uncomplicated appendectomies and drainage of soft tissue abscesses, limiting further the variable's usability. It would be great if NSQIP tracked degree of trainee and attending involvement in each case; however that is not the case, and at present, we have only very granular information on whether a staff surgeon was simply in the OR.

I have to reiterate that our study was a retrospective look at a database that was not specifically designed to look at the effect of trainee participation in clinically relevant outcomes, and our findings have to be interpreted with caution.

DISCUSSANTS

G. Fried (Montreal, QC):

I have no disclosures. Although large data set analyses like these are vulnerable to methodologic criticisms and the assumptions upon which they're based, I would like to comment more on the philosophic aspects of this.

Resident education is a fundamental responsibility that we own to ensure the provision of quality care to the next generation of our patients. The data that you have presented here show that engaging residents with responsibilities for patient care, although clearly necessary for their training, comes at a cost, both in terms of outcomes and finances. Some dimensions of these costs have been documented in this article.

Unfortunately, development of a strategy to shift the focus of learning from the clinical setting partially to the simulation setting requires a substantial investment in human and financial resources to fulfill these needs. These data, and those from other similar studies, must be captured and used to provide some evidence for us so that we could advocate for the resources that will allow us to provide better educational programs and to protect the best interests of our patients during the learning phase.

Residents provide inexpensive care. There is a cost to that. We need to provide some evidence that really gets the attention of people that will provide some of the resources to balance that out. I think such an approach makes sense both economically and ethically.

Response From G. Kasotakis:

I don't think I could agree more with your remarks. With our findings, we believe that the majority of the additional morbidity can be attributed to 2 major factors that all of us recognize in our daily practice: The prolonged OR time cases take whenever residents participate, and, 2, perhaps to their less than perfect surgical skills. Factors like these may also come into play when junior attendings operate, compared with their senior colleagues.

However, we do have to better prepare residents before they come to the OR and tackle those emergency operations. Also, both surgeons and residents have to be mindful of the time. We have to be mindful of the clock. We have to aim to be more efficient, so that we can improve our outcomes overall.

Again, this is by no means a prospective clinical trial that randomizes emergency cases to resident involvement versus not, and one has to be mindful of the limitations of our analysis. But there is no smoke without fire, and we can certainly use these findings, whatever their limitations, to advocate for more resources in better training the surgery resident of the future.

DISCUSSANTS

M.T. Hawn (Birmingham, AL):

I echo Dr. Sosa's concerns about confounding. I wondered whether you looked at admission source and excluded transfer patients to try and deal with the referral bias to tertiary care centers.

My second question is more philosophical. Why are we ascribing all the increased complications to the trainee, and why isn't your title "Academic Surgeons Affect Outcomes in Emergency General Surgery"? Aren't we ultimately responsible?

Response From G. Kasotakis:

Unfortunately, we did not look at the transfer status of patients. So, of course, that is something that can be taken into consideration when we look further and deeper into the effect of trainee participation in emergency operations.

Also, yes, we do believe that there is some residual confounding that we were not able to address because of the nature of the data set and, again, the main limit was that this was a large data set that was not designed to address this specific question. So, we do acknowledge our limitations, and we are looking to work with something better in the future.

With regard to the effect of academic surgeons or attendings' seniority on outcomes, it's a great idea for future research projects.

DISCUSSANTS

H. Polk (Louisville, KY):

I would like to join the skeptics' lobby, skeptical with the work and skeptical about your conclusions. I have worked with enough large databases to know that you can take any nonsensical comparison you like, and if one makes the denominator big enough, they are all $P < 0.0001$. There's a lot of that in here.

What determined whether or not a resident scrubbed on a case in this comparison? Why didn't you use the Charlson index for some easily understood, reliable comparison of degree of comorbidities? Why did you have more transfusions in the OR when a resident was present but more transfusions postoperatively when a resident was absent? And, finally, what does the presence or absence of residents have anything to do with DVT or pulmonary embolism?

There are lots of failings in this work.

Response From G. Kasotakis:

Deciding what factors come into play on whether residents and of what postgraduate level will scrub into certain cases is a difficult task. Again, that is a variable that is unfortunately missing from the NSQIP data set, so we cannot really comment on that all that much.

I'm not really sure we could have used a Charlson index, given patients' preexisting comorbidities are coded in such a way in NSQIP that precludes reliably calculating such a parameter. However, we did use the preoperatively calculated risk for morbidity and mortality to risk-stratify patients and control for baseline risk factors. This risk calculator contains information on the indication for surgical intervention and more than 22 clinical parameters that are known risk factors for adverse outcomes. It is also one of the key parameters that NSQIP uses to risk-adjust and compare outcomes and is well-validated for that purpose. So, I'm not sure a Charlson index would have performed better. Again, as we did not design the database, we can only use what's already available.

With regard to the association of resident participation and intraoperative transfusions and venous thromboembolic events, I think the common denominator is longer operative time. The longer you dissect through tissue, the more bleeding you are likely to cause, with case complexity controlled for. The longer you leave someone sedated and immobilized, the more likely it is for them to develop venous thrombosis.

DISCUSSANTS

O. Kirton (Hartford, CT):

Obviously, as everyone knows, the ACGME is requiring that we provide some type of clinical assessment back to our residents. And we are looking at NSQIP, all of us, hoping that this could be an appropriate tool. But there are very many limitations to NSQIP, particularly when you look at emergency surgery, there are many variables—whether surgery occurs during the day, whether it occurs at night, is it performed by the lesser experienced junior faculty who are doing the emergency surgery, or whether it is senior faculty. And who is actually operating with the general surgeon faculty when it's not the resident assisting? The Advanced Practitioner, or the surgical technologist.

I think we just need to be cautious, particularly, as brought up by Dr. Hiram Polk, when you look at some of these complications, is it a reflection of the resident, or is it a reflection of supervision and training?

Again, I think more and more, we need to be able to give better feedback and better supervision to our residents because the time we have with them is increasingly limited.

Response From G. Kasotakis:

Again, yes, we do have to be cautious about how we are going to interpret these findings, but I think this is a first look at how resident participation may have a slightly increased risk of complications that do matter. We have to take that in mind before we look toward the future.

Again, we did not attempt to answer the question on whether staff surgeon seniority and level of involvement, or degree of resident supervision, are associated with certain outcomes. These parameters likely contribute to some residual confounding and would have to be addressed in the future. We wanted to address if resident participation, as documented in the NSQIP data set, is associated with certain outcomes after emergency surgery.

DISCUSSANTS

G. Velmahos (Boston, MA):

Even if I must confess that I do not fully agree with your conclusions, and I'm sure that your popularity with the residents must be at an all-time low, I recognize the usual scientific rigor with which you perform your studies.

My question pertains to legal and maybe informed consent aspects, particularly relevant to this audience, because almost everybody here works at a teaching institution.

Are we now, based on this study, supposed to inform patients that there is a resident in the OR with us, and, therefore, the outcomes of the patient may be compromised and they better choose another place?

Response From G. Kasotakis:

It is true, the residents at our training program were not happy to hear our findings, when we presented them recently to them. From a medicolegal aspect, I think that's an important concern that we have to keep in mind. Perhaps training institutions should consider adding a clause to surgical consents that state that residents of varying seniority and skillset will be participating and assisting in a case, explaining to the patients that if staff surgeons believe that their skills are not up to par of what is expected for that case, they can take over.

DISCUSSANTS

M. Malangoni (Philadelphia, PA):

This is an interesting article. Yours is not the first to show an association between resident participation with operating room cases as increasing the operative time or increasing the risk of complications. What's been demonstrated before is that the complications have been relatively minor.

I share others' concerns about attributing this association the operative participation of a resident as opposed to potential factors.

What is surprising about your results is that these increased complications did not result in an increased length of stay. So, that raises the question of whether or not, in your matching, you adjusted for institutional characteristics. Despite the strict definitions within the NSQIP database for each of these complications, there may be differences in screening techniques that could influence the results. An example would be VTE, where an intensive screening technique will pick up more cases of VTE than are necessarily clinically significant, which would support the fact that there's no change in length of stay.

Did you account for institutional differences in your matching?

Response From G. Kasotakis:

Unfortunately, NSQIP does not include information from what institution the patients come. So, it was not possible for us to do that. It would be great if we had that information at hand and could perform a multilevel analysis.

However, I do agree that we are not the first to demonstrate an association with adverse outcomes whenever residents participate. That may be just because of the fact that patients are typically referred to a higher level of training institutions, where residents are more likely to be available and help out in those cases. Now, again, we are very careful to not mention the word "causation" in our article or in our discussion today, because we can obviously not do that.

But coming back to the hospital length of stay, we did not see a direct, independent association between resident participation and length of stay. However, when we compared the 2 groups, the group that had their procedures performed by residents stayed longer in the hospital, just because they had more complications. And, as you mentioned, there was no independent resident effect on overall length of stay, because it's the baseline comorbidities and the case complexity and potentially the complications that would prolong the hospital stay; not the presence of residents in the OR, something that kind of makes sense.

DISCUSSANTS

L. Ratner (New York, NY):

I was struck very much by things like reintubation, failure to wean, transfusion requirements, OR time, and anesthetic time that were all increased in the resident group. As someone who has spent my entire career in academic medical centers, I'm struck by how often the surgical residents are getting a whole lot more supervision than the anesthesia residents. As someone who likes to share the wealth, is there any data in the database regarding the anesthesia residents and whether or not they were supervised or whether or not they participated?

Response From G. Kasotakis:

That is a great observation also. We noted that the anesthesia time was prolonged by a greater amount than the total operative time, suggesting that anesthesia trainees may play a small role somewhere there and perhaps contribute to some small extent to the unmeasured confounding. That is definitely something that we have to keep in mind. However, this information on whether anesthesia trainees participate, how they are supervised, and how aggressive they are with giving IV fluids and blood products, we simply don't have the data to test whether an association exists, but I personally believe that it does.