Patient Outcomes with Teaching Versus Nonteaching Healthcare: A Systematic Review

Panagiotis N. Papanikolaou[1], **Georgia D. Christidi**[1], **John P. A. Ioannidis**[1,2*]

1 Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina, Greece, 2 Institute for Clinical Research and Health Policy Studies, Department of Medicine, Tufts University School of Medicine, Boston, Massachusetts, United States of America

Funding: The authors received no specific funding for this study.

Competing Interests: The authors have declared that no competing interests exist.

Academic Editor: Andrew D. Oxman, Norwegian Directorate for Health and Social Welfare, Norway

Citation: Papanikolaou PN, Christidi GD, Ioannidis JPA (2006) Patient outcomes with teaching versus nonteaching healthcare: a systematic review. PLoS Med 3(9): e341. DOI: 10.1371/journal.pmed.0030341

Received: February 7, 2006
Accepted: June 9, 2006
Published: September 12, 2006

DOI: 10.1371/journal.pmed.0030341

Abbreviations: CI, confidence interval; RR, relative risk

* To whom correspondence should be addressed. E-mail: jioannid@cc.uoi.gr

ABSTRACT

Background

Extensive debate exists in the healthcare community over whether outcomes of medical care at teaching hospitals and other healthcare units are better or worse than those at the respective nonteaching ones. Thus, our goal was to systematically evaluate the evidence pertaining to this question.

Methods and Findings

We reviewed all studies that compared teaching versus nonteaching healthcare structures for mortality or any other patient outcome, regardless of health condition. Studies were retrieved from PubMed, contact with experts, and literature cross-referencing. Data were extracted on setting, patients, data sources, author affiliations, definition of compared groups, types of diagnoses considered, adjusting covariates, and estimates of effect for mortality and for each other outcome. Overall, 132 eligible studies were identified, including 93 on mortality and 61 on other eligible outcomes (22 addressed both). Synthesis of the available adjusted estimates on mortality yielded a summary relative risk of 0.96 (95% confidence interval [CI], 0.93–1.00) for teaching versus nonteaching healthcare structures and 1.04 (95% CI, 0.99–1.10) for minor teaching versus nonteaching ones. There was considerable heterogeneity between studies ($I^2 =$ 72% for the main analysis). Results were similar in studies using clinical and those using administrative databases. No differences were seen in the 14 studies fully adjusting for volume/experience, severity, and comorbidity (relative risk 1.01). Smaller studies did not differ in their results from larger studies. Differences were seen for some diagnoses (e.g., significantly better survival for breast cancer and cerebrovascular accidents in teaching hospitals and significantly better survival from cholecystectomy in nonteaching hospitals), but these were small in magnitude. Other outcomes were diverse, but typically teaching healthcare structures did not do better than nonteaching ones.

Conclusions

The available data are limited by their nonrandomized design, but overall they do not suggest that a healthcare facility's teaching status on its own markedly improves or worsens patient outcomes. Differences for specific diseases cannot be excluded, but are likely to be small.

The Editors' Summary of this article follows the references.

Introduction

A large number of studies, including many in leading medical journals (see Protocol S1 for references for the studies themselves) [1,2] have tried to address whether medical teaching settings obtain better patient outcomes than nonteaching ones; theoretically a teaching environment may also entail unnecessary risks for patients. Superior outcomes in teaching settings have been reported in some studies, but others have claimed the opposite [3].

The pertinent evidence is derived entirely from non-randomized studies. However, patient populations in different settings may have different case-mixes—for example, academic centres often receive the most difficult cases [4–6]. Teaching versus nonteaching settings may differ also in structure, e.g., availability of technology or volume of patients [7–11]. To avoid confounding, these covariates need to be accounted for. Healthcare structures may also differ in process characteristics, i.e. measures that address the appropriate implementation of healthcare. These are not patient outcomes, but may translate into differential outcomes, e.g., if the right treatment is used more frequently, patients should do better. A comparison of outcomes in teaching versus nonteaching units should adjust for differences in patients and structures. However, adjustment for process would dilute any true outcome differences, since process explains the outcomes. To further complicate matters, data on outcomes and on adjusting covariates may be derived from different administrative or clinical sources [2]. Clinical data are usually more accurate than administrative data that depend on utilization databases.

To address this issue, it is necessary to examine the impact of these issues across different studies and to try to generate a systematic picture of the available evidence on the comparison between teaching and nonteaching healthcare structures. We set to do this by examining the evidence for diverse patient outcomes.

Methods

Eligibility

We considered eligible all controlled studies in which any teaching healthcare structures were compared against nonteaching counterparts on any subjective or objective patient outcome. We considered English-language studies regardless of the unit of the healthcare structure (e.g., hospital, service, physician, health system) and regardless of whether teaching status was a primary or secondary analysis.

Teaching hospitals are sometimes further divided into major and minor teaching ones. This distinction is well defined for US hospitals [1,2]. However, to accommodate non-US studies, we allowed contrasts that may be defined with various terms: e.g., the teaching unit(s) may be termed university, academic, medical school-affiliated, resident service, or service with house staff; and the comparator unit(s) may be termed nonuniversity/community, nonacademic, nonaffiliated, or service without residents and students. Overall, we considered all studies in which the compared units differed in teaching status, regardless of the terminology employed. We did not consider studies where compared units differed on whether they had an academic affiliation or not, but were all employed in teaching (e.g., university versus

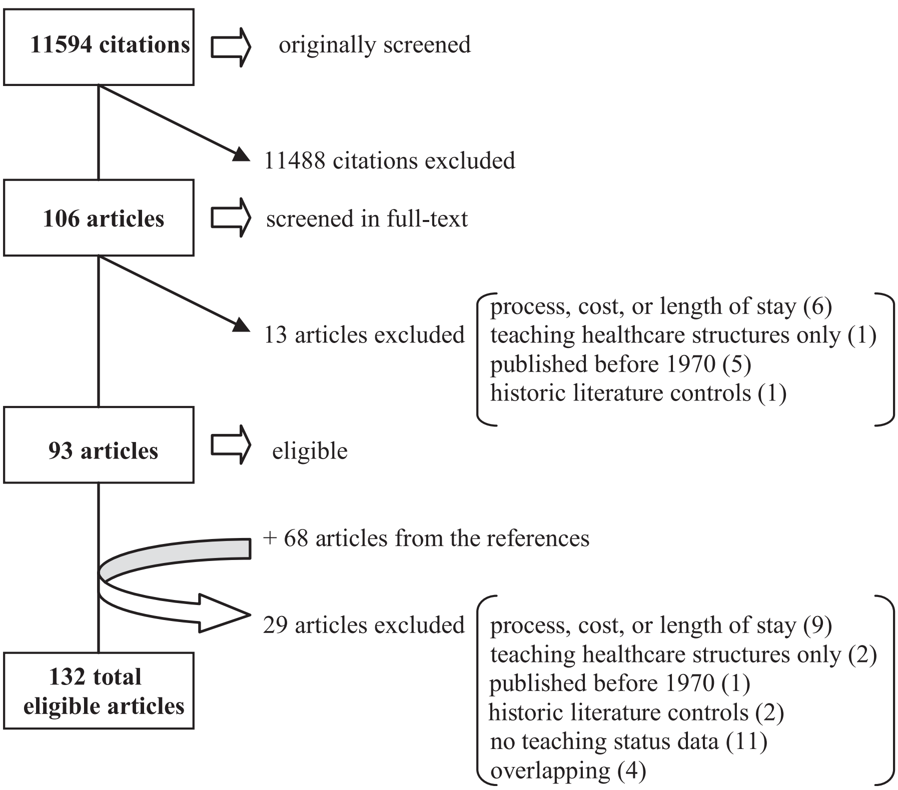

Figure 1. Flow Figure for Screened, Included, and Excluded Articles
DOI: 10.1371/journal.pmed.0030341.g001

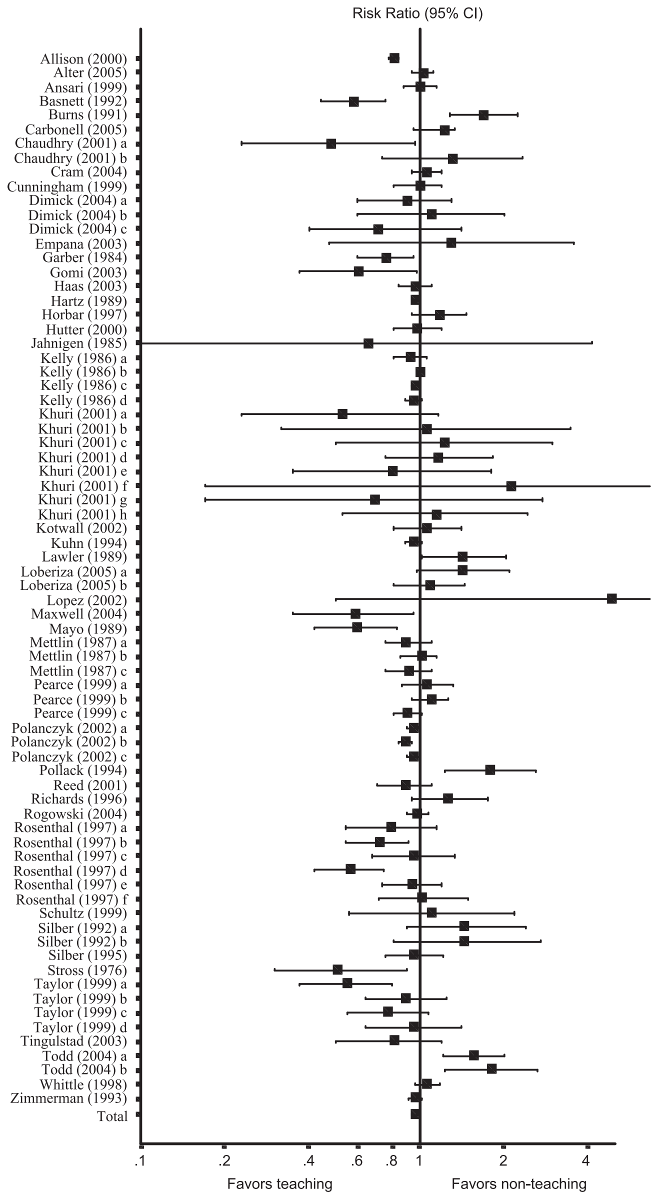

Figure 2. Relative Risk Estimates

RR estimates and 95% CIs across studies are shown that address mortality along with summary RR by random effects calculations (Total). The order of the presented estimates is the same as in Table 2. Articles that include estimates on various diagnoses are presented with separate estimates according to type of diagnosis and lettered in the order of Table 2.

DOI: 10.1371/journal.pmed.0030341.g002

teaching community hospital). Studies of tertiary versus primary/secondary care or specialists versus generalists were not eligible, unless the distinction also coincided with teaching status.

We considered studies regardless of patient outcomes addressed, but excluded length of stay, cost or financial parameters, and process measures from our consideration. We also excluded overlapping articles (retaining only the one with the most complete information), studies with historical literature controls, and studies published before 1970, since these would be largely irrelevant for current healthcare. We excluded nonoriginal articles (editorials, opinion pieces, reviews), and meeting abstracts.

Searches

We searched PubMed (last search updated June 2005). Given the multifarious nature of eligible studies, we experimented with a variety of different search strategies and compared their yield for selected articles with capture-recapture methods. Our final search was "(university OR academic OR teaching OR faculty OR medical school OR affiliation) AND (community [title word] OR non-academic OR non-academic OR non-teaching OR non-teaching OR affiliation)." Search of the Cochrane Central Registry of Controlled Trials yielded no additional references. We also contacted experts. First we screened titles and abstracts. Articles deemed potentially relevant were screened in the full text. We also perused reviews and the reference lists of identified eligible articles. Two investigators performed the searches. A third investigator settled discrepancies.

Outcomes

We defined mortality as the primary outcome, since it is the most important and objective outcome although some variability may still be encountered across studies, e.g., the time frame during which mortality is captured. Furthermore, studies addressed mortality more than any other outcome. We set no restriction on eligibility for secondary patient outcomes, and accepted both objective (e.g., morbidity) and subjective (e.g., patient satisfaction) outcomes.

Data Extraction

For each eligible study, we extracted information on authors, year of publication, type of data (clinical data where information was obtained directly from medical records versus administrative data from utilization databases), reported academic affiliation of corresponding author (yes, no, unclear), time period for the analysed databases, number of hospitals or healthcare structures, number of patients, and definitions of teaching and nonteaching status. Academic affiliation was derived from addresses of health science-related schools, universities, hospitals, or other institutions at which teaching is a key mission; nonacademic addresses included research or nonresearch institutions and hospitals at which teaching is not a key mission.

For all studies that addressed mortality, we recorded whether any adjustment was made for nonprocess covariates. If so, we recorded the adjusting covariates as well as the adjusted estimate of the relative risk (RR) and 95% confidence interval (CI) for the comparison between teaching versus nonteaching healthcare structures. RRs and 95% CIs were derived from the presented data in various forms depending on the model used for analysis (e.g., Cox hazard ratios, odds ratios from logistic regressions, Poisson relative risks, standardized mortality ratios, etc.) and were computed from the presented information when not directly stated by the authors. When both major and minor (other) teaching categories were available, we recorded separate RR and 95% CI for major teaching versus nonteaching and other teaching versus nonteaching comparisons. When multiple adjusted models were available, we preferred the one with the most extensive nonprocess adjustments. Data were recorded separately for different health conditions, whenever such

Table 1. Characteristics of Eligible Studies

Author (Year)	Country of Origin	Time Period	Healthcare Structures[a] (Patients)	Contrasted Groups
Agabiti (2003)[b]	Italy	1996–1997	5 (1,603)	Teaching, nonteaching
Aiken (1994)	US	1988	234 (ND)	Teaching "magnet" hospitals (with good nursing care), nonteaching magnet hospitals, control
Akkersdijk (1994)	Holland	1990	ND (1998)	University, other
Allison (2000)	US	1994–1995	4,361 (11,4411)	Major teaching, minor teaching, nonteaching
Alter (2005)	Canada	1997–2000	ND (139,484)	Academically affiliated, non-academic
Alter (2001)	Canada	1992–3	190 (25,697)	Teaching, nonteaching
Alter (2003)	Canada	1994–8	201 (15,166)	Academically affiliated, non-academically affiliated
Andersen (2000)	Norway	1996–8	14 (32,248)	University, community
Anderson (1992)	US	1988–9	16 (1,231)	Teaching, nonteaching
Ansari (1999)[b]	Australia	1987–1998, 1994–1995	36 (ND)	Teaching, nonteaching
Ansari (1998)[b]	Australia	1989–1995	ND (24,534)	Teaching, nonteaching
Ashley (1971)	UK	1964–1999	5 (932)	Teaching, regional-board
Bach (1998)	US	1995–1996	1 (118)	University service, community service
Barry (2003)	US	1991–1997	28 (156,136)	Major teaching, minor teaching, nonteaching
Basnett (1992)[b]	UK	1982–1986	3 (999)	Teaching, nonteaching
Beebe (2000)	US	1997–1998	3 (536)	University, community
Benitez (2001)[b]	Spain	1996	31 (2,240)	Teaching, nonteaching
Bottsford (1997)[b]	US	1980–1995	1 (362)	Resident service, non-resident service (same surgeon)
Brennan (1991)	US	1984	51 (31,429)	Primary teaching, other teaching, nonteaching
Bruster (1994)	UK	1992–1993	36 (5,150)	Teaching, nonteaching
Burns (1991)	US	1988	47 (54,571)	Teaching, nonteaching
Carbonell (2005)[b]	US	2000	994 (93,578)	Teaching, community
Carpenter (1972)	US	ND	5 (1,117)	University, community
Chaudhry (2001)	Canada	1991–1996	ND (938)	Teaching, nonteaching
Chen (1999)	US	1994–1995	4,672 (149,177)	Teaching, nonteaching, small urban, small rural
Cohen (1989)	Canada	1972–1984	ND (37,383)	Urban teaching, urban nonteaching, large rural, small rural
Cole (1995)[c]	Scotland	1986–1990	ND (326,259)	Teaching, nonteaching specialist, GP maternity
Cooper (1996)	US	1987–1989	ND (81,579)	Teaching, nonteaching
Cooper (1998)	US	1994	13 (1,031)	Major teaching, minor teaching, nonteaching
Coyte (1999)	Canada	1984–1991	ND (18,530)	Teaching, nonteaching
Cram (2004)	US	1998	441 (641,860)	Major teaching, minor teaching, nonteaching
Cunningham (1999)	US	1994	333 (7,901)	Teaching, nonteaching
Davis (1997)	US	1993–1994	47 (780)	Academic, community-based
Desai (2005)	US	ND	3 (533)	Strong, close, no medical school affiliation
Dimick (2004)	US	1996–1997	ND (66,85)	Teaching, nonteaching
Diringer (2001)	US	1996–1999	42 (1,038)	Affiliated with medical school, non-affiliated with medical school
Dubois (1995)	Canada	1990–1992, 1991–1993	2 (500)	Teaching, community
Duggirala (2004)	US	1990–1996	3,818 (ND)	Major teaching, other teaching, nonteaching
Elhanan (1995)[c]	Israel	1990–1991	2 (1,489)	University, community
Empana (2003)	France	1997	19 (585)	Teaching, nonteaching level III facilities
Espehaug (1999)	Norway	1988–1996	70 (39,505 hips)	University, central (also training), local (non-training)
Feagan (2000)[b]	Canada	1996–1997	20 (858)	Teaching, community
Feigenson (1978)[b]	US	mid-1970s	84 (439)	Teaching, community
Finkelstein (1998)	US	1992–1994	18 (1,6051)	Teaching, nonteaching
Fleming (1981)	US	1976	300 (490)	Teaching, nonteaching
Fleming (1991)	US	1985	657 (ND)	Teaching, nonteaching
Flood (1984)	US	ND	1,224 (494,051)	Teaching, nonteaching
Fox (2002)[c]	14 countries	1999–2000	95 (11,543)	Teaching, nonteaching
Garber (1984)[b]	US	1981	1 (2,025)	Faculty service, community service
Garcia (2001)	US	1996	52 (63,143)	Academic, community teaching, community nonteaching
Gilchrist (1999)	US	1993–1994	82 (4,010)	Teaching (academic), nonteaching (nonacademic or private)
Goldacre (1976)	UK	1969–1973	ND (687)	Teaching, general nonteaching (regional board), infectious disease
Gomi (2003)[b]	Japan	1998–2001	76 (220)	Academic, non-academic
Gorst (1992)[c]	UK	1982–1989	13 (953)	Tertiary centres, district general hospitals
Gregory (1999)	US	1991	78 (92,798)	Private teaching, private nonteaching, HMO, public
Gregory (2001)	US	1988–1991	ND (10,400)	Private teaching, private nonteaching, HMO, public
Gregory (2001)	US	1995	288 (463,196)	Teaching, nonteaching
Haas (2003)	US	1996	357 (54,874)	Teaching, public or district, rural, other
Hartz (1989)	US	1986	3,100 (ND)	Public teaching, private teaching, nonteaching
Heck (1998)	US	1985–1990	ND (208,980)	Teaching, nonteaching
Horbar (1997)	US	1991–1992	62 (7,672)	Teaching, nonteaching
Howard (2004)	US	1988, 1991	252 (313,669)	Teaching, nonteaching
Huber (2001)[c]	US	1994–1996	900 (16,450)	Teaching, nonteaching
Hutter (2000)	US	1990–1996	357 (5,696)	University, affiliated teaching, nonteaching
Jahnigen (1985)	US	1981–1983	6 (124)	Teaching, nonteaching
Jensen (2004)[b]	US	1997–2001	8 (231)	Academic, VA, private

Table 1. Continued

Author (Year)	Country of Origin	Time Period	Healthcare Structures[a] (Patients)	Contrasted Groups
Johnston (2001)	US	1997–1999	2 (569)	Academic, community
Kazmers (1996)	US	1991–1993	116 (3,687)	Large hospitals with strong affiliation with medical schools, middle-sized hospitals with strong affiliation with medical schools, middle-sized hospitals, small rural hospitals, psychiatric hospitals
Keeler (1992)[c]	US	1981–1982, 1985–1986	297 (14,008)	Major teaching, moderate/limited teaching, nonteaching
Kelly (1986)[b]	US	1977	160 (5,694)	Teaching, medical-school affiliated, non-medical-school affiliated
Khuri (2001)	US	1997–1999	128 (690,811)	Teaching, nonteaching
Kingston (1992)[b]	UK	1981–1983	12 (567)	Teaching, nonteaching
Kotwall (2002)	US	1988–1995	720 (24,926)	Urban teaching, urban nonteaching, rural
Krupski (1985)	US	1980–1983	3 (300)	University, private community, university-affiliated VA
Kuhn (1994)	US	1988	3,782 (ND)	Osteopathic, for-profit, public nonteaching, public teaching, private not-for-profit nonteaching, private not-for-profit teaching
Lawler (1989)	US	1985–1986	1 (523)	Academic, community physicians whose patients are admitted to an ICU
Lee-Feldstein (1994)	US	1984–1990	126 (5,892)	Teaching, HMO, large community, small community
Litwin (1992)[b]	Canada	1990–1991	31 (2,201)	Teaching, nonteaching
Loberiza (2005)	US	1998–2000	163 (4,285)	Medical school affiliation with students/residents, fellows, both vs. non-affiliated
Lopez (2002)	US	1999	44 (933)	University-based, community-based surgeons
Maxwell (2004)	US	1988–1999	ND (4,601)	Urban teaching, urban nonteaching, rural
Mayo (1989)	Canada	1984–1985	27 (3,045)	Teaching, affiliated, nonteaching/nonaffiliated
Mazer (1993)	US	1988	4 (335)	Teaching, nonteaching
Mettlin (1987)[b]	US	1977, 1982, 1984	ND (ND)	Medical school affiliated, non-medical school affiliated
Meyers (1998)	US	1988–1994	134 (107,648)	Academic medical centers, nonteaching
Mylotte (2001)	US	1995	2 (249)	Teaching, nonteaching
Oei (2002)[c]	US	1997–1998	2 (362)	University, community
Oleske (1991)	US	1986	ND (130,249)	Teaching, nonteaching
Pearce (1999)	US	1992–1996	835 (90,331)	Teaching, nonteaching
Peterson (1994)	US	1988–1990	158 (33,641)	Medical school affiliated, non-medical school affiliated
Polanczyk (2002)	US	1993–1995	248 (388,964)	Major teaching, minor teaching, nonteaching
Pollack (1994)	US	1989–1992	16 (5,415)	Teaching, nonteaching
Reed (2001)	US	1998–1999	137 (23,058)	Teaching community, nonteaching community
Richards (1996)	UK	1984–1988	49 (1,324)	Teaching, nonteaching
Richards (1998)[b]	US	1992–1997	93 (227,451)	Teaching, nonteaching
Richards (2000)[b]	US	1992–1998	205 (498,998)	Major teaching, graduate teaching, limited teaching, nonteaching
Rock (1993)	US	1983, 1988	288 (7,523)	Teaching, nonteaching
Rogowski (2004)[b]	US	1995–2000	332 (94,110)	Teaching, nonteaching
Rosenthal (1997)	US	1991–1993	30 (89,851)	Major teaching, minor teaching, nonteaching
Schultz (1999)	US	1992	373 (ND)	Major teaching, graduate, limited teaching, nonteaching
Selker (1991)	US	1979–19981	6 (4,099)	University teaching, medical-school affiliated teaching, nonteaching
Shah (1971)	India	1961–1969	2 (ND)	City teaching, rural community
Shaughnessy (1995)	US	1984–1988	17 (ND)	Teaching, nonteaching
Sheng (2005)	Taiwan	2002	3 (273)	University, community
Silber (1995)	US	1991–1992	57 (16,673)	Teaching, nonteaching
Silber (1992)	US	1985–1986	531 (5,972)	With housestaff, without housestaff
Singh (1996)	India	ND	ND (2,000)	Medical college hospital, community hospitals
Sloan (2000)	US	1995	85 (32,593)	Teaching, nonteaching
Stiller (1994)	UK	1971–1988	ND (1,258)	Teaching, nonteaching
Stiller (1999)	UK	1984–1994	ND (879)	Teaching, nonteaching
Stone (1992)	US	1987–1988	40 (300)	High-intensity teaching, low-intensity teaching, nonteaching
Stross (1976)	US	1973–1974	23 (4,980)	University, urban, community
Tanisada (2000)	Japan	1992–1994	37 (336)	Academic, non-academic
Taylor (1999)	US	1984–1994	1,378 (3,206)	Major teaching, minor teaching, nonteaching
Teres (1983)	US	1978	2 (558)	Private and teaching, private
Thomas (2000)	US	1992	28 (2,820)	Major teaching, minor teaching, nonteaching
Thomas (2000)	US	1992	27 (15,000)	Major teaching, minor teaching, nonteaching
Thompson (2002)	US	1991–1992	27 (ND)	With or without orthopedic residency
Tingulstad (2003)	Norway	1992–1997	8 (115)	Teaching, nonteaching
Tingulstad (2003)	Norway	1987–1996	8 (571)	Teaching, nonteaching
Todd (2004)	US	1998–2000	ND (14,901)	Urban teaching, urban nonteaching, rural (only primary care residencies)
Udvarhelyi (1990)	US	1985	1 (180)	Teaching service, nonteaching service
Wade (1994)	US	1987–1991	327 (1,964)	University affiliated, non-affiliated
Wade (1996)	US	1989–1994	ND (130)	Teaching, nonteaching
Warren (1998)[b]	US	1993–1995	ND (8,549)	Heavy, moderate, light, no teaching load
Wells (1998)	US	1994–1995	1 (154)	Resident service, private community service
Whitehouse (2002)	US	1997–1998	2 (118)	Teaching, community

Table 1. Continued

Author (Year)	Country of Origin	Time Period	Healthcare Structures[a] (Patients)	Contrasted Groups
Whitsel (2000)	US	1997–1999	3 (5,705)	University, community
Whittle (1998)	US	1990	ND (22,294)	Teaching, nonteaching
Wolfe (1997)	UK	1991	ND (118)	Teaching, nonteaching with support, nonteaching without support
Yuan (2000)	US	1984–1993	5,127 (16,983,000)	For-profit, not-for-profit, osteopathic, nonteaching public, teaching not-for-profit, teaching public
Zimmerman (1993)	US	1988–1990	35 (15,297)	Teaching, nonteaching

[a]All refer to hospitals, except Bootsford (1997), Davis (1997), Diringer (2001), Kingston (1992), Lopez (2002), and Shuaghnessy (1995), which refer to, respectively, a single surgeon, endoscopists, intensive care units, surgeons, surgeons, and nursing homes.
[b]Nonacademic affiliated corresponding author.
[c]Unclear whether corresponding author has academic affiliation.
ND, no data.
DOI: 10.1371/journal.pmed.0030341.t001

separate information was provided. When only unadjusted estimates were given, these were recorded but not further analysed because of the problems of unadjusted estimates. For each study that provided adjusted estimates we recorded whether adjustments had addressed volume/experience, severity, and comorbidity—beyond simple considerations of age, gender, urgency of a procedure, and (for multiple-diagnosis studies) diagnosis-related group.

