REVIEW

Evidence on Interventions to Reduce Medical Errors

An Overview and Recommendations for Future Research

John P.A. Ioannidis, MD, Joseph Lau, MD

OBJECTIVE: To critically review the existing evidence on interventions aimed at reducing errors in health care delivery.

DESIGN: Systematic review of randomized trials on behavioral, educational, informational, and management interventions relating to medical errors. Pertinent studies were identified from MEDLINE, EMBASE, the Cochrane Clinical Trials Registry, and communications with experts.

SETTING: Both inpatients and outpatients qualified. No age or disease restrictions were set.

MEASUREMENTS: Outcomes were medical errors, including medication, prescription, and diagnostic errors, and excluding preventive medicine errors and simple ordering of redundant tests.

MAIN RESULTS: Thirteen randomized studies qualified for evaluation. The trials varied extensively in their patient populations (mean age, 2 weeks to 83 years), study setting, definition of errors, and interventions. Most studies could not afford masking and rigorous allocation concealment. In 9 of 13 studies, error rates in the control arms were very high (10% to 63%), and large treatment benefits from the studied interventions were demonstrated for the main outcome. Interventions were almost always effective in a sample of 24 nonrandomized studies evaluated for comparison. Actual patient harm from serious errors was rarely recorded.

CONCLUSIONS: Medical errors were very frequent in the studies we identified, arising sometimes in more than half of the cases where there is an opportunity for error. Relatively simple interventions may achieve large reductions in error rates. Evidence on reduction of medical errors needs to be better categorized, replicated, and tested in study designs maximizing protection from bias. Emphasis should be placed on serious errors.

KEY WORDS: medical errors; meta-analysis; behavioral interventions; randomized trials; study design.

J GEN INTERN MED 2001;16:325-334.

Medical errors are a common cause of morbidity and mortality in a variety of health care settings.¹⁻³ Their importance has been increasingly recognized, as reflected in a recent report by the Institute of Medicine that drew widespread public attention.^{4,5} It is estimated that in the United States alone, medical errors result in 44,000 to 98,000 unnecessary deaths and approximately 1 million excess injuries each year.^{4,6} Extremes of age, complex and/or urgent care, and prolonged hospital stay are associated with more errors, but significant errors may also occur in the outpatient setting and for patients of all ages.^{4,6} Despite the increasing recognition of the importance of this phenomenon, there is limited knowledge of which interventions may be used to effectively reduce the incidence and impact of errors on medical care. There is optimism that both simple behavioral interventions, system approaches, and information technology-based approaches may alter the incidence and consequences of medical errors.^{5,7} However, much of the data pertaining to medical error have been generated using epidemiologic observational studies or intervention studies with before-after comparisons. Randomized evidence has not been carefully scrutinized.

In the present systematic review, we undertook to retrieve and critically evaluate the available randomized evidence on interventions specifically aimed at reducing medical errors. Given the intricacy of defining what constitutes a medical error and the multifaceted dimensions of this phenomenon, we anticipated that this field of research would pose important study design issues. Therefore, in addition to summarizing the available evidence, the aim of this overview was to evaluate study designs and limitations in order to make recommendations for improving future research in this field. Pertinent nonrandomized studies were also evaluated briefly for comparison and to obtain complementary information.

METHODS

Eligibility Criteria

We considered all randomized, controlled trials that examined an intervention versus placebo or no intervention and specified the aim of reducing medical errors as a primary or secondary outcome. It is conceivable that several other interventions tested in randomized trials

From the Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina, Greece (JPAI); Department of Medicine, Tufts University School of Medicine, Boston, Mass (JPAI, JL); and Division of Clinical Care Research, New England Medical Center, Boston, Mass (JL).

Address correspondence to Dr. Lau: Division of Clinical Care Research, New England Medical Center, Box 63, 750 Washington St., Boston, MA 02111 (e-mail: JLau1@Lifespan.org).

for other reasons may incidentally reduce errors, but unless error reduction has been set as a primary or secondary endpoint, this is impossible to know with any certainty. Furthermore, an exact definition of medical errors is difficult. For the purposes of this overview, we included medication errors (including prescription, dosing, and omission errors), prescription of inappropriate/harmful diagnostic tests or omission of necessary orders/prescriptions directly related to patient safety, and misdiagnosis errors beyond the inherent limitations of applied diagnostic tests (limitations existing even when these tests are applied and interpreted appropriately). Behavioral, educational, information, and management interventions qualified, including computerized interventions. Simple ordering of redundant tests was not an eligible outcome. Studies with emphasis on patient compliance were excluded, unless emphasis was entirely on errors (not simply missed doses) made by patients or parents because of inadequate information given by health care providers. Studies evaluating only the omission of orders or actions suggested by preventive medicine guidelines were not included in this overview, because they usually constitute omission of potential benefit rather than direct harm to the patient. Similarly, we did not consider studies of computerized systems for altering physician behavior where the emphasis was not specifically on errors—there is an extensive literature that has already been reviewed previously.⁸ We also excluded studies where 1) the outcome was only interrater variability in the absence of a gold standard; 2) the intervention was a different imaging or laboratory test that could improve diagnosis beyond the diagnostic technology used in the control arm (the emphasis of the overview was on missed and wrong diagnosis with existing technologies, not on diagnostic accuracy of new diagnostic technologies); 3) trainees rather than professional staff or the patients themselves were involved; 4) fictitious or simulated cases were considered; or 5) only time to response of caregivers was assessed without actual errors being counted. We also excluded studies where imaging studies were read with different approaches (e.g., blinded vs unblinded, use of comparative images, etc.) where the focus would be on the optimal reading and interpretation of a diagnostic test. Finally, only randomized evidence was considered for the main evaluation. However, nonrandomized studies were also collected for a complementary assessment of the evidence they may provide.

Literature Search and Data Extraction

The literature search was based on MEDLINE (1966–2000) and EMBASE searches. The main search was conducted in June 1999 and updated subsequently until March 2000. The search was based on the terms “medical errors,” “prescription errors,” “diagnostic er-

rors,” and “medication errors” in conjunction with an array of terms characteristic of randomized controlled trials (e.g., randomized controlled trial, randomized clinical trial, random, placebo). We used no language restrictions. The abstracts were screened, and studies that might qualify were retrieved in full for further screening. We also screened the references of the retrieved papers and communicated with experts and colleagues. Finally, the Cochrane Clinical Trials Registry was also screened.

Data extraction was performed in duplicate and disagreements were discussed in a consensus conference. The following data were extracted: year of publication, sample size (number of patients per arm and number of opportunities for error, if the latter was different from number of patients), study setting and study population characteristics, quality components (including double blinding, allocation concealment,⁹ details on the mode and adequacy of randomization, and adequacy of description of withdrawals), definition of errors and whether errors were a primary or secondary outcome, number or score of errors per arm and comparative statistics. In some studies, the number of opportunities for errors was equal to the number of patients, while in others where several decision and action items might be available per patient, the opportunities for errors count might be substantially larger. For example, in a study of computer generated reminders, it is possible that many reminders may be generated for various orders pertaining to the same patient. Finally, we also collected data on whether the definition of errors used as outcomes was adequately clear and presented in sufficient detail, whether the errors were clinically serious or at least the number of clinically serious errors was mentioned, whether data were reported on the number of patients who were actually clinically harmed and on the number of deaths due to errors, and how the information on error outcomes had been collected. Errors were categorized into diagnosis errors, medication errors (including prescription, omission, and dosing errors), and other management errors.

Nonrandomized Studies

In order to evaluate the pertinent nonrandomized evidence, a separate search was performed without the requirement for studies to be randomized. Otherwise, the same search strategy was used as for randomized studies. The search was based on MEDLINE and did not include perusal of references or communication with experts. The aim was not to ensure that every single observational study in the field would be retrieved, but to complement our assessment of the randomized evidence by also assessing nonrandomized studies in the field. We summarized key attributes of the retrieved studies, including the following: year of publication, type of design, definition of errors, definition of interventions,

sample size (in patients or opportunities for errors), clarity of error definition, whether data were collected also on clinically serious errors, and whether the intervention was found to be statistically significantly effective.

Statistical Analysis

Given the large heterogeneity of study designs and definitions of outcomes, we did not attempt a formal quantitative synthesis by meta-analytic methods. Instead, we expressed the results of each randomized study using the same treatment effect metrics, including risk ratios, risk differences, and numbers needed to treat, whenever applicable.¹⁰ Numbers needed to treat are based on patients or opportunities for error, depending on the study outcome definition. Selected attributes of randomized and nonrandomized studies were compared with Fisher's exact test and odds ratios (ORs) were estimated. Analyses were conducted in SPSS 10.0 (SPSS, Inc., Chicago, Ill) and *P*-values are 2-tailed.

RESULTS

For the identification of randomized trials, 52 papers were retrieved and reviewed at full length. Of those, 39 were excluded as ineligible (nonrandomized design, improper randomization, or no control arm [*n* = 14], interrater variability outcome only and/or no medical error outcomes [*n* = 5], trainees only being involved and fictitious/simulated cases only being considered [*n* = 1], intervention consisting of new technology for drug delivery or mode of reading a diagnostic test [*n* = 6], focus only on redundant testing [*n* = 5], focus only on preventive interventions [*n* = 8]).

Characteristics of the 13 eligible randomized studies¹¹⁻²³ are summarized in Table 1. These studies addressed very diverse aspects of errors in health care delivery, ranging from medication, prescription, and dosage errors to misdiagnosis/mismanagement and lack of recognition of the seriousness of illness. Errors could be made by the physicians, nurses, pharmacists, patients themselves, or (in the case of sick children) their parents. The interventions were equally diverse, including leaflets, automated systems and computerized reminders, instructive protocols, multidisciplinary approaches, algorithms, pharmacist interventions, ergonomic changes (illumination in the workplace), and performance of aspects of health care by health professionals other than physicians. Outcomes were measured either on the basis of patients or opportunities for errors. Three studies involved more than 1,000 patients and 2 more had more than 1,000 opportunities for error. The settings were extremely diverse, offering a sample of the multifarious nature of current medical care and the mean age of the patient populations varied from 2 weeks to 83 years, when reported.

Only 1 randomized study scored positively on all 4 quality items. Masking was implemented in only 2 trials, and only 1 of the 2 provided detailed information on how allocation concealment was ensured. Masking and allocation concealment were usually impractical or even impossible. All studies offered appropriate details on their randomization procedures. The unit of randomization could be the patient (or parent of a sick child), the health care provider, the ward, or even a time unit (in one study, 3 different levels of illumination were randomized to be applied in the workplace for 7 random days each in a sample of 21 days). Ten of the 13 studies offered adequate data on withdrawals and exclusions.

As shown in Table 2, the definition of errors used as outcomes was adequately clear, with the exception of one study which simply stated that a serious error is one which has a potential health risk, without further details. Four studies targeted diagnostic errors, 9 studies targeted medication errors, and 3 also targeted other management errors. In 10 of the 13 studies, the targeted errors were not necessarily clinically serious, and in most cases they were probably not serious at all. Only 2 studies^{17,19} clearly reported clinically significant harm to the patients; no study recorded any information on deaths. It is unclear whether any deaths due to errors occurred at all in any of the 13 studies, although 17 errors were coded as life-threatening in one study.

The data collection methods varied across protocols. Several studies used chart reviews or benefited from the electronic medical records they had set up as part of the intervention. Direct observation of health care delivery by an independent investigator was also commonly used.

The main results of the 13 randomized trials are summarized in Table 3. In 9 of the 13 studies, the interventions were found to be effective in reducing error rates. However, even in the 4 cases where the randomized intervention did not reduce error rates significantly, the conclusion of the study authors was (at least in part) favorable for the tested intervention. In one study,¹⁶ the intervention (a structured protocol for assessing the need for obtaining a radiograph after lower extremity injury) actually increased the error rate; however, it significantly decreased the emergency department waiting time. In another study, investigators demonstrated that nurse practitioners did not make more errors than junior doctors in evaluating minor injuries in the emergency department; this was an equivalence rather than superiority design.²³ In a third study without statistical significance,¹⁸ errors were not the primary endpoint of the study—the intervention was superior to control in terms of interrater concordance of the parent with the pediatrician in recognizing the child's severity of illness. Finally, one study¹⁹ found no additional benefit from adding a team approach to a computerized physician order entry for reducing medication errors, but in a before-after comparison (in the same report), the computerized physician order entry system already seemed to provide a large benefit.

Table 1. Characteristics of Randomized Studies

Study (Year)	Error Definition	Intervention	Sample Size	Setting	Mean Age	Quality Components			
						Mask	Rand	AC	WD
Klassen ¹⁶ (1993)	Missed fracture, dislocation or effusion after trauma	Brand protocol by triage nurse	991 patients	Pediatric emergency room	11 y	+	+(P)	+	+
McCarthy ¹⁸ (1990)	Parent not recognizing illness as being serious	Teaching acute illness observation scales to mothers	369 mothers	Pediatric outpatient	2 w	—	+(P)	—	+
Attard ¹⁷ (1992)	Wrong management (operate or not)	Pain relief with papaveretum for acute abdominal pain	100 patients	Surgical	45 y	+	+(P)	—	—
Sakr ²³ (1999)	Significant clinical management error	Nurse practitioner vs junior doctor providing care	1,453 patients	Adult emergency room	36 y	—	+(P)	—	+
Overhage ¹² (1997)	Error of omission*	Computerized reminders of corollary orders	2,181 patients	Inpatient medical	54 y	—	+(S)	—	+
Owens ¹¹ (1990)	Inappropriate drug choice [†]	Multidisciