Therapeutics and Clinical Risk Management

Dovepress
open access to scientific and medical research

Open Access Full Text Article

REVIEW

Is There a Doctors' Effect on Patients' Physical Health, Beyond the Intervention and All Known Factors? A Systematic Review

Christoph Schnelle[1], Justin Clark[1], Rachel Mascord[2], Mark A Jones[1]

[1]Institute of Evidence-Based Healthcare, Bond University, Robina, Queensland, Australia; [2]General Dentist, BMA House, Sydney, New South Wales, Australia

Correspondence: Christoph Schnelle, Institute of Evidence-Based Healthcare, Bond University, Robina, Queensland, Australia, Email christoph.schnelle@student.bond.edu.au

Video abstract

Point your SmartPhone at the code above. If you have a QR code reader the video abstract will appear. Or use: https://youtu.be/uXjR7VOXTwQ

Purpose: Despite billions of doctor visits worldwide each year, little is known on whether doctors themselves affect patients' physical health after accounting for intervention and confounders such as patients' and doctors' data, hospital effects, nor how strong that doctors' effect is. Knowledge of surgeons' and psychotherapists' effects exists, but not for 102 other medical specialties notwithstanding the importance of such knowledge.

Methods: Eligibility Criteria: Randomized controlled trials (RCTs), case-control, and cohort studies including medical doctors except surgeons for any intervention, reporting the proportion of variance in patients' outcomes owing to the doctors (random effects), or the fixed effects of grading doctors by outcomes, after multivariate adjustment. Exclusions: studies of <15 doctors or solely reporting doctors' effects for known variables.

Sources: Medline, Embase, PsycINFO, inception to June 2020. Manual search for papers referring/referred to by resulting studies.

Risk of Bias: Using Newcastle–Ottawa scale.

Results: Despite all medical interventions bar surgery being eligible, only thirty cohort papers were found, covering 36,239 doctors, with 10 specialties, 21 interventions, 60 outcomes (17 unique). Studies reported doctors' effects by grading doctors from best to worst, or by diversely calculating the doctor-attributed percentage of patients' outcome variation, ie the intra-class correlation coefficient (ICC). Sixteen studies presented fixed effects, 18 random effects, and 3 another approach. No RCTs found. Thirteen studies reported exceptionally good and/or poor performers with confidence intervals wholly outside the average performance. ICC range 0 to 33%, mean 3.9%. Highly diverse reporting, meta-analysis therefore not applicable.

Conclusion: Doctors, on their own, can affect patients' physical health for many interventions and outcomes. Effects range from negligible to substantial, even after accounting for all known variables. Many published cohorts may reveal valuable information by reanalyzing their data for doctors' effects. Positive and negative doctor outliers appear regularly. Therefore, it can matter which doctor is chosen.

Keywords: physicians' effect, practice effect, physicians' practice pattern, clinical competence, professional practice gap, delivery of health care, quality of health care, physicians

Strengths and Limitations

A strength of our review is the comprehensive literature search, using a complex and complete combination of terms for the search strategy to identify most of the relevant studies; furthermore, we screened the articles' list of references and studies citing the article for further eligible studies, with no limitations regarding the language or timeframe. In addition, it is the first systematic review providing detailed and clear reporting of the effect size, and that the doctors' effect is often substantial.

Conversely, there is a trio of limitations. First, although the scoring of the Newcastle-Ottawa Scale (NOS) for assessing the risk of bias showed that the majority of included studies scored a value of 8 or 9 (9 is the maximum total), that scale has been critiqued for being "difficult to use and [having] vague decision rules"[1] leading to poor or fair inter-rater reliability among reviewers. However, The Cochrane Collaboration[2] has endorsed its implementation in systematic reviews that

Received: 12 May 2022
Accepted: 11 July 2022
Published: 21 July 2022

include nonrandomized studies. Second, since all of the included studies were set in Europe and North America, our findings may not be applicable to other locations, particularly developing nations. Finally, among our included studies, data was reported too heterogeneously in content and presentation to allow meta-analysis.

1. What is already known on this topic: psychotherapists and surgeons are well known to have a substantial effect on patients' physical health. However, the scale of the influence of (non-surgical) medical doctors on patients' physical health, after accounting for all known confounders, is less understood. In other words, is there a doctors' effect which there is currently no explanation for?
2. What this study adds: this systematic review is considered to be the first to address the unexplained doctors' effect on patients' physical health, showing that medical doctors can be effect modifiers of interventions. Findings are highly variable, ranging from little effect through to large effects, where the latter can result in significant differences in patient's physical health outcomes, depending on the doctor, which means that it can matter which doctor is chosen.

Rationale

Each year, patients worldwide visit medical doctors billions of times, with 800 million visits in the United States[3] and 150 million visits in Australia[4] alone. However, apart from a classic[5] 1955 essay[6] that states "[T]he most frequently used drug in general practice was the doctor himself", there has been limited research on whether medical doctors, on their own, can represent an intervention or an effect modifier of interventions, ie whether different doctors who use the same intervention have differing patient's physical health outcomes, even after accounting for all known variables, including doctor demographics and patient risk factors. It is well-known that psychotherapists can have a significant effect on their patients' mental health, an effect that equals the strength of pharmaceutical interventions and is mentioned in training manuals.[7] It is also known that surgeons, after accounting for all known information,[8] do have a widely varying effect on patients' physical health. Therefore, it would be useful to know whether this applies to other medical doctors, as a fundamental question in medical research is what effect the medical practitioner has on patients' physical health. The doctor certainly has an effect by choosing and applying the intervention, but it is less clear whether the effect goes beyond the intervention, and whether doctors constitute an intervention in their own right.

Research on general doctors' performance has concluded that it is difficult to assess practice variation among doctors and therefore, it is often not worthwhile to direct quality improvement efforts at this level of medical services.[9,10] However, some doctors were found to be more effective than others at employing interventions, owing, for example, to a substantial volume or practice effect in many surgical specialties.[11,12] Recent evidence also proposes that patients' outcomes can be substantially affected by provider expectations.[13] In other non-surgical specialties, research conducted on doctors' effects is scarcer, with evidence limited to primary care,[14,15] obstetrics,[16] and acute care,[17] in which physicians' factors point to a sizeable effect on patients' health outcomes. Thus, a significant doctors' effect detected indicates that there are doctors who perform better than others. Many initiatives aimed at improving medical standards aim to identify underperformers to either remove them from medical practice or propose strategies to improve their standards.[18–20] However, there seems to be no systematic review that answers a more basic question: Are there differences among doctors which contribute to creating an effect on patients' physical health outcomes, even when all known factors have been accounted for?

In a kitchen, it would be obvious that cooks using the same ingredients have widely varying outcomes. In law, practitioners charge widely varying rates, with clients presumably assuming that the most expensive lawyers are so much better than the average lawyer that they are worth their higher fees. No such presumption of substantial differences between doctors seems to exist in medicine as an established research fact.

If we know whether medical doctors can differ widely in their performance, then we can find out under what circumstances the effect is large or small, important, or unimportant. In addition, we can check whether there are positive and negative outliers among doctors, allowing health care services to support the negative outliers to improve, if possible, and to learn from the positive outliers, and, if needed, make sure that they are treated with the care and respect such exceptionally good doctors deserve.

This systematic review gives the answer to precisely this question: What research has been published that shows whether doctors, on their own, have an effect on patients' physical health outcomes, after taking into account all known information? Known information can consist of patient demographics and risk factors, intervention, doctor demographics such as age, specialization, education and experience, and hospital or area effects such as county or country effects.

This review further looks at the quality of the publications and their heterogeneity, and whether reporting on doctors' performance can be improved and prepared for meta-analysis. It may seem ambitious to cover 102 non-surgery medical specialties[21] in a single publication but such is the paucity of this material – despite the billions of interactions of medical doctors each year – that the number of publications found do fit into a single systematic review. Future reviews may be more focused, but an overarching review is the first step, due to the current lack of any review.

What is the Current State of Research?

In 2002, the British Medical Journal (BMJ) devoted an entire issue to the following question: "What's a good doctor and how do you make one?"[22] assuming that it would be useful to know what a good doctor is. In this special edition, one article presented letters from doctors and others attempting to answer these questions. One quote stated: "There is not a single piece of evidence or the means to measure whether a doctor is good or bad."[23] The editorial of that 2002 issue stated

> (…) defining a good doctor, I suggest, lies in degree of difficulty somewhere between defining a good composer and a good human being. In fact, it's impossible.

Hospitals are known to substantially influence patients' physical health outcomes and hospital performances regarding patients' physical health outcomes vary widely.[24–29] The same is true for larger entities like regions or countries where mortality rates can differ substantially.[30]

Recent research has investigated 10 surgical trials, in which the effect size of surgeons was analyzed to assess the surgeon intra-cluster correlation coefficients (ICCs), ie the percentage of the whole patient outcome variation due to the surgeon. It revealed that surgeons alone are responsible for a range of effects on patients' health outcomes, which vary between different surgical specialties.[31]

Objectives

This systematic review examines the existing literature on measuring and reporting doctors' effects on patients' physical health after adjusting for known factors for medical doctors that are not surgeons. Psychotherapists are here not considered to be medical doctors.

Methods

Eligibility Criteria

This systematic review follows the standards set for Synthesis Without Meta-analysis (SWiM).[32] Only studies that investigated actual patients' physical outcomes were included. Scientific publications that reported patients' opinions or their satisfaction levels were excluded as these are not patients' physical health outcomes and often less reliable measurements.[33] The study PICO is as follows:

Population	**P**	**Medical Doctors That are Not Surgeons**
Intervention	I	Any
Comparison	C	Not applicable
Outcome	O	Practitioners' effect on patients' physical health outcome

https://doi.org/10.2147/TCRM.S372464

Information Sources and Search Strategy

We conducted a comprehensive search on the following databases: Embase, Medline via PubMed, and PsycINFO, to retrieve pertinent studies that investigate the doctors' effect on patients' physical health outcomes, from inception until June 2020. The search strategy was designed and developed for each database by JMC, a search specialist (Supplemental File 1). In addition, using the references lists of the selected articles and former reviews we manually searched for potentially related studies that may have been missed in the initial literature search. Furthermore, systematic review registries including PROSPERO and Cochrane's CENTRAL register were searched for similar reviews.

Selection Process and Further Eligibility Criteria

Two review authors independently screened the titles and abstracts of all retrieved records. Any disagreements were resolved via discussions and consultation with a third reviewer. We included any case-control study, retrospective or prospective cohort study, or randomized controlled trial (RCT) that graded individual doctors according to their performance regarding the patients' physical health outcomes, or where the percentage of the variance in patients' outcomes is explained by differences between doctors. All outcomes related to patients' physical health were eligible, for example survival/mortality rate, repair reoperations, hospitalization rates, length of post-procedure stay, readmission rate, post-operative complications, pain, infection rate, embryo transfer rate, blood pressure, cholesterol, and glycemic control. Surgeons were excluded from this review as they were reviewed in a separate paper.[8] No restrictions were placed on publication date or language.

We excluded studies that address only doctors' effects related to specific known doctor-related variables, such as the doctor's specialty or the volume of procedures performed. Studies including fewer than 15 doctors and cross-sectional studies were also excluded, due to their increased risk of bias.

No authoritative source was found to provide a reference for the smallest number of clusters required for a reliable ICC estimation. Here the number of referred-to clusters is the minimum number of practitioners to warrant inclusion. We used 15 as a minimum number but realize this is somewhat arbitrary (Figure 1).

Data Collection Process and Data Items

We used Endnote 9 for exporting the titles and abstracts of retrieved records, which were then uploaded into Rayyan for screening. Then the potentially eligible records were marked as members of a group in the original Endnote library and their full text documents added to the library for further full-text screening.

From each final included study, CS and a second extractor independently and in duplicate, extracted the relevant data into an excel sheet including the following variables:

- Study ID consisting of the first author's last name and year of publication
- Type of study (RCT, Cohort)
- Country of origin
- Medical specialty
- Type of intervention(s)
- Patients or procedures
- Number of doctors
- Number of hospitals or institutions
- Outcome type(s)
- Number of positive and negative outliers
- Authors' evaluation of significant doctors' effect Y/N
- Multivariate analysis Y/N
- ICC (intra-class correlation coefficient)

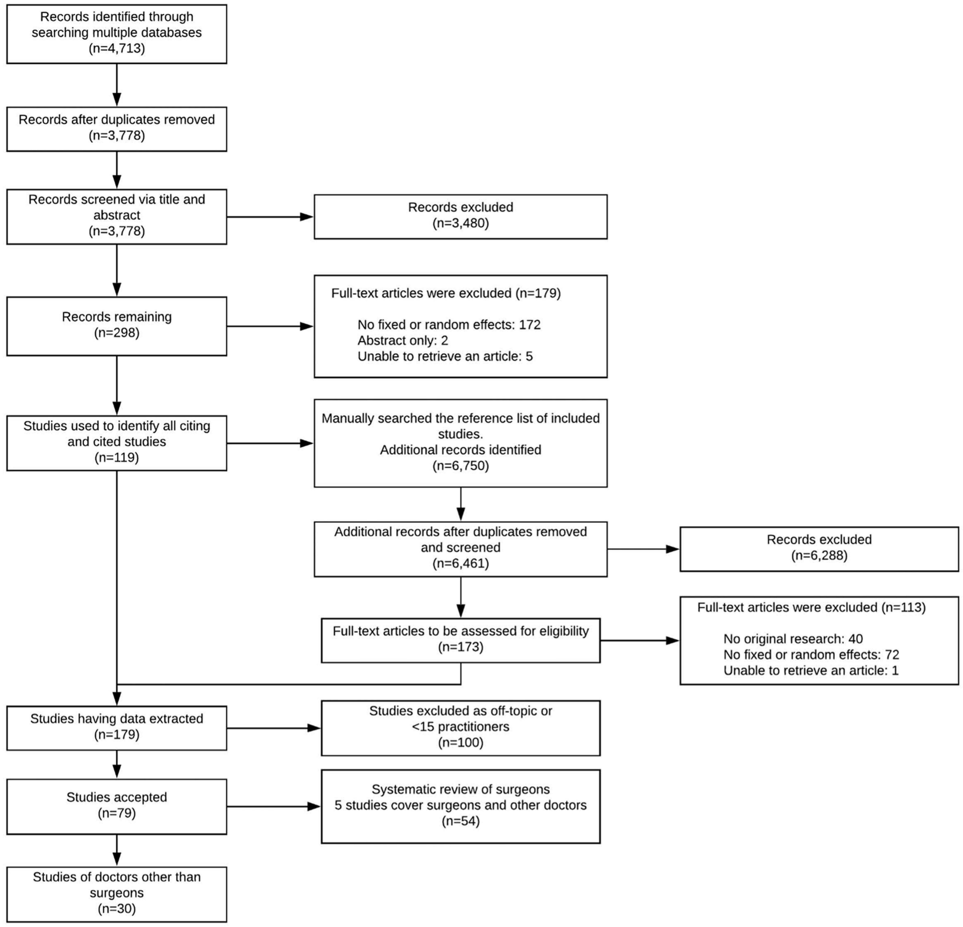

Figure 1 Flow diagram of selection of included documents.

Risk of Bias Assessment

Two review authors independently assessed the risk of bias of all included cohort studies, using the Newcastle-Ottawa Scale (NOS).[34,35]

Effect Measures

One pathway of evaluating doctors' performance is to measure fixed effects, which are covered as a statistical technique by Allison.[36] Fixed effects allow the identification of high and low outliers and give an impression of how heterogeneously doctors perform in a particular area. Grading doctors also shows whether the variation in effect is consistent with chance or bigger than that. The metric for the fixed effects in this study is the percentages of negative and positive outliers, as defined and reported per each individual study.

The other method of assessing a doctors' effect is by measuring random effects, also explained by Allison.[36] Random effects measure the variation in patient outcomes that is due to the doctor beyond known factors, such as their level of experience. Likewise, these effects cannot be explained by differences in diagnostic prowess or choosing more or less suitable interventions. Random effects allow the discernment of how much doctors may constitute an intervention in their own right. That measurement is called the intra-class correlation coefficient (ICC). Examples are mortality in intensive care,[37] or levels of uncontrolled hypertension,[38,39] or high HgA1c levels[18,38,40–42] among patients of family medicine doctors or general practitioners.

The ICC is here described as the proportion of patients' health outcomes that resulted from the doctor's effect, in the form of a percentage of the total patient outcome variation. The significance of even small ICCs is covered in the Discussion.

Synthesis Methods

The identification of doctors' effects on patients' health outcomes is presented in many different ways that can be classified into two methods. Both methods either use hierarchical regression or multilevel mixed effects regression modelling to understand both doctor and higher-level variation.[43,44]

Percentage of Variation in Patient's Health Outcome

The percentage of variation in patients' health outcome owing to the doctor is reported as the intra-class correlation coefficient (ICC). The ICC, which can be identified through post-regression estimation, is a number ranging from 0 to 1 representing the percentage of variation in a particular outcome due to each level in the regression model. Therefore, to enable the allocation of the percentage of variation owing to the doctor, random effects for doctors and occasionally for hospitals or other higher-level aggregators such as county, are included in the studies.[38,39,45–47]

The regression analyses included patient risk scores and other known confounders such as doctors' demographics as fixed effects. There was a pronounced variance in the depth and quality of the analysis between different studies, with Papachristofi et al as a high quality example.[47] In addition, further extensive literature is available addressing the ICC.[48–53]

Grading Doctors from Best to Worst

Regarding this approach, doctors are ordered according to the patients' physical health outcomes, typically with a 95% confidence interval (CI). This CI is calculated using, for example, cluster-robust standard errors,[54,55] or other means such as simulation,[16] or the delta method.[38] Doctors are considered to be outliers when their 95% CI is wholly above or below the mean rate of the patients' outcomes. Consequently, results are reported by listing the outliers in order, or as a funnel or caterpillar plot,[56] with the latter constituting an outcome-ordered forest plot.

Reporting Bias and Certainty Assessment

Since meta-analysis was not applicable, we did not assess the reporting bias nor conducted certainty assessments.

Results

Study Selection

We retrieved 4713 records from electronic searches, reduced to 3778 after removing duplicates, and 119 after screening. Manually searching the reference list of these studies yielded an additional 6750, reduced to 60 after screening. The resulting 179 studies were reviewed in full, yielding 79 accepted studies of which 30 applied to doctors other than surgeons. These 30 studies with 36,239 doctors met our pre-specified criteria for inclusion in the final synthesis (Figure 1).

Study Characteristics

The final 30 included studies either graded individual doctors from best to worst according to their performance (N=9),[16,18,57–63] or recorded a residual variation owing to doctors in a multivariate multi-level analysis yielding an ICC (N=11),[38–42,45,46,64–67] or both (N=7),[37,47,68–72] or used a different way to describe their results (N=3).[17,73,74] Jemt et al[74] used a different approach but also listed one positive and two negative outliers.

All 30 studies were observational cohort studies that included doctors from multiple specialties, such as general practitioners, family doctors, or primary care physicians (N=11),[18,38–42,45,46,59,62,65] anesthesiologists (N=4),[47,68,70,71] cardiologists (N=4),[58,60,61,67] hospitalists or residents (N=7),[17,37,62,63,66,69,72] and one each of dentistry,[74] gynaecology,[16] pathology,[64] paediatrics,[46] radiology,[73] and reproductive medicine.[57] (N=18) studies were conducted in the USA,[18,37–42,45,58,59,61,62,64,67–69,72,73] (N=7) in the UK,[16,17,47,60,66,70,71] and one each in Canada,[63] Italy,[57] Netherlands,[65] Spain,[46] and Sweden.[74] The number of included doctors ranges from 21 to 4230. Table 1 summarizes the characteristics of the included studies.

Risk of Bias Assessment

Among the 59 outcomes in the included 30 studies, (N=48) scored 9 stars, (N=10) 8 stars, and (N=1) 7 stars, with a maximum possible score of 9 stars on the Newcastle-Ottawa Scale.[34,35] Those of 7 and 8 stars scored either 0 or 1 on the aspect of

Table 1 Characteristics of Included Studies

Publication	Practitioner	Specialty	Detailed Intervention	Doctors	Patients/ Procedures	Institutions	Outcome	NOS*
Becerra, 2017[64]	Pathologist	Colorectal surgery/ Pathology	Lymph node examination after colectomy	814	12,332	187	Suboptimal care	9
Beckett, 2018[17]	Hospitalist	Acute care	Acute care	22	21,570	1	Mortality	8
							Readmission	8
Brown, 2016[18]	GP General Practitioner	Primary care	Diabetes glucose control	133	14,033	84	Avoiding uncontrolled diabetes	9
Cirillo, 2020[57]	OB-GYN doctor or senior residents	Obstetrics	Embryo transfer	32	19,824	1	Ongoing pregnancy	9
Davenport, 2020[73]	Radiologist	Radiography	Headache CT	55	25,596	1	Mortality	9
							Readmission	9
Eijkenaar, 2013[65]	GP General Practitioner	Primary care	Primary care	447	26,684		COPD**-related admissions	8
				537	37,832		Diabetes-related admissions	8
Glance, 2016[68]	Anesthesiologist	Cardiac surgery	Cardiac surgery	357	55,436	40	Major complications or mortality	9
Goodwin, 2013[69]	Hospitalist	Acute care	Acute care	1099	129,491	268	Length of stay	9
				1099	131,710	268	Mortality	9
Gossl, 2013[58]	Cardiologist	Cardiology	Percutaneous coronary intervention	21	7838	3	MACE Major adverse cardiac event inc. death	9
							Mortality	9
Gutacker, 2018[66]	Hospitalist	Emergency Care	AMI Acute myocardial infarction	1746	138,044	148	Length of stay	9
							Mortality	9
							Readmission	9
		Cardiac surgery	CABG*	212	24,505	30	Length of stay	9
							Mortality	9
							Readmission	9
		Pneumonia	Pneumonia	3760	405,671	152	Length of stay	9
							Mortality	9
							Readmission	9
		Stroke	Stroke	1214	144,114	144	Length of stay	9
							Mortality	9
							Readmission	9
		Orthopedic surgery	Hip fracture	1735	156,145	148	Length of stay	9
							Mortality	9
							Readmission	9
			Hip replacement	1325	170,678	229	Length of stay	9
							Mortality	9
							Readmission	9
Hannan, 2017[67]	Cardiologist	Cardiology	Percutaneous coronary intervention	403	27,560	60	Incomplete revascularization	9
Harley, 2005[16]	Gynecologist	Obstetrics	Gynecologists' performance	143	Not stated	Multiple	7-item composite measure	7
Hofer, 1999[45]	GP General Practitioner	Primary care	Diabetes glucose control	232	3642	3	Hospitalizations	9

(*Continued*)

Table 1 (Continued).

Publication	Practitioner	Specialty	Detailed Intervention	Doctors	Patients/ Procedures	Institutions	Outcome	NOS*
Holmboe, 2010[38]	GP General Practitioner	Primary care	Cholesterol control	236	22,526	13 states	Avoiding high cholesterol	9
			Diabetes glucose control	236	22,526	13 states	Avoiding uncontrolled diabetes	9
			Hypertension control	236	22,526	13 states	Avoiding uncontrolled Hypertension	9
Jemt, 2016[74]	Dental Surgeon	Dental Implants	Dental implants	23	8808	1	Implant failure	9
Kaplan, 2009[59]	GP General Practitioner	Primary care	Diabetes glucose control	210	7574		10 quality measures	8
			Cholesterol control	210	7574		10 quality measures	8
Krein, 2002[40]	GP General Practitioner	Primary care	Cholesterol control	258	12,110	9/13	Avoiding high cholesterol	8
			Diabetes glucose control	258	12,110	12/13	Avoiding uncontrolled diabetes	8
Kunadian, 2009[60]	Cardiologist	Cardiology	Percutaneous coronary intervention	261	149,888	48	Mortality	9
Navar-Boggan, 2012[61]	Cardiologist	Cardiology	Hypertension control	47	5979	1	Avoiding uncontrolled hypertension	9
O'Connor, 2008[41]	GP General Practitioner	Primary care	Diabetes glucose control	120	2589	18	Avoiding uncontrolled diabetes	8
Orueta, 2015[46]	GP (Family doctors)	Primary care	Avoidable hospitalization	1193	2,207,175	130	Hospitalization rates	9
	Pediatrician	Primary care	Avoidable hospitalization	286			Hospitalization rates	9
Papachristofi, 2014[70]	Anesthetist	Cardiac surgery	Cardiac surgery	24	18,426	1	Mortality	9
Papachristofi, 2016[71]	Anesthetist	Cardiac surgery	Cardiac surgery	190	110,769	10	Mortality	9
Papachristofi, 2017[47]	Anesthetist	Cardiac surgery	Cardiac surgery	190	107,038	10	Length of stay	9
Prasad-Kerlin, 2018[37]	Hospitalist	Acute care	Mechanical ventilation	345	11,268	104	Mortality	9
Selby, 2010[39]	GP General Practitioner	Primary care	Cholesterol control	1005[a]	169,156	35	Avoiding high cholesterol	9
			Hypertension control	1049[b]	232,053	35	Avoiding uncontrolled hypertension	9
Singh, 2015[72]	Hospitalist	Primary care	Primary care	525	48,883	143	Readmission	9
Singh, 2019[62]	GP General Practitioner	Primary care	Primary care	4230	565,579		Hospital readmission	9
Tuerk, 2008[42]	GP General Practitioner	Primary care	Diabetes glucose control	42	1381	1	Avoiding uncontrolled diabetes	8
Verma, 2020[63]	Hospitalist	Acute care	Emergency admissions, inpatient care	135	103,085	7	Length of stay	9
							Mortality	9
							Readmission	9

Notes: If "Institutions" is blank, then the number is not applicable (GPs, General Practitioners, for example), or not given and most likely greater than one. *CABG is coronary artery bypass graft. **COPD is chronic obstructive pulmonary disease. For cholesterol, diabetes, and hypertension management, outcomes were standardized to avoiding high cholesterol/HbA1C/blood pressure. [a]Single year numbers. Totals for 6 years are 6,832 doctors, 1,588,407 patients. [b]Single year numbers. Totals for 6 years are 6,995 doctors, 2,021,935 patients.
Abbreviation: NOS, Newcastle-Ottawa Scale for assessing risk of bias of cohort studies.

comparability, whereas the studies with 9 stars scored 2. All included studies scored the maximum points regarding the selection and the outcome criteria (Table 1).

Results of Individual Studies

Altogether 15 studies with 21 outcomes published caterpillar plots or plots that gave the same information.[17,18,37,47,57,58,61–63,65,68–72] One paper showed funnel plots.[60] Such plots represent and sort the doctors' performance for a specific patient outcome, usually showing a 95% confidence interval (CI) for each doctor and whether that CI was wholly below or above the mean performance rate. Results varied from no over- or underperformer[70,71] up to substantial numbers of both.[37,59,61,63,72]

Of the 16 studies that show fixed effects, 11 reported one or more exceptional performers after accounting for all known confounders, including doctors' demographic variables such as their years of experience and volume of procedures/patients, and the at times substantial hospital effects.[16,18,37,47,58–61,63,69,74] Two papers[57,62] found only negative outliers. Three papers found no positive or negative outliers.[68,70,71] (Table 2).

A few papers (N=18) presented a random effect, reported in many different ways, which express the intra-class correlation coefficient (ICC), ie the variation due to the doctor as a percentage of the whole variation in patient physical health outcomes, with that variation calculated while accounting for all available patient, doctor, or institution variables.[37–42,45–47,62,64–71] Reported random effects ranged from approximately zero (ICC of 0.0%) to substantial (ICC up to 33%, median of 1.9%, mean of 3.9%, inter-quartile range 1.0–4.2%) (Figures 2 and 3, Table 2).

Only cholesterol,[38–40] diabetes,[40–42] and hypertension[38,39] control outcomes had more than one study each for the same medical specialty and intervention. ICCs range from 0% to 2%, except Holmboe et al[38] who found much higher ICCs of 12% and 9%. The main difference between this and the other studies is that Holmboe's cohort consisted of doctors who volunteered to participate (Table 2). In nine instances, the ICC was between 9% and 33%.

Reporting Bias, Syntheses, and Certainty of Evidence

Not applicable since there was no statistical synthesis of the results.

Discussion

The findings from this systematic review indicate that doctors have an effect on patients' physical health, even after taking into account all known variables or confounders. This effect ranges from zero to substantial with nine instances where the doctor was associated with at least 9% of the total variation in patient health.

In terms of the effect of even small ICCs, a randomized controlled trial[75] that established the prophylactic value of aspirin was halted early as it was considered to be unethical to withhold aspirin from the control group, even though aspirin only accounted for 1% of the variability in outcomes, ie the trial was halted for a treatment with an ICC of 1%. Further, even a "small" doctors' effect makes a substantial difference in patient health as that difference is applied billions of times each year in each doctor-patient interaction. The value and importance of even small ICCs is further outlined in these three publications.[7,76,77]

At times doctors can be identified whose performance is substantially above or below the average performer. Therefore, a possible answer to the question, "What's a good doctor and how do you make one?"[78] is, "A good doctor is a doctor with significantly better patient physical health outcomes than the average doctor." In addition, a possible answer to, "and how do you make one?" could be,

> Good doctors already exist and can be identified. Unless good doctors' abilities are wholly innate, more good doctors can be made by learning from those who already are good doctors, and exceptionally good doctors also exist.

The key here is that an effect with an unknown cause has been identified. The cause could be anything unmeasured in the included cohort studies, such as doctors' communication skills, their level of care for patients, their physical or mental health, the time they give to a patient, their ability to listen to a patient, their diagnostic ability (as a more suitable intervention is more likely to yield better outcomes), their ability to perform under stress etc. This is an avenue for further research.[79,80]

It is noteworthy that no included study identifying exceptionally good doctors made recommendations on how to use this resource. The substantial number of positive outliers are at times not mentioned in the text, only shown in the graph.

Table 2 Publications by Outcome and Numerical Results

Outcome	Practitioner	Specialty	Publication	ICC^	Outliers %	
					Negative	**Positive**
Avoiding high cholesterol	GP General Practitioner	Primary care	Holmboe, 2010[38]	12.0%		
			Kaplan, 2009[59]	9.0%		
			Krein, 2002[40]	1.0%		
			Selby, 2010[39]	1.9%		
Avoiding uncontrolled diabetes	GP General Practitioner	Primary care	Brown, 2016[18]		6.0%	6.8%
			Holmboe, 2010[38]	9.0%		
			Kaplan, 2009[59]	33.0%		
			Krein, 2002[40]	0.0%		
			O'Connor, 2008[41]	0.8%		
			Tuerk, 2008[42]	2.0%		
Avoiding uncontrolled hypertension	Cardiologist	Cardiology	Navar-Boggan, 2012[61]		6.4%	12.8%
	GP General Practitioner	Primary care	Holmboe, 2010[38]	9.0%		
			Selby, 2010[39]	1.9%		
Complications						
MACE Major adverse cardiac event inc. death	Cardiologist	Cardiac surgery	Gossl, 2013[58]		0.0%	4.8%
Major complications or mortality	Anesthesiologist	Cardiac surgery	Glance, 2016[68]	0.5%	0.0%	0.0%
Hospitalizations						
COPD**-related admissions	GP General Practitioner	Primary care	Eijkenaar, 2013[65]	2.5%		
Diabetes-related admissions	GP General Practitioner	Primary care	Eijkenaar, 2013[65]	0.6%		
Hospitalizations	GP General Practitioner	Primary care	Hofer, 1999[45]	1.0%		
			Orueta, 2015[46]	6.1%		
	Pediatrician	Primary care	Orueta, 2015[46]	10.3%		
Length of stay	Anesthetist	Cardiac surgery	Papachristofi, 2017[47]	0.2%	2.1%	0.5%
	Hospitalist	Acute care	Goodwin, 2013[69]	2.6%	19.5%	18.0%
			Verma, 2020[63]		18.5%	14.8%
		Cardiac surgery	Gutacker, 2018[66] (Heart attack)	6.5%		
			Gutacker, 2018[66] (CABG*)	5.2%		
		Pneumonia	Gutacker, 2018[66]	2.1%		
		Stroke	Gutacker, 2018[66]	1.5%		
		Orthopedic surgery	Gutacker, 2018[66] (Hip fracture)	3.2%		
			Gutacker, 2018[66] (Hip replacement)	12.7%		

Mortality	Anesthetist	Cardiac surgery	Papachristofi, 2014[70]	0.1%	0.0%	0.0%
			Papachristofi, 2016[71]	0.3%	0.0%	0.0%
	Cardiologist	Cardiology	Gossl, 2013[58]		0.0%	4.8%
			Kunadian, 2009[60]***	0.2%	0.0%	0.4%
	Hospitalist	Acute care	Beckett, 2018[17]* (different presentation)			
			Goodwin, 2013[69]	0.8%	1.5%	0.6%
			Prasad-Kerlin, 2018[37]	1.8%	22.6%	25.5%
			Verma, 2020[63]		1.5%	5.2%
		Cardiac surgery	Gutacker, 2018 (Heart attack)	1.4%		
			Gutacker, 2018[66] (CABG*)	0.9%		
		Pneumonia	Gutacker, 2018[66]	1.2%		
		Stroke	Gutacker, 2018[66]	1.1%		
		Orthopedic surgery	Gutacker, 2018[66] (Hip fracture)	1.2%		
			Gutacker, 2018[66] (Hip replacement)	0.3%		
	Radiologist	Radiography	Davenport, 2020[73]			
Pregnancy	Reproductive doctor	Obstetrics	Cirillo, 2020[57]		3.1%	0.0%
Readmission	Hospitalist	Acute care	Beckett, 2018[17] (different presentation)			
			Verma, 2020[63]		0.7%	3.0%
		Cardiac surgery	Gutacker, 2018[66] (Heart attack)	0.4%		
			Gutacker, 2018[66] (CABG*)	0.8%		
		Pneumonia	Gutacker, 2018[66]	0.4%		
		Primary care	Singh, 2015[72]	15.0%	12.8%	12.5%
			Singh, 2019[62]		0.02%	0.00%
		Stroke	Gutacker, 2018[66]	0.8%		
		Orthopedic surgery	Gutacker, 2018[66] (Hip fracture)	0.7%		
			Gutacker, 2018[66] (Hip replacement)	2.5%		
	Radiologist	Radiography	Davenport, 2020** (different presentation)			
Suboptimal care	Pathologist	Colorectal surgery	Becerra, 2017[64]	22.5%		
Success or failure						
Implant failure	Dental Surgeon	Dental Implants	Jemt, 2016[74]		8.7%	
Incomplete revascularization	Cardiologist	Cardiac surgery	Hannan, 2017[67]	12.0%		
Multiple measures						
10 measures	GP General Practitioner	Primary care	Kaplan, 2009[59]		27.8%	43.8%
7-item composite measure	Gynecologist	Obstetrics	Harley, 2005[16]		6.3%	2.1%

Notes: The leftmost column is by patients' physical outcome with summarized outcome bold. GP is General Practitioner or Primary Care Physician. ^ICC is Intra-class correlation coefficient that shows percentage of variance owing to the practitioner in the form of percentage of total variance after taking into account all known confounders. *CABG is coronary artery bypass graft. **COPD is chronic obstructive pulmonary disease. Outliers are listed for studies which ordered the doctors according to their effect on patient's physical health outcomes from worst to best or vice versa. The percentages listed represent those practitioners whose 95% CI is wholly below or above the mean. Beckett et al,[17] Davenport et al,[73] and Jemt et al[74] presented their data in a way that only fits partially or not at all in the table. Blank entries under ICC or Outliers mean that the Publication did not report those measures. All studies that reported Outliers except Verma et al[63] adjusted their results for other factors like patient risk. *Beckett et al[17] reported fixed effects before and after case-mix but no 95% confidence intervals. **Davenport et al[73] reported fixed effects with no effect on mortality but other, more indirect measures. ***Kunadian et al[60] did not publish the number of cardiologists or outliers directly, though Figure 2 in the paper is a funnel plot. The paper's reference 7 provides the original data https://www.health.ny.gov/statistics/diseases/cardiovascular/docs/pci_2002-2004.pdf which shows 146,775 cases with 261 cardiologists of whom 7 were underperformers and 5 outperformers. With one exception, all cardiologists with fewer than 31 cases were grouped as "all others". Cardiologists had one entry in the table for each hospital they worked in. The authors of this systematic review calculated the ICC for this dataset to be 0.17%, 95% CI 0.11%, 0.26%. Mortality is 945/146,775 or 0.64%.

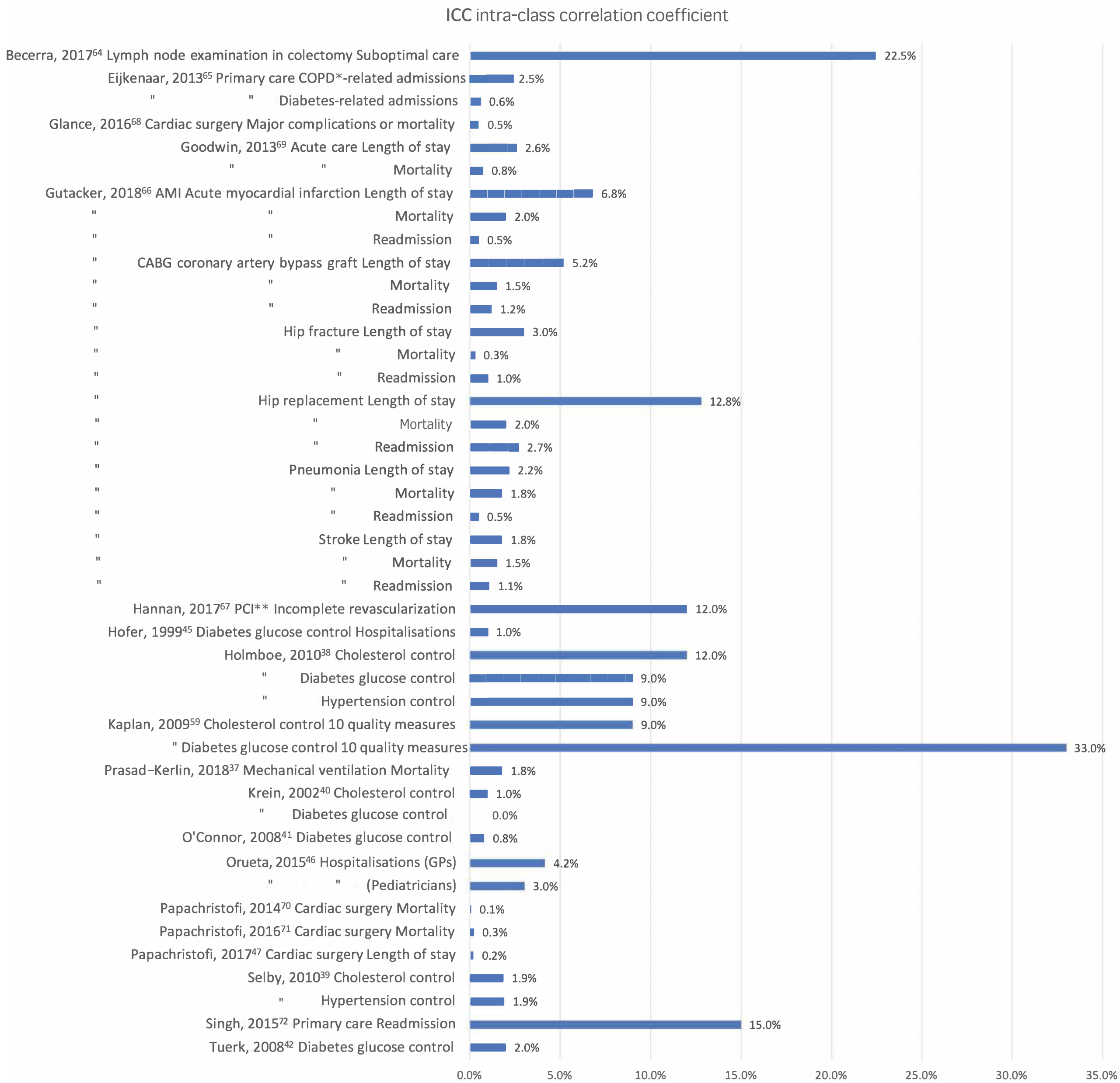

Figure 2 Intra-class correlation coefficients (ICCs) by paper, intervention, and patient outcome. *COPD is chronic obstructive pulmonary disease. **PCI is percutaneous coronary intervention.

The closest to an investigation of high performers was presented by Brown et al[18] who found that in diabetes control, high performing doctors were more likely to be female and underperforming doctors' patients were more likely to be male. Goodwin et al[69] found that hospitalists' patients' length of stay did not affect other patient outcomes. In other words, hospitalists whose patients had shorter lengths of stay in hospital had the same outcome as patients of hospitalists who were underperformers, but no further investigation was undertaken. As one contributory factor to doctors' performance, recent research has proposed that even health care provider expectations can have a substantial placebo effect on patient outcomes, ie patient outcomes can be affected through "social transmission".[13]

Many of the publications excluded for this systematic review among the approximately 10,000 studies were large-scale cohort studies where doctors' effects were attributed to one or more characteristics. However, this attribution was done without reporting the variation in patients' physical health outcomes that was due to the doctor after accounting for

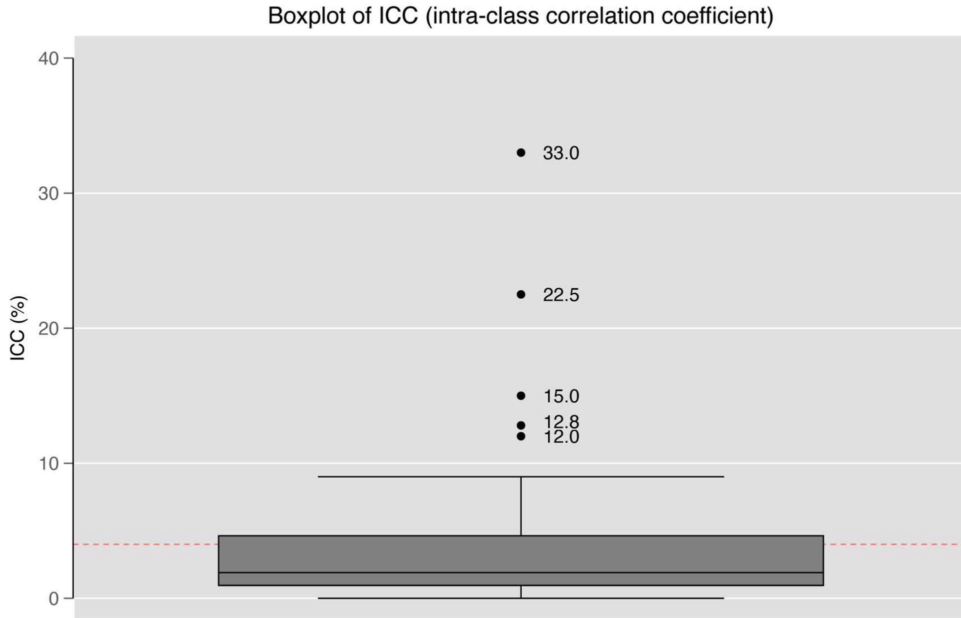

Figure 3 Boxplot of ICC (intra-class correlation coefficient).

all known risk factors. It would be relatively simple to re-analyze these and other already cleaned up and prepared datasets for such a residual effect. Publishing ICCs, ie the amount of variation due to doctors in a consistent way, will make future meta analyses possible. The authors have prepared a methodological review for this purpose.[81]

To the authors' surprise, re-analyzing existing data is not useful for many randomized controlled trials as no data register the authors contacted had any way to identify trials that showed a doctors' effect. Further, a clinical trial specialist privately told the authors that in large randomized trials, with many treatment centers, only the center identifier and not the individual doctor identifier is recorded, making it difficult or impossible to extract a doctors' effect from the data even though it would substantially affect the sample size needed for clinical trials when there are differences among medical doctors, as this would subsequently affect the RCTs statistical power.[82]

Research that addresses the doctors' effect on patients' health outcomes seems to be a form of investigation that is in its infancy. There are no established standards on how to report a doctors' effect, and results are heterogeneous indeed.

The authors found very little systematic research on the probability that doctors, in their own right, may be an intervention whose effect on patients' outcomes can be measured and be more or less effective. This is surprising since there is a well-known clustering effect with patients who have the same doctor tending to have more similar outcomes than patients of a different doctor.[31,83] Likewise, it is well established in psychotherapy that psychotherapists, in their own right, can constitute an intervention, which is independent of the actual intervention used.[7,84]

Summary

Given the increasing difficulty with identifying effective new interventions[85–87] and the increasing cost of research, it may be worth looking beyond the intervention to the other two components of a medical treatment, viz. the doctor and the patient. If there are substantial differences between doctors in patients' physical health outcomes, then identifying those doctors who perform well below or well above average could be a relatively simple way to increase the standards of healthcare. This could be done by bringing low performers closer to average and by learning from high performers, which could provide improved healthcare at a relatively low cost. It would certainly be another option for policy makers: to improve the performance in their healthcare system beyond evaluating existing and potential new interventions for suitability.

Once outstanding performers have been identified,[16,18,37,58,59,61,72] it may be possible to have them as role models, mentors, or teachers of other practitioners. Current literature considers standards to still be elusive[88] and identifies outstanding teachers of medicine by acclaim rather than any objective standards.[89] Once identified, excellent role models have been associated by Wright et al[89] as "stressing the importance of the doctor-patient relationship in one's teaching and teaching the psychosocial aspects of medicine" – ie they stress the doctor-patient relationship aspects that go beyond identifying and applying the intervention. Other characteristics may have contributed to exceptional performances, such as their ability to employ easy-to-emulate techniques like putting the patient at ease, willingly listen to the patient to the end, a harmonious lifestyle, a strong sense of purpose, or that they are very rarely exhausted, or have higher expectations of the effectiveness of their intervention,[13] or any of a myriad of other possibilities.

The benefit of research investigating outstanding performers could be large as the differences between exceptional and average performers may be substantial, when simple choices made, or techniques used at work or out of work, that contributed to the outstanding performance then become available to other practitioners. As an exceptional performer is often no more expensive to employ than an average or below average performer, there could be very substantial beneficial effects on public health if many other doctors are given the possibility to improve.

Previous attempts at improving standards of care through profiling have run into difficulties. Krein et al[40] in 2002 argued that despite large profiling campaigns of individual healthcare providers in order to contain costs and improve quality of care, the evidence of effecting change that way has been mixed, expensive, adversely affected careers, tended to ignore the systems the healthcare providers worked in, and, when done badly, profiling can be meaningless, providing incentives that worsen the quality of care.

A word of caution is that in a number of studies the raw patient physical health outcome numbers showed very large differences between doctors but this difference was strongly reduced or even eliminated after taking into account other factors such as patient risk or patients' demographics.[57,60,63] Even after a risk assessment it may be clear that many members of the worst performing group of doctors produce substandard work but the data available lacks statistical power and precludes identifying individuals with certainty. In such a case, disciplining or evicting individual practitioners may not be justifiable without further investigation. However, the more available data there is for each practitioner, the higher the possibility to misuse such data or to disempower practitioners by limiting the opportunity to use their ability and experience or by adding more and more rules and regulations.

Conclusions and Implications

Doctors have an effect on patients' physical health for many interventions and outcomes and after accounting for all known data such as doctor demographics and patient risk. This effect ranges from negligible to substantial and therefore, it is worth investigating further whether these effects and their scale persist for other medical specialties and interventions, which at present is not clear due to the small number of studies found and the lack of consistency in their measurements. Many available RCTs and cohort studies could be reanalyzed to address and estimate the doctors' effects.[81] When grading doctors by patients' physical health outcomes, it is at times possible to identify positive and negative outliers whose confidence interval ranges wholly above or below the average performance. Therefore, it can matter greatly which doctor is chosen.

Support

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Data Sharing Statement

No additional data available.

Ethical Approval

As this is a systematic review of published studies, no ethical approval was required.

Author Contributions

All authors made a significant contribution to the work reported, whether in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal the article has been submitted to; and

agree to be accountable for all aspects of the work. The authors thank Dr Aya Ashraf Ali and Tulia Gonzalez Flores for their excellent editorial contributions. The lead author affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as originally planned (and, if relevant, registered) have been explained.

Disclosure

All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare: no support from any organization for the submitted work; no financial relationships with any organizations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work.

References

1. Hartling L, Milne A, Hamm MP, et al. Testing the Newcastle Ottawa Scale showed low reliability between individual reviewers. *J Clin Epidemiol*. 2013;66(9):982–993. doi:10.1016/j.jclinepi.2013.03.003
2. Higgins TJ, Chandler J, Cumpston M, Li T, Page MJ, Welch VA. Cochrane handbook for systematic reviews of interventions version 6.1 (updated September 2020). Cochrane 2020; 2020. Available from: https://training.cochrane.org/cochrane-handbook-systematic-reviews-interventions. Accessed January 17, 2021.
3. National ambulatory medical care survey: 2018 national summary tables. 2018. Available from: https://www.cdc.gov/nchs/fastats/physician-visits.htm. Accessed July 15, 2022.
4. Health AGDo. Medicare in Australia. Australians make more than 150 million visits to a GP every year. 2022.
5. Curran J. The doctor, his patient and the illness. *BMJ*. 2007;335(7626):941. doi:10.1136/bmj.39384.467928.94
6. Balint M. The doctor, his patient, and the illness. *Lancet*. 1955;265(6866):683–688. doi:10.1016/S0140-6736(55)91061-8
7. Baldwin SA, Imel Z. Therapist effects: findings and methods. *Bergin Garfields Handbook Psychother Behav Change*. 2013;6:258–297.
8. Schnelle C, Clark J, Mascord R, Jones M. Is there a surgeons' effect on patients' physical health, beyond the intervention, that requires further investigation? A systematic review. *Ther Clin Risk Manag*. 2022;1(18):467–490. doi:10.2147/TCRM.S357934
9. Fung V, Schmittdiel JA, Fireman B, et al. Meaningful variation in performance: a systematic literature review. *Med Care*. 2010;48(2):140–148. doi:10.1097/MLR.0b013e3181bd4dc3
10. Mercuri M, Gafni A. Medical practice variations: what the literature tells us (or does not) about what are warranted and unwarranted variations. *J Eval Clin Pract*. 2011;17(4):671–677. doi:10.1111/j.1365-2753.2011.01689.x
11. Wilt TJ, Shamliyan TA, Taylor BC, MacDonald R, Kane RL. Association between hospital and surgeon radical prostatectomy volume and patient outcomes: a systematic review. *J Urol*. 2008;180(3):820–829. doi:10.1016/j.juro.2008.05.010
12. Maruthappu M, Gilbert BJ, El-Harasis MA, et al. The influence of volume and experience on individual surgical performance: a systematic review. *Ann Surg*. 2015;261(4):642–647. doi:10.1097/SLA.0000000000000852
13. Chen P-HA, Cheong JH, Jolly E, Elhence H, Wager TD, Chang LJ. Socially transmitted placebo effects. *Nat Human Behav*. 2019;3(12):1295–1305. doi:10.1038/s41562-019-0749-5
14. Moreau A, Boussageon R, Girier P, Figon S. The "doctor" effect in primary care. *Presse Med*. 2006;35(6):967–973. doi:10.1016/S0755-4982(06)74729-7
15. Alruthia Y, Sales I, Almalag H, et al. The relationship between health-related quality of life and trust in primary care physicians among patients with diabetes. *Clin Epidemiol*. 2020;12:143–151. doi:10.2147/CLEP.S236952
16. Harley M, Mohammed MA, Hussain S, Yates J, Almasri A. Was Rodney Ledward a statistical outlier? Retrospective analysis using routine hospital data to identify gynaecologists' performance. *Br Med J*. 2005;330(7497):929–932. doi:10.1136/bmj.38377.675440.8F
17. Beckett DJ, Spears M, Thomson E. Reliable consultant level data from an Acute Medical Unit: a powerful tool for improvement. *J R Coll Physicians Edinb*. 2018;48(2):108–113. doi:10.4997/jrcpe.2018.202
18. Brown EC, Robicsek A, Billings LK, et al. Evaluating primary care physician performance in diabetes glucose control. *Am J Med Qual*. 2016;31(5):392–399. doi:10.1177/1062860615585138
19. Oomen RJ, Biesaart MC. Doorhalingen in het BIG-register [Removal from the Dutch healthcare professionals register: considerations taken into account by the disciplinary tribunal from 2006 to 2011]. *Ned Tijdschr Geneeskd*. 2012;156(49):548. Dutch.
20. Weenink J. Back on track. Addressing poor performance of healthcare professionals. 2018.
21. AAMC. Careers in medicine - specialty profiles. Available from: https://www.aamc.org/cim/explore-options/specialty-profiles. Accessed March 17, 2022.
22. Hurwitz B. What's a good doctor, and how can you make one? *BMJ*. 2002;325(7366):667–668. doi:10.1136/bmj.325.7366.667
23. Rizo CA, Jadad AR, Enkin M. What's a good doctor and how do you make one?: Doctors should be good companions for people. *BMJ*. 2002;325(7366):711. doi:10.1136/bmj.325.7366.711
24. De Vries EN, Ramrattan MA, Smorenburg SM, Gouma DJ, Boermeester MA. The incidence and nature of in-hospital adverse events: a systematic review. *Qual Saf Health Care*. 2008;17(3):216–223. doi:10.1136/qshc.2007.023622
25. Leppin AL, Gionfriddo MR, Kessler M, et al. Preventing 30-day hospital readmissions: a systematic review and meta-analysis of randomized trials. *JAMA Intern Med*. 2014;174(7):1095–1107. doi:10.1001/jamainternmed.2014.1608
26. Tam VC, Knowles SR, Cornish PL, Fine N, Marchesano R, Etchells EE. Frequency, type and clinical importance of medication history errors at admission to hospital: a systematic review. *CMAJ*. 2005;173(5):510–515. doi:10.1503/cmaj.045311
27. Tjarda Van Heek N, Kuhlmann KFD, Scholten RJ, et al. Hospital volume and mortality after pancreatic resection: a systematic review and an evaluation of intervention in The Netherlands. Conference Paper. *Ann Surg*. 2005;242(6):781–790. doi:10.1097/01.sla.0000188462.00249.36
28. Van Walraven C, Bennett C, Jennings A, Austin PC, Forster AJ. Proportion of hospital readmissions deemed avoidable: a systematic review. *CMAJ*. 2011;183(7):E391–E402. doi:10.1503/cmaj.101860

29. Nilsen SM, Bjørngaard JH, Carlsen F, et al. Hospitals´ discharge tendency and risk of death-an analysis of 60,000 Norwegian Hip fracture patients. *Clin Epidemiol*. 2020;12:173–182. doi:10.2147/CLEP.S237060
30. Ferroni E, Rossi PG, Alegiani SS, et al. Survival of hospitalized COVID-19 patients in Northern Italy: a population-based cohort study by the ITA-COVID-19 network. *Clin Epidemiol*. 2020;12:1337–1346. doi:10.2147/CLEP.S271763
31. Cook JA, Bruckner T, MacLennan GS, Seiler CM. Clustering in surgical trials - database of intracluster correlations. *Trials*. 2012;13:132. doi:10.1186/1745-6215-13-2
32. Campbell M, McKenzie JE, Sowden A, et al. Synthesis without meta-analysis (SWiM) in systematic reviews: reporting guideline. *BMJ*. 2020;368: l6890. doi:10.1136/bmj.l6890
33. Gill L, White L. A critical review of patient satisfaction. *Leadersh in Health Serv*. 2009;22(1):8–19. doi:10.1108/17511870910927994
34. Stang A. Critical evaluation of the Newcastle-Ottawa scale for the assessment of the quality of nonrandomized studies in meta-analyses. *Eur J Epidemiol*. 2010;25(9):603–605. doi:10.1007/s10654-010-9491-z
35. Wells GA, Shea B, O'Connell D, et al. *The Newcastle-Ottawa Scale (NOS) for Assessing the Quality of Nonrandomised Studies in Meta-Analyses*. Oxford; 2000.
36. Allison PD. Fixed effects regression models. In: *Quantitative Applications in the Social Sciences*. Vol. 160. SAGE publications; 2009.
37. Prasad-Kerlin MP, Epstein A, Kahn JM, et al. Physician-level variation in outcomes of mechanically ventilated patients. *Ann Am Thorac Soc*. 2018;15(3):371–379. doi:10.1513/AnnalsATS.201711-867OC
38. Holmboe ES, Weng W, Arnold GK, et al. The comprehensive care project: measuring physician performance in ambulatory practice. *Health Serv Res*. 2010;45(6 Pt 2):1912–1933. doi:10.1111/j.1475-6773.2010.01160.x
39. Selby JV, Schmittdiel JA, Lee J, et al. Meaningful variation in performance: what does variation in quality tell us about improving quality? *Med Care*. 2010;48(2):133–139. doi:10.1097/MLR.0b013e3181c15a6e
40. Krein SL, Hofer TP, Kerr EA, Hayward RA. Whom should we profile? Examining diabetes care practice variation among primary care providers, provider groups, and health care facilities. *Health Serv Res*. 2002;37(5):1159–1180. doi:10.1111/1475-6773.01102
41. O'Connor PJ, Rush WA, Davidson G, et al. Variation in quality of diabetes care at the levels of patient, physician, and clinic. *Prev Chronic Dis*. 2008;5(1):A15.
42. Tuerk PW, Mueller M, Egede LE. Estimating physician effects on glycemic control in the treatment of diabetes. *Diabetes Care*. 2008;31 (5):869–873. doi:10.2337/dc07-1662
43. Rasbash J, Goldstein H. Efficient analysis of mixed hierarchical and cross-classified random structures using a multilevel model. *J Educ Behav Stat*. 1994;19(4):337–350. doi:10.2307/1165397
44. Goldstein H, McDonald RP. A general model for the analysis of multilevel data. *Psychometrika*. 1988;53(4):455–467. doi:10.1007/BF02294400
45. Hofer TP, Hayward RA, Greenfield S, Wagner EH, Kaplan SH, Manning WG. The unreliability of individual physician 'report cards' for assessing the costs and quality of care of a chronic disease. *J Am Med Assoc*. 1999;281(22):2098–2105. doi:10.1001/jama.281.22.2098
46. Orueta JF, Garcia-Alvarez A, Grandes G, Nuno-Solinis R. The origin of variation in primary care process and outcome indicators: patients, professionals, centers, and health districts. *Medicine*. 2015;94(31):e1314. doi:10.1097/md.0000000000001314
47. Papachristofi O, Klein AA, Mackay J, Nashef S, Fletcher N, Sharples LD. Effect of individual patient risk, centre, surgeon and anaesthetist on length of stay in hospital after cardiac surgery: Association of Cardiothoracic Anaesthesia and Critical Care (ACTACC) consecutive cases series study of 10 UK specialist centres. *BMJ Open*. 2017;7(9):e016947. doi:10.1136/bmjopen-2017-016947
48. Gulliford MC, Adams G, Ukoumunne OC, Latinovic R, Chinn S, Campbell MJ. Intraclass correlation coefficient and outcome prevalence are associated in clustered binary data. *J Clin Epidemiol*. 2005;58(3):246–251. doi:10.1016/j.jclinepi.2004.08.012
49. Chakraborty H, Hossain A. R package to estimate intracluster correlation coefficient with confidence interval for binary data. *Comput Methods Programs Biomed*. 2018;155:85–92. doi:10.1016/j.cmpb.2017.10.023
50. Adams G, Gulliford MC, Ukoumunne OC, Eldridge S, Chinn S, Campbell MJ. Patterns of intra-cluster correlation from primary care research to inform study design and analysis. Article. *J Clin Epidemiol*. 2004;57(8):785–794. doi:10.1016/j.jclinepi.2003.12.013
51. Turner RM, Omar RZ, Thompson SG. Constructing intervals for the intracluster correlation coefficient using Bayesian modelling, and application in cluster randomized trials. *Stat Med*. 2006;25(9):1443–1456. doi:10.1002/sim.2304
52. Donner A. A review of inference procedures for the intraclass correlation coefficient in the one-way random effects model. *Int Stat Rev*. 1986;54 (1):67–82. doi:10.2307/1403259
53. Eldridge SM, Ukoumunne OC, Carlin JB. The intra-cluster correlation coefficient in cluster randomized trials: a review of definitions. *Int Stat Rev*. 2009;77(3):378–394. doi:10.1111/j.1751-5823.2009.00092.x
54. Cameron AC, Trivedi PK. *Microeconometrics Using Stata*. TX, USA: Stata press College Station; 2009.
55. Anderson BR, Ciarleglio AJ, Cohen DJ, et al. The Norwood operation: relative effects of surgeon and institutional volumes on outcomes and resource utilization. *Cardiol Young*. 2016;26(4):683–692. doi:10.1017/s1047951115001031
56. Fernández-Castilla B, Declercq L, Jamshidi L, Beretvas N, Onghena P, Van den Noortgate W. Visual representations of meta-analyses of multiple outcomes: extensions to forest plots, funnel plots, and caterpillar plots. *Methodology*. 2020;16(4):299–315. doi:10.5964/meth.4013
57. Cirillo F, Patrizio P, Baccini M, et al. The human factor: does the operator performing the embryo transfer significantly impact the cycle outcome?. *Hum Reprod*. 2020;35(2):275–282. doi:10.1093/humrep/dez290
58. Gossl M, Rihal CS, Lennon RJ, Singh M. Assessment of individual operator performance using a risk-adjustment model for percutaneous coronary interventions. *Mayo Clin Proc*. 2013;88(11):1250–1258. doi:10.1016/j.mayocp.2013.07.017
59. Kaplan SH, Griffith JL, Price LL, Pawlson LG, Greenfield S. Improving the reliability of physician performance assessment: identifying the "physician effect" on quality and creating composite measures. *Med Care*. 2009;47(4):378–387. doi:10.1097/MLR.0b013e31818dce07
60. Kunadian B, Dunning J, Roberts AP, Morley R, de Belder MA. Funnel plots for comparing performance of PCI performing hospitals and cardiologists: demonstration of utility using the New York hospital mortality data. *Catheter Cardiovasc Interv*. 2009;73(5):589–594. doi:10.1002/ccd.21893
61. Navar-Boggan AM, Boggan JC, Stafford JA, Muhlbaier LH, McCarver C, Peterson ED. Hypertension control among patients followed by cardiologists. *Circ Cardiovasc Qual Outcomes*. 2012;5(3):352–357. doi:10.1161/circoutcomes.111.963488
62. Singh S, Goodwin JS, Zhou J, Kuo YF, Nattinger AB. Variation among primary care physicians in 30-day readmissions. *Ann Intern Med*. 2019;170 (11):749–755. doi:10.7326/m18-2526

63. Verma AA, Guo Y, Jung HY, et al. Physician-level variation in clinical outcomes and resource use in inpatient general internal medicine: an observational study. *BMJ Qual Saf*. 2020. doi:10.1136/bmjqs-2019-010425
64. Becerra AZ, Aquina CT, Berho M, et al. Surgeon-, pathologist-, and hospital-level variation in suboptimal lymph node examination after colectomy: compartmentalizing quality improvement strategies. *Surgery*. 2017;161(5):1299–1306. doi:10.1016/j.surg.2016.11.029
65. Eijkenaar F, van Vliet RC. Profiling individual physicians using administrative data from a single insurer: variance components, reliability, and implications for performance improvement efforts. *Med Care*. 2013;51(8):731–739. doi:10.1097/MLR.0b013e3182992bc1
66. Gutacker N, Bloor K, Bojke C, Walshe K. Should interventions to reduce variation in care quality target doctors or hospitals? *Health Policy*. 2018;122(6):660–666. doi:10.1016/j.healthpol.2018.04.004
67. Hannan EL, Zhong Y, Jacobs AK, et al. Incomplete revascularization for percutaneous coronary interventions: variation among operators, and association with operator and hospital characteristics. *Am Heart J*. 2017;186:118–126. doi:10.1016/j.ahj.2017.01.015
68. Glance LG, Hannan EL, Fleisher LA, et al. Feasibility of report cards for measuring anesthesiologist quality for cardiac surgery. *Anesth Analg*. 2016;122(5):1603–1613. doi:10.1213/ane.0000000000001252
69. Goodwin JS, Lin YL, Singh S, Kuo YF. Variation in length of stay and outcomes among hospitalized patients attributable to hospitals and hospitalists. *J Gen Intern Med*. 2013;28(3):370–376. doi:10.1007/s11606-012-2255-6
70. Papachristofi O, Mackay JH, Powell SJ, Nashef SAM, Sharples L. Impact of the anesthesiologist and surgeon on cardiac surgical outcomes. *J Cardiothorac Vasc Anesth*. 2014;28(1):103–109. doi:10.1053/j.jvca.2013.07.004
71. Papachristofi O, Sharples LD, Mackay JH, Nashef SAM, Fletcher SN, Klein AA. The contribution of the anaesthetist to risk-adjusted mortality after cardiac surgery. *Anaesthesia*. 2016;71(2):138–146. doi:10.1111/anae.13291
72. Singh S, Lin YL, Nattinger AB, Kuo YF, Goodwin JS. Variation in readmission rates by emergency departments and emergency department providers caring for patients after discharge. *J Hosp Med*. 2015;10(11):705–710. doi:10.1002/jhm.2407
73. Davenport MS, Khalatbari S, Keshavarzi N, et al. Differences in outcomes associated with individual radiologists for emergency department patients with headache imaged with CT: a retrospective cohort study of 25,596 patients. *AJR Am J Roentgenol*. 2020;1–8. doi:10.2214/ajr.19.22189
74. Jemt T, Olsson M, Renouard F, Stenport V, Friberg B. Early implant failures related to individual surgeons: an analysis covering 11,074 operations performed during 28 years. *Clin Implant Dent Relat Res*. 2016;18(5):861–872. doi:10.1111/cid.12379
75. Steering Committee of the Physicians' Health Study Research Group. Final report on the aspirin component of the ongoing physicians' health study. *N Engl J Med*. 1989;321(3):129–135. doi:10.1056/nejm198907203210301
76. Wampold BE. *The Great Psychotherapy Debate: Models, Methods, and Findings*. Lawrence Erlbaum Associates, Inc; 2001.
77. Kraemer HC, Kupfer DJ. Size of treatment effects and their importance to clinical research and practice. *Biol Psychiatry*. 2006;59(11):990–996. doi:10.1016/j.biopsych.2005.09.014
78. McMullin P. What's a good doctor and how do you make one? *BMJ*. 2002;325(7366):711. doi:10.1136/bmj.325.7366.711
79. Schnelle C, Jones MA. Protocol for a qualitative study on doctors' opinions on and experiences of exceptionally good doctors. *Adv Med Educ Pract*. 2022;13:103–109. doi:10.2147/AMEP.S343554
80. Schnelle C, Jones MA. Protocol for a Qualitative Study on Doctors' Opinions on and Experiences of Exceptionally Good Doctors. *Adv Med Educ Pract*.2022;13:103–109. doi:10.2147/AMEP.S343554
81. Schnelle C, Jones MA. The doctors' effect on patients' physical health outcomes beyond the intervention. A methodological review. *Clin Epidemiol*. 2022;18:467. doi:10.2147/CLEP.S357927
82. Walwyn R, Roberts C. Meta-analysis of absolute mean differences from randomised trials with treatment-related clustering associated with care providers. *Stat Med*. 2015;34(6):966–983. doi:10.1002/sim.6379
83. Lee KJ, Thompson SG. Clustering by health professional in individually randomised trials. *BMJ*. 2005;330(7483):142–144. doi:10.1136/bmj.330.7483.142
84. Wampold BE, Imel ZE. *The Great Psychotherapy Debate: The Evidence for What Makes Psychotherapy Work*. Second ed. Taylor and Francis Inc; 2015:1–323.
85. Ioannidis JPA. Evidence-based medicine has been hijacked: a report to David Sackett. *J Clin Epidemiol*. 2016;73:82–86. doi:10.1016/j.jclinepi.2016.02.012
86. Smaldino PE, McElreath R. The natural selection of bad science. *Royal Soc Open Sci*. 2016;3(9):160384. doi:10.1098/rsos.160384
87. Sertkaya A, Wong -H-H, Jessup A, Beleche T. Key cost drivers of pharmaceutical clinical trials in the United States. *Clin Trials*. 2016;13(2):117–126. doi:10.1177/1740774515625964
88. Kenny NP, Mann KV, MacLeod H. Role modeling in physicians' professional formation: reconsidering an essential but untapped educational strategy. *Acad Med*. 2003;78(12):1203–1210. doi:10.1097/00001888-200312000-00002
89. Wright SM, Kern DE, Kolodner K, Howard DM, Brancati FL. Attributes of excellent attending-physician role models. *N Engl J Med*. 1998;339(27):1986–1993. doi:10.1056/NEJM199812313392706