For all studies that addressed any other outcome besides mortality, the same considerations applied for recording adjusted estimates for informative comparisons, selecting adjusted models, and presenting data separately according to each outcome and condition of interest. Unadjusted analyses were simply recorded, as above.

Data extraction was performed by two investigators and further checked by a third senior investigator for accuracy. Discrepancies were discussed to reach consensus.

Data Synthesis

We summarized descriptive characteristics of the studies. We had anticipated that quantitative synthesis of the retrieved information may be precarious given the diversity in study designs, outcomes, measurements, and adjustments. Nevertheless, mortality was a very common, unambiguous outcome and was reasonably amenable to exploratory quantitative synthesis. Given the expected between-study heterogeneity, all formal meta-analyses were performed using random effects calculations [12] using general variance models. Each study was weighted by the inverse of its variance plus the estimated between-study variance. Between-study heterogeneity was estimated with the Q statistic [12] and the I^2 statistic [13]. The former is a chi square-based test for the statistical significance of heterogeneity, while the latter provides a measure of the extent of heterogeneity and values above 75 suggest very large heterogeneity.

We performed an overall synthesis using all data on mortality across all studies and separate analyses according to subgroups defined by type of data (clinical versus administrative), year of publication (per decade), location of study (US versus other), affiliation of corresponding author (academic, nonacademic, unclear), and whether the compared healthcare structures were named as teaching versus non-teaching or otherwise. We also performed separate analyses for each health condition where at least three RR estimates were available for data synthesis. A sensitivity analysis synthesized only the more rigorously adjusted studies where adjustments had considered volume/experience, severity, and comorbidity (as defined above). Finally, we examined whether observed effect sizes were related with the precision of the estimates [14]. When less-precise studies show more prominent effects than more precise studies, this may reflect bias (including publication bias), but may also hint to study-design differences or other genuine sources of heterogeneity.

Whenever data compared major teaching, minor (other) teaching, and nonteaching status, we used the major teaching versus nonteaching comparison. We also performed a separate analysis that considered all comparisons of minor teaching versus nonteaching status to address specifically whether healthcare structures that perform limited teaching have different mortality rates than nonteaching ones. No studies specifically defined and compared only minor teaching versus nonteaching healthcare.

Finally, for nonmortality outcomes, we simply described the range of estimates across studies. Analyses were performed in SPSS 12.0 (SPSS, Chicago, Illinois, United States).

Results

Retrieval of Articles and Eligible Studies

Of 11,594 originally screened articles (Figure 1), 132 articles were eventually eligible (Table 1): 93 addressed mortality, and 61 addressed other eligible outcomes (22 addressed both mortality and other outcomes). Most eligible studies were performed in the US ($n = 94$, 71.2%). Studies had also been done in Canada ($n = 10$), European countries ($n = 19$), Australia ($n = 2$), and Asia ($n = 6$), while one study was multinational. Studies covered a wide range of time periods, but 75 (56.8%) studied patient databases after 1991. The word "teaching" was explicitly used in the naming of the compared groups in 97 studies (73.5%), while in the other 35 the teaching versus nonteaching comparison could be indirectly inferred from the definition of the groups.

Of the 93 articles containing mortality data, 28 (30.1%) provided only unadjusted estimates, and 17 (18.3%) provided adjusted data that were not usable because of lack of sufficient detail in reporting or because the contrast was

Table 2. Comparisons of Teaching versus Nonteaching Healthcare on Patient Mortality

Author (Year)	Data	Adjusting Factors	Conditions	Adjusted RR (95% CI)[a]
Allison (2000)	C	Age, sex, race, cardiac arrest, CHF, SBP, creatinine, WBC, AMI location	AMI	MT-NT 0.80 (0.77–0.84), OT-NT 0.91 (0.84–0.95)
Alter (2005)	A	Age, sex, neighbourhood characteristics (6), on-site catheterization facilities, on-site revascularization facilities	AMI	1.03 (0.94–1.12)
Ansari (1999)	A	Age distribution, length of stay, type of admission, type of prostatectomy, location of hospital, cancer of prostate, other malignancy (except prostate), CVD, respiratory illness, endocrine disorders (including DM)	Prostatectomy	1.00 (0.87–1.14)
Basnett (1992)	C	Age, stage	Primary breast cancer	0.57 (0.44–0.75)
Burns (1991)	A	Age, sex, comorbidity, hospital ownership, bedsize, volume, location, board certification	AMI	1.70 (1.28–2.25)
			CVA, respiratory infection, respiratory failure, COPD, pneumonia, CABG, aorta repair/replacement, CHF, atrial fibrillation, colectomy, GI bleeding, major joint procedure, hip fracture, dehydration, UTI	NS in multivariate model
Carbonell (2005)	A	Age, sex, hospital bed size, laparoscopy, cholangiogram, complications, emergency surgery, race, primary expected payer, hospital region	Cholecystectomy	1.23 (0.95–1.34)
Chaudhry (2001)	C	Age, tumor size, estrogen receptor status, radiotherapy	Breast cancer—small tumor	0.47 (0.23–0.96)
			Breast cancer—large tumor	1.32 (0.73–2.32)
Cram (2004)	A	Age, race, sex, comorbid conditions	50 common diagnoses	MT-NT 1.06 (0.94–1.19), OT-NT 1.03 (0.97–1.09)
Cunningham (1999)	A	Age, sex, race/ethnicity, insurance status, hospital type, severity stage, comorbidity level, any previous admission, AIDS experience	AIDS	1.0 (0.8–1.2)
Dimick (2004)	A	Age, sex, race, indication for surgery, extent of resection, admission type, Romano modification of Charlson comorbidity score, hospital volume	Pancreatic resection	0.9 (0.6–1.3)
			Hepatic resection	1.1 (0.6–2.0)
			Esophageal resection	0.7 (0.4–1.4)
Diringer (2001)	C	Race, admitting service, primary reason for admission, GSC, number of ICH patients admitted, patient management policy, number of ICU beds, neuro vs. general ICU, ACGME fellowship, CCM fellowship	ICU	NS in multivariate model
Empana (2003)	C	Antenatal corticoids, IUGR, inborn centre, gestational age, gender, CRIB, RDS, intraventricular haemorrhage grade 3–4, assisted ventilation	NICU	1.30 (0.47–3.57)
Fleming (1991)	A	Risk-adjusted mortality index	All diagnoses	NS in multivariate model
Garber (1984)	A	Age, sex, race, residence, urgency of admission, discharge during the previous 6 months, diagnosis related group	Twelve diagnoses groups	0.75 (0.6–0.95) approximate
Gomi (2003)	C	Age, sex, performance status, grade, stage, lymphatic dissection, residual disease, radiotherapy characteristics, timing and dose, chemotherapy	Esophageal cancer	0.60 (0.37–0.98)
Haas (2003)	A	Age, sex, income, past discharges, clinical risk factors, ethnicity	Community-acquired pneumonia	0.96 (0.84–1.10)
Hartz (1989)	A	Severity of illness, occupancy rate, ownership, high-technology index, hospital beds	All diagnoses	0.96 (0.95–0.98)
Horbar (1997)	C	Apgar, birth weight, prenatal care, antenatal steroid use, gender, race, hospital volume	NICU low birth-weight	1.18 (0.94–1.47)
Hutter (2000)	A	Age, sex, diagnosis type, procedure, hospital volume	Major pancreatic resection	0.97 (0.8–1.2) approximate
Jahnigen (1985)	C	Prognosis and date of discharge matching	Long nursing home stay	0.65 (0.10–4.13)
Kelly (1986)	A	Procedure volume (hospital), procedure volume (physician), board certification, location, public hospital, expenses per day, total admissions, urban, disease stage, number of diagnoses, age, sex, insurance	Stomach operation for cancer	0.92 (0.8–1.05)
			Stomach operation for ulcer	1.00 (0.97–1.03)
			Colorectal cancer operation	0.96 (0.93–0.99)
			AAA repair	0.95 (0.88–1.02)
Khuri (2001)	A	Severity index	CEA	0.52 (0.23–1.17)
			AAA repair	1.05 (0.32–3.44)
			Infrainguinal vascular surgery	1.22 (0.50–2.97)

Table 2. Continued

Author (Year)	Data	Adjusting Factors	Conditions	Adjusted RR (95% CI)[a]
			Colectomy	1.17 (0.75–1.82)
			Open cholecystectomy	0.79 (0.35–1.79)
			Laparoscopic cholecystectomy	2.14 (0.17–27.4)
			THR	0.68 (0.17–2.75)
			Lobectomy/pneumonectomy	1.14 (0.53–2.44)
Kotwall (2002)	A	Hospital volume, age, sex, emergency operation, rural, year	Whipple procedure	1.05 (0.8–1.4) for urban
Kuhn (1994)	A	HCFA predicted mortality, other factors associated with patient risk	All diagnoses	0.95 (0.88–1.01)
Lawler (1989)	C	Age, race, physiological component of APACHE score, number of concomitant diagnoses, digestive diseases, sex, illnesses, insurance status	ICU	1.43 (1.01–2.03)
Loberiza (2005)	C	Clinical severity index, physician-patient case load/y, initial contact for after office or emergency calls	HSCT for AML, ALL, CML	MT-NT 1.43 (0.98–2.09), OT1-NT 1.43 (0.92–2.22), OT2-NT 2.35 (1.17–4.74)
			HSCT for HL, NHL	MT-NT 1.08 (0.81–1.45), OT1-NT 1.27 (0.91–1.77), OT2-NT 1.79 (1.08–2.95)
Lopez (2002)	A	Level of risk	Bariatric operations	5.0 (0.5–50)
Maxwell (2004)	A	Age	Carotid body tumor surgery	0.58 (0.35–0.95) for urban
Mayo (1989)	A	Age	Hemorrhagic stroke	MT-NT 0.59 (0.42–0.83)
Mettlin (1987)	A	Age, race, stage, histology, hospital bed size, metropolitan size, residency program, medical school affiliation, treatment modalities (specific per cancer type)	Breast cancer	0.89 (0.75–1.1) approximate
			Prostate cancer	1.01 (0.85–1.15) approximate
			Hodgkin's disease	0.91 (0.75–1.1) approximate
Pearce (1999)	A	Year, age, sex, emergency admission, LOS, type of hospital (for profit, not for profit, government, church), hospital beds, admissions per year with specific surgery, admissions per surgeon/year, vascular surgery certification)	AAA repair	1.06 (0.86–1.31)
			Lower extremity bypass graft	1.10 (0.94–1.27)
			CEA	0.90 (0.81–1.01)
Polanczyk (2002)	A	Sociodemographic, comorbid conditions	CHF	MT-NT 0.95 (0.90–0.99), OT-NT 0.98 (0.94–1.03)
			AMI	MT-NT 0.89 (0.84–0.94), OT, NT 1.04 (0.99–1.10)
			CVA	MT-NT 0.95 (0.90–1.0), OT-NT 1.08 (1.03–1.14)
Pollack (1994)	C	Pediatric risk of mortality score, endocrine disease, postoperative, intensivist hospital, pre-ICU care area, multisystem disease, oncologic disease, prehospital CPR, chromosomal anomaly	PICU	1.79 (1.23–2.61)
Reed (2001)	A	Age, sex, race, risk of death, physician specialty, location, hospital experience, treatment with IV tPA	CVA	0.88 (0.70–1.10)
Richards (1996)	A	District, age	Primary breast cancer	1.27 (0.93–1.74)
Rogowski (2004)	C	Annual admissions, NICU level, large metropolitan area, hospital ownership, insurance, gestational age, 1 min Apgar score, SGA, multiple birth, congenital malformation, vaginal delivery, prenatal care, sex, race, median income, mean education	NICU	0.98 (0.893–1.075)
Rosenthal (1997)	C	Diagnosis-tailored predicted risk of death from clinical and sociodemographic data	AMI	0.78 (0.54–1.14), OT-NT 1.23 (0.96–1.57)
			CHF	0.71 (0.54–0.91), OT-NT 1.01 (0.80–1.28)
			GI bleeding	0.95 (0.67–1.34), OT-NT 1.04 (0.66–1.63)
			COPD	0.56 (0.42–0.74), OT-NT 1.14 (0.88–1.47)
			Pneumonia	0.93 (0.73–1.20), OT-NT 1.07 (0.83–1.37)
			CVA	1.02 (0.71–1.48), OT-NT 1.18 (0.94–1.47)
Schultz (1999)	A	Percent board certified physicians, registered nurse hours/patient day, AMI volume, CABG/PTCA resources, urban density, profit status, total operating expenses/patient day	AMI	1.10 (0.56–2.18)

Table 2. Continued

Author (Year)	Data	Adjusting Factors	Conditions	Adjusted RR (95% CI)[a]
Selker (1991)	C	Age, SBP, SBPSQ, T waves, Q waves, HRT	AMI	NS in multivariate model
Silber (1992)	C	Age, sex, operation type, prior medical history, admission severity score, high technology hospital, number of hospital beds, board certified anesthesiologists, board certified surgeons	Cholecystectomy/prostatectomy	1.44 (0.9–2.4) for surgical, 1.44 (0.8–2.7) for anesthesia
Silber (1995)	C	48 patient and 11 hospital variables	CABG	0.95 (0.75–1.21)
Stross (1976)	C	Age, sex, admitting BP, temperature, WBC count, previous history of AMI, angina pectoris, CHF, DM, HTN	AMI in CCU	0.5 (0.3–0.9) approximate
Taylor (1999)	A	Age, sex, residence in the community, cognitive status, risk-adjustment score, ischemic stroke, hemorrhagic stroke, AMI, CHF, CHF with hypertension or renal disease; also peritrochanteric fracture for hip fracture analysis	Hip fracture	MT-NT 0.54 (0.37–0.79), OT-NT 0.81 (0.56–1.14)
			CVA	MT-NT 0.89 (0.64–1.24), OT-NT 1.09 (0.79–1.49)
			CAD	MT-NT 0.76 (0.55–1.07), OT-NT 0.81 (0.58–1.13)
			CHF	MT-NT 0.95 (0.64–1.41), OT-NT 1.22 (0.84–1.77)
Tingulstad (2003)	C	Residual disease	Ovarian cancer surgery	0.8 (0.5–1.2) approximate
Tingulstad (2003)	C	Age, FIGO stage, residual disease, time period, histologic type and grade, comorbidity score, CA 125	Ovarian cancer	NS in multivariate model
Todd (2004)	A	Age, sex, injury severity score, preexisting medical conditions, head trauma, primary spleen injury, assault injury, operation	Splenic trauma	1.56 (1.21–2.00) for urban, 1.82 (1.23–2.63) for rural
Whittle (1998)	A	Age, race, gender, hospital, physician, season of admission, etiology, complex score, DRG	Pneumonia 30-d mortality	1.06 (0.96–1.18)
Zimmerman (1993)	C	Reason for admission, first day APACHE III, location and duration of hospital stay prior to ICU admission, emergency surgery, LOS and hospital discharge practices	ICU	0.96 (0.91–1.01)

Based on the above, concomitant adjustment for volume/experience, severity, and comorbidity was performed in the studies by Carbonell (2005), Cunningham (1999), Dimick (2004, three comparisons), Kelly (1986, four comparisons), Pollack (1994), Rogoski (2004), Silber (1992, two comparisons), and Silber (1995).

[a]All adjusted relative risks refer to comparisons between teaching and nonteaching healthcare structures unless otherwise specified (NS in multivariate model means that the reports simply stated the lack of statistical significance, but did not provide any useful numerical values for the relative risk and its uncertainty).

A, administrative; AAA, abdominal aortic aneurysm; ACGME, American College for Graduate Medical Education; AIDS, acquired immunodeficiency syndrome; ALL, acute lymphoblastic leukemia; AMI, acute myocardial infarction; AML, acute myeloid leukemia; APACHE, Acute Physiology and Chronic Health Evaluation; C, clinical; CA 125, cancer antigen 125; CABG, coronary artery bypass grafting; CAD, coronary artery disease; CCM, Critical Care Medicine; CCU, coronary care unit; CEA, carotid endarterectomy; CHF, congestive heart failure; CI, confidence interval; CML, chronic myeloid leukemia; COPD, chronic obstructive pulmonary disease; CPR, cardiopulmonary resuscitation; CRIB, Clinical Risk Index for Babies; CVA, cerebrovascular accident; CVD, cardiovascular disease; DM, diabetes mellitus; DRG, diagnosis-related group; FIGO, International Federation of Gynecology and Obstetrics; GI, gastrointestinal; GSC, Glasgow Coma Scale; HCFA, Health Care Financing Administration; HL, Hodgkin lymphoma; HRT, heart rate; HSCT, hematopoietic stem cell transplantation; HTN, hypertension; ICH, intracerebral hemorrhage; ICU, intensive care unit; IUGR, intrauterine growth retardation; IV, intravenous; LOS, length of stay; MT, major teaching; NHL, non-Hodgkin lymphoma; NICU, neonatal intensive care unit; NS, not significant; NT, nonteaching; OT, other teaching; PICU, pediatric intensive care unit; PTCA, percutaneous transluminal coronary angioplasty; RDS, respiratory distress syndrome; RR, relative risk; SBP, systolic blood pressure; SBPSQ, squared systolic blood pressure; SGA, small for gestational age; THR, total hip replacement; tPA, tissue plasminogen activator; UTI, urinary tract infection; WBC, white blood count.

DOI: 10.1371/journal.pmed.0030341.t002

not directly relevant to the teaching status (e.g., comparison of major teaching versus minor teaching and nonteaching combined). Data on the 48 eligible articles containing adjusted mortality and their adjusting covariates are shown in Table 2. Less than half of them (20 [41.7%]) used clinical data sources. Adjusting covariates varied from 1 to 59 per study. A wide variety of diseases were represented in these studies (Table 2). The 48 articles contained 74 eligible comparisons. Of those comparisons, specific adjustments for volume/experience were performed in 30, specific adjustments for severity were performed in 59, and specific adjustments for comorbidity were performed in 35 comparisons. In 14 comparisons, all three aspects were considered in the adjustments.

Of the 61 studies that addressed other patient outcomes, 23 (37.7%) provided only unadjusted estimates and ten (15.9%) had nonusable adjusted analyses. Adjusted estimates were available for 28 (45.9%) studies (Table 3). Clinical data sources were used in 11 of them (39.3%).

Mortality

Overall synthesis of all usable adjusted mortality data yielded a summary RR of 0.96 (95% CI, 0.93–1.00; $p = 0.024$ [Figure 2]) for teaching versus nonteaching healthcare structures. Between-study heterogeneity was considerable ($p < 0.001$; $I^2 = 72\%$). However, there was no evidence that studies with smaller weight had different estimates from studies with larger weight (tau correlation coefficient –0.03; $p = 0.71$), and the same was true when administrative database studies were examined separately from clinical database

Table 3. Comparisons of Teaching versus Nonteaching Healthcare on Other Patient Outcomes Besides Mortality

Author (Year)	Data	Outcomes	Adjusted RR (95% CI)[a]
Alter (2005)	A	Readmission after AMI	0.83 (0.77–0.90)
Alter (2001)	A	Cardiac admission or ED visit after discharge	0.71 (0.56–0.90)
Ansari (1998)	A	Adverse events after prostatectomy	1.28 (1.15–1.43)
Basnett (1992)	C	Relapse of breast cancer	0.68 (0.53–0.92)
Brennan (1991)	C	All adverse events	MT-NT 2.29 (1.1–4.8) approximately; OT-NT 1.52 (??CI)
		Adverse events due to negligence	MT-NT 0.26 (0.08–0.9) approximately; OT-NT 0.92(??CI)
Cohen (1989)	A	Readmission after cholecystectomy	1.29 (1.10–1.45) for urban hospitals
Coyte (1999)	A	Revision of knee replacement	1.24 (1.03–1.50) for best scenario; 1.60 (1.08–2.36) for worst scenario
Desai (2005)	C	Satisfaction with mental health care	MT-NT $p = 0.48$; OT-NT $p = 0.92$ on continuous scale outcome
Duggirala (2004)	A	Postoperative DVT/PE	MT-NT 1.14 (1.06–1.22); OT-NT 1.08 (1.03–1.12)
		Postoperative pulmonary compromise	MT-NT 1.10 (0.97–1.24); OT-NT 1.13 (1.04–1.22)
		Postoperative pneumonia	MT-NT 1.31 (1.12–1.52); OT-NT 1.05 (0.96–1.16)
		Postoperative UTI	MT-NT 1.28 (0.86–1.91); OT-NT 1.39 (1.08–1.80)
Espehaug (1999)	A	Revision of cemented primary THR	MT-NT 1.2 (1.02–1.47); OT-NT 0.8 (0.67–0.95)
		Revision of uncemented primary THR	MT-NT 1.6 (1.13–2.19); OT-NT 1.1 (0.80–1.51)
Finkelstein (1998)	C	Patient satisfaction: physician care	−1.5 (−6.9–9.8) on 100-point scale
		Patient satisfaction: nursing care	−4.6 (−14.0–4.9) on 100-point scale
		Patient satisfaction: provision of information	−3.3 (−12.5–5.9) on 100-point scale
		Patient satisfaction: discharge preparation	−6.8 (−16.9–3.3) on 100-point scale
		Patient satisfaction: global assessments	0.3 (−13.8–14.4) on 100-point scale
Fleming (1981)	C	Patient satisfaction	$p > 0.05$ on continuous scale outcome
Garcia (2001)	A	Cesarean delivery	MT-NT 0.66 (0.55–0.76); OT-NT 1.23 (1.12–1.34)
		Primary cesarean delivery	MT-NT 0.64 (0.52–0.75); OT-NT 1.17 (1.05–1.29)
		Episiotomy	MT-NT 0.46 (0.38–0.54); OT-NT 0.88 (0.81–0.95)
		Laceration (vaginal delivery)	MT-NT 1.22 (1.07–1.37); OT-NT 1.08 (1.00–1.16)
		Laceration (any type)	MT-NT 0.86 (0.76–0.96); OT-NT 0.95 (0.88–1.01)
		Serious complications index	MT-NT 0.77 (0.61–0.93); OT-NT 0.98 (0.84–1.11)
Heck (1998)	A	Revision of knee replacement	NS in multivariate model
Jahnigen (1985)	C	Rehospitalization from nursing home	0.33 (0.11–0.96)
Jensen (2004)	C	HCV no virological response to interferon	0.7 (0.3–2.0) for private; 0.8 (0.2–3.3) for veterans
Khuri (2001)	A	Complications of CEA	1.06 (0.69–1.65)
		Complications of AAA repair	0.81 (0.45–1.45)
		Complications of infrainguinal vascular reconstruction	1.70 (1.16–2.48)
		Complications of colectomy	1.55 (1.21–1.97)
		Complications of open cholecystectomy	1.79 (1.26–2.56)
		Complications of laparoscopic cholecystectomy	2.27 (1.45–3.53)
		Complications of THR	1.09 (0.73–1.65)
		Complications of lobectomy/pneumonectomy	0.87 (0.53–1.42)
Meyers (1998)	A	Surgical complications of hysterectomy	1.86 (1.74–1.99)
		Medical complications of hysterectomy	3.28 (3.06–3.52)
Oleske (1991)	A	Primary cesarean delivery	0.76 (0.73–0.79)
Richards (1998)	A	Central line-associated bloodstream infection	NS in multivariate model
		Ventilator-associated pneumonia	NS in multivariate model
		Catheter-associated UTI	NS in multivariate model
Shaughnessy (1995)	C	Rehospitalization from nursing home	0.37 (0.2–0.7) approximately
Silber (1995)	C	Failure to rescue	1.04 (0.82–1.32)
		Complication	0.89 (0.82–0.98)
Silber (1992)	C	Adverse outcomes from cholecystectomy or prostatectomy	0.97 (0.8–1.2) for surgical; 1.05 (0.8–1.3) for anesthesia
		Failure to rescue after cholecystectomy or prostatectomy	2.05 (1.1–3.9) for surgical; 1.25 (0.6–2.4) for anesthesia
Sloan (2000)	A	Complications or death from laparoscopic cholecystectomy	1.46 (0.76–2.82)
		Complications or death from open cholecystectomy	1.05 (0.60–1.83)
		Complications or death from stomach operations	3.38 (1.19–9.64)
		Complications or death from intestinal operations	2.73 (1.82–4.08)
		Complications or death from hysterectomy	3.69 (2.54–5.37)
		Complications or death from THR	4.58 (3.04–6.63)
Thompson (2002)	C	Operative complications from cemented THR	1.12 (0.39–3.19)
		Operative complications from cementless THR	2.13 (0.83–5.44)
		General complications from cemented THR	1.02 (0.50–2.06)
		General complications from cementless THR	1.38 (0.59–3.22)
		Pain on follow-up from cemented THR	0.82 (0.48–1.40)
		Pain on follow-up from cementless THR	1.43 (0.88–2.33)
Todd (2004)	A	Laparotomy	0.72 (0.63–0.84) for urban hospitals
		Splenectomy	0.70 (0.60–0.82)
Warren (1998)	A	Rehospitalization within 7 days after simple mastectomy	HTL-NTL 0.88 (0.55–1.40); MTL-NTL 0.77 (0.43–1.38); LTL-NTL 0.91 (0.62–1.35)
		Rehospitalization within 30 days after simple mastectomy	HTL-NTL 1.06 (0.80–1.40); MTL-NTL 0.88 (0.62–1.25); LTL-NTL 0.92 (0.72–1.16)

Table 3. Continued

Author (Year)	Data	Outcomes	Adjusted RR (95% CI)[a]
		Rehospitalization within 7 days after modified radical mastectomy	HTL-NTL 1.05 (0.88–1.26); MTL-NTL 1.10 (0.89–1.36); LTL-NTL 1.03 (0.90–1.19)
		Rehospitalization within 30 days after modified radical mastectomy	HTL-NTL 1.05 (0.94–1.17); MTL-NTL 1.03 (0.90–1.18); LTL-NTL 1.01 (0.92–1.10)
Whittle (1998)	A	Readmission after pneumonia	1.02 (0.90–1.15)

[a]All adjusted relative risks refer to comparisons between teaching and nonteaching healthcare structures unless otherwise specified (NS in multivariate model means that the reports simply stated the lack of statistical significance, but did not provide any useful numerical values for the relative risk and its uncertainty).
A, administrative; AAA, abdominal aortic aneurysm; AMI, acute myocardial infarction; C, clinical; CEA, carotid endarterectomy; CI, confidence interval; DVT, deep vein thrombosis; ED, emergency department; HCV, hepatitis C virus; HTL, heavy teaching load; LTL, light teaching load, NS, not significant; MT, major teaching; MTL, moderate teaching load; NT, nonteaching; NTL, "none" teaching load; OT, other teaching; PE, pulmonary embolism; RR, relative risk; THR, total hip replacement; UTI, urinary tract infection.
DOI: 10.1371/journal.pmed.0030341.t003

studies. Comparisons between minor teaching healthcare structures and nonteaching ones yielded a RR of 1.04 for mortality (95% CI, 0.99–1.10), also with significant between-study heterogeneity ($I^2 = 60\%$).

Subgroup Analyses

With one exception, no evidence indicated that various subgroup estimates differed among themselves (Table 4). The exception was subgroups defined according to year of publication. The single study in the "year of publication" category 1971–80 gave a large benefit in favour of teaching hospitals, while summary effects from subsequent decades indicated no difference between teaching and nonteaching healthcare structures. Results were similar overall in studies using clinical versus administrative data and were not influenced by the affiliation of the corresponding author, the exact naming of the comparison, or the study location, except for a small nonsignificant trend for superior outcomes with teaching institutions in non-US studies. Heterogeneity was sizable within all of these subgroups.

In diagnosis-focused analyses, there was no between-study heterogeneity for some diagnoses, but heterogeneity persisted for others (Table 4). Certain diagnoses seemed to differ significantly between teaching and nonteaching healthcare structures: Teaching institutions seemed to perform significantly better than nonteaching ones for breast cancer, cerebrovascular accidents, and mixed diagnoses, but for the latter two diagnosis groups the magnitude of the differences was very small. Conversely, a small significant superiority in favour of nonteaching hospitals was seen for cholecystectomy. For most diagnoses, the 95% CIs excluded major differences between teaching and nonteaching healthcare structures.

Table 4. Subgroup Analyses for Mortality

Category	Subgroup	Comparisons (I^2)	RR (95% CI)
Data sources	Clinical	25 (83)	0.93 (0.84–1.03)
	Administrative	49 (59)	0.98 (0.95–1.00)
Year of publication	1971–1980	1 (not applicable)	0.50 (0.30–0.90)[a]
	1981–1990	12 (57)	1.00 (0.97–1.03)
	1991–2000	29 (80)	0.96 (0.88–1.03)
	2001–2005	32 (57)	1.00 (0.95–1.06)
Location of study	US	64 (72)	0.97 (0.94–1.00)
	Other	10 (76)	0.84 (0.70–1.02)
Affiliation of corresponding author	Academic	61 (73)	0.97 (0.93–1.02)
	Nonacademic	13 (68)	0.95 (0.90–1.00)[a]
Definition/naming of comparison	Teaching	61 (74)	0.96 (0.93–0.99)[a]
	Other naming	13 (63)	0.99 (0.87–1.12)
Types of diagnoses	Pneumonia[b]	3 (0)	1.01 (0.93–1.09)
	Pancreatic surgery	3 (0)	0.98 (0.84–1.14)
	NICU/PICU	6 (72)	1.11 (0.97–1.27)
	CVA[b]	4 (0)	0.95 (0.90–0.99)[a]
	Cholecystectomy	3 (0)	1.21 (1.02–1.43)[a]
	CHF[b]	3 (57)	0.88 (0.72–1.06)
	Breast cancer	5 (80)	0.85 (0.78–0.93)[a]
	AMI[b]	7 (87)	0.94 (0.82–1.08)
	AAA repair	3 (0)	0.96 (0.90–1.03)
	Mixed diagnoses[b]	3 (28)	0.97 (0.93–1.00)[a]

AAA, abdominal aortic aneurysm; AMI, acute myocardial infarction; CHF, congestive heart failure; CVA, cerebrovascular accident; NICU, neonatal intensive care unit; PICU, pediatric intensive care unit; RR, relative risk.
[a]$p < 0.05$.
[b]One study with nonsignificant results in multivariate model, not included.
DOI: 10.1371/journal.pmed.0030341.t004

Sensitivity Analysis

An analysis limited to the 14 comparisons in which volume/experience, severity, and comorbidity had all been specifically adjusted for yielded a summary RR of 1.01 (95% CI, 0.94–1.07). However, considerable between-study heterogeneity existed ($I^2 = 60\%$).

Among the 14 comparisons, statistically significant diagnosis-specific differences in mortality were seen in two studies. One study found significantly increased mortality in teaching paediatric intensive care units (RR 1.79; 95% CI, 1.23–2.61). Another study found a small nominally significant reduction in mortality risk in teaching hospitals for colorectal cancer surgery (RR 0.96; 95% CI, 0.93–0.99). None of the conditions for which significant benefits had been seen in teaching healthcare in the overall analysis (cerebrovascular disease, breast cancer, mixed diagnoses) were addressed by any studies with full concomitant adjustment for volume/experience, severity, and comorbidity. Conversely, one study on cholecystectomy outcomes had performed these adjustments and showed similar trends for increased death risk in teaching hospitals (RR 1.21; 95% CI, 0.95–1.34), while another study with two comparisons addressing mortality on cholecystectomy/prostatectomy procedures also found a nonsignificantly increased risk of death in teaching hospitals (RR 1.44 in both comparisons).

Other Outcomes

Types of outcomes that were addressed in the included studies varied greatly, precluding any formal quantitative synthesis. Overall, among estimates presented with odds ratios, 36 were in favour of nonteaching hospitals and 21 were in favour of teaching hospitals, but most of these estimates were not statistically significant (Table 3). None of the six continuous outcome estimates pertaining to patient satisfaction reached statistical significance.

Discussion

This systematic review of 132 studies revealed little evidence for a difference in healthcare outcomes between teaching and nonteaching settings. Observational study designs have limitations, but our results do indicate that differences in mortality outcomes between teaching and nonteaching healthcare structures, if they exist, appear small, and may not exist at all for most diseases and circumstances. Summary RR increase or decrease of 4% is well within the range of error expected from observational studies. Focusing on formal statistical significance would be misleading here [15], and the precision of effect sizes should not be overinterpreted. Given the wide diversity of these studies, the quantitative results should only be seen as suggestive, not conclusive. Results on nonmortality outcomes were even more diverse, further limiting quantitative inferences, but nonteaching hospitals did not seem to have inferior performance in most cases. Our results suggest in broad terms that teaching hospital status does not in and of itself result in major benefits nor risks for patient outcomes. In addition to the results of the combined analysis, which should be interpreted with great caution, our systematic review highlights some of the major problems in this literature.

We observed large between-study heterogeneity. This is expected, given the nonrandomized design of these studies and the variability in settings, diseases, adjusting factors, and databases used. Multiple comparisons, selective reporting, publication bias, and other study-specific biases are an additional threat to this literature and it would be difficult to probe their exact depth. We observed some small differences between teaching and nonteaching institutions for certain diagnoses. One may argue that differences in patient outcomes with teaching versus nonteaching healthcare are expected to exist only for specific diseases and settings. Focusing on subgroup analyses, however, can lead to misleading claims when the overall data are unavoidably weak due to inherent design problems.

Allowing for this caveat, better survival was seen for breast cancer and possibly cerebrovascular accidents in teaching healthcare. Unfortunately, these superior outcomes with teaching hospitals were seen in studies that did not adjust for volume/experience, severity, as well as comorbidity. Teaching centres may have better experience and closer adherence to guidelines for treating some types of cancer patients [16] and may utilize treatment more appropriately for some vascular diseases [17]. Conversely, for cholecystectomy the mortality rates were lower in nonteaching healthcare. Involvement of inexperienced trainees may not be beneficial for patients undergoing a common operation, such as cholecystectomy, in which experience is the most important factor. In another study, lack of experience of residents was also felt to underlie the relatively poor outcomes of teaching hospitals in paediatric intensive care patients [18]. However, such differential results should be interpreted very cautiously. It is impossible to adjust for all potential confounders in such studies.

Additional caveats should be discussed. We believe that some eligible studies are still missing from our evaluation, since it is very difficult to identify all articles that have attempted incidentally a cursory comparison of teaching versus nonteaching healthcare. Two relatively recent systematic reviews in this field found fewer than 25 eligible articles each, probably because of this limitation [2,3]. Focusing on studies in which teaching effects are claimed to be primary findings may bias the results in favour of teaching institutions. However, even if we missed some studies in our assessment, the data that we managed to find represent substantial evidence, and conclusions are unlikely to change.

Several studies provided unadjusted estimates of the RRs. Analyzing unadjusted estimates is problematic, since they do not consider differences in case-mix, baseline severity, and other patient characteristics. Thus, we used only adjusted estimates of the RRs for our data synthesis. Even these estimates may be biased. There is no way to correct for all possible confounding in observational designs. Volume and experience with the management of a disease are also important variables to adjust for [7–11,19]. However, even for confounding factors that seem to be very important, their exact impact is not yet fully known, and recent better-quality studies have begun to reveal the inadequacies of previous work [20]. Studies with adjustments for the most important covariates yielded similar results to the overall meta-analysis, but even here residual confounding cannot be fully excluded. Furthermore, the quality of the data may sometimes have been less than optimal, in particular when the data sources were administrative rather than clinical. Nevertheless, we found no major differences in the results of studies using clinical versus administrative data sources.

The results of this study may provide enough evidence to fuel the debate on the prospects of academic medicine [21]. Various scenarios for the future of academic medicine have been proposed, according to which academic medicine may eventually be abolished; may become more driven by public dictates; may become more privatized and corporate; may acquire a more global outlook; or may try to be as fully engaged as possible [21]. For those proposing that academic medicine can be abolished, our systematic review may be interpreted as evidence that abolishing teaching versus nonteaching distinctions likely will not affect patient outcomes on average. If public pressure becomes more important, academic medicine may focus more on neglected outcomes of indigent populations. A privatized academic medicine scenario may cause the outcomes of certain unprofitable procedures to receive little attention, while other, more profitable conditions may receive a disproportionately large amount of attention.

The net outcome effects of any change in direction of academic medicine are not easy to predict. Our review is largely limited to data from developed countries. Thus, for example, proposing that strengthening academic medicine in developing countries will not improve patient outcomes should not be done lightly. Teaching in a healthcare setting does not affect only patients and only in the immediate term; it is an integral part of medicine with benefits to patients in the long term as well. As such, teaching should be fostered to create better practitioners in the future for both academic and nonacademic centres.

Supporting Information

Protocol S1. References for the 132 Eligible Studies Cited in This Analysis

Found at DOI: 10.1371/journal.pmed.0030341.sd001 (66 KB DOC).

Table S1. Excluded Studies and Reasons for Exclusion

Found at DOI: 10.1371/journal.pmed.0030341.st001 (61 KB DOC).

Table S2. Eligible Studies Containing Only Unadjusted or Nonusable Data on Mortality

Found at DOI: 10.1371/journal.pmed.0030341.st002 (41 KB DOC).

Acknowledgments

This overview was conducted under the auspices of the International Campaign to Revitalise Academic Medicine (ICRAM) as part of the effort to appraise the evidence base on academic medicine. Members of the ICRAM Working Party are listed at http://www.bmj.com.

Author contributions. PNP, GDC, and JPAI designed the study. PNP and JPAI analyzed the data. PNP, GDC, and JPAI contributed to writing the paper.

References

1. Ayanian JZ, Weissman JS (2002) Teaching hospitals and quality of care: A review of the literature. Milbank Q 80: 569–593.
2. Kupersmith J (2005) Quality of care in teaching hospitals: A literature review. Acad Med 80: 458–466.
3. International Working Party to Promote and Revitalise Academic Medicine (2004) Academic medicine: The evidence base. BMJ 329: 789–792.
4. Berman RA, Green J, Kwo D, Safian KF, Botnick L (1986) Severity of illness and the teaching hospital. J Med Educ 61: 1–9.
5. Wyatt SM, Moy E, Levin RJ, Lawton KB, Witter DM, et al. (1997) Patients transferred to academic medical centers and other hospitals: Characteristics, resource use, and outcomes. Acad Med 72: 921–930.
6. Draper EA, Wagner DP, Knaus WA (1981) The use of intensive care: A comparison of a university and community hospital. Health Care Financ Rev 3: 49–64.
7. Halm EA, Lee C, Chassin MR (2002) Is volume related to outcome in health care? A systematic review and methodologic critique of the literature. Ann Intern Med 137: 511–520.
8. Killeen SD, O'Sullivan MJ, Coffey JC, Kirwan WO, Redmond HP (2005) Provider volume and outcomes for oncological procedures. Br J Surg 92: 389–402.
9. Gandjour A, Bannenberg A, Lauterbach KW (2003) Threshold volumes associated with higher survival in health care: A systematic review. Med Care 41: 1129–1141.
10. Hillner BE, Smith TJ, Desch CE (2000) Hospital and physician volume or specialization and outcomes in cancer treatment: Importance in quality of cancer care. J Clin Oncol 18: 2327–2340.
11. Sowden AJ, Deeks JJ, Sheldon TA (1995) Volume and outcome in coronary artery bypass graft surgery: True association or artefact? BMJ 311: 151–155.
12. Lau J, Ioannidis JP, Schmid CH (1997) Quantitative synthesis in systematic reviews. Ann Intern Med 127: 820–826.
13. Higgins JP, Thompson SG (2002) Quantifying heterogeneity in a meta-analysis. Stat Med 21: 1539–1558.
14. Begg CB, Mazumdar M (1994) Operating characteristics of a rank correlation test for publication bias. Biometrics 50: 1088–1101.
15. Egger M, Schneider M, Davey Smith G (1998) Spurious precision? Meta-analysis of observational studies. BMJ 316: 140–144.
16. Ray-Coquard I, Philip T, de Laroche G, Froger X, Suchaud JP, et al. (2005) Persistence of medical change at implementation of clinical guidelines on medical practice: A controlled study in a cancer network. J Clin Oncol 23: 4414–4423.
17. Chen J, Radford MJ, Wang Y, Marciniak TA, Krumholz HM (1999) Do "America's Best Hospitals" perform better for acute myocardial infarction? N Engl J Med 340: 286–292.
18. Pollack MM, Cuerdon TT, Patel KM, Ruttimann UE, Getson PR, et al. (1994) Impact of quality-of-care factors on pediatric intensive care unit mortality. JAMA 272: 941–946.
19. Bennett CL, Deneffe D (1993) Does experience improve hospital performance in treating patients with AIDS? Health Policy 24: 35–43.
20. Sheldon TA (2004) The volume-quality relationship: Insufficient evidence for use as a quality indicator. Qual Saf Health Care 13: 325–326.
21. Awasthi S, Beardmore J, Clark J, Hadridge P, Madani H, et al. (2005) Five futures for academic medicine. PLoS Med 2: e207. DOI: 10.1371/journal.pmed.0020207

Editors' Summary

Background. When people need medical treatment they may be given it in a "teaching hospital." This is a place where student doctors and other trainee healthcare workers are receiving part of their education. They help give some of the treatment that patients receive. Teaching hospitals are usually large establishments and in most countries they are regarded as being among the very best hospitals available, with leading physicians and surgeons among the staff. It is usually assumed that patients who are being treated in a teaching hospital are lucky, because they are getting such high-quality healthcare. However, it has sometimes been suggested that, because some of the people involved in their care are still in training, the patients may face higher risks than those who are in nonteaching hospitals.

Why Was This Study Done? The researchers wanted to find out which patients do best after treatment—those who were treated in teaching hospitals or those who were in nonteaching hospitals. This is a difficult issue to study. The most reliable way of comparing two types of treatment would be to decide at random which treatment each patient should receive. (For more on this see the link below for "randomized controlled trials.") In practice, it would be difficult to set up a study where the decision on which hospital a patient should go to was made at random. One problem is that, because of the high reputation of teaching hospitals, the patients whose condition is the most serious are often sent there, with other patients going to nonteaching hospitals. It would not be a fair test to compare the "outcome" for the most seriously ill patients with the outcome for those whose condition was less serious.

What Did the Researchers Do and Find? The researchers conducted a thorough search for studies that had already been done, which met criteria which the researchers had specified in advance. This type of research is called a "systematic review." They found 132 studies that had compared the outcomes of patients in teaching or nonteaching hospitals. None of these studies was a trial. (They were "observational studies" where researchers had gathered information on what was already taking place, rather than setting up an experiment.) However, in 14 studies, extensive allowances had been made for differences in such factors as the severity of the patients' condition, and whether or not they had more than one type of illness when they were treated. There was a great deal of variability in the results between the studies but, overall, there was no major difference in the effectiveness of treatment provided by the two types of hospital.

What Do These Findings Mean? There is no evidence to support that it is better to be given treatment in a teaching or a nonteaching hospital. The authors do note that a limitation in their analysis is that it was based on studies that were not randomized controlled trials. They also raise the question that differences might be found if considering specific diseases one by one, rather than putting information on all conditions together. However, they believe that any such difference would be small. Their findings will be useful in the continuing debate on the most effective ways to train doctors, while at the same time providing the best possible care for patients.

Additional Information. Please access these Web sites via the online version of this summary at http://dx.doi.org/10.1371/journal.pmed.0030341.

- Wikipedia entry on teaching hospitals (note: Wikipedia is a free online encyclopedia that anyone can edit)
- Information on randomized clinical trials from the US National Institutes of Health
- A definition of systematic reviews from the Cochrane Collaboration, an organization which produces systematic reviews

All of the above include links to other Web sites where more detailed information can be found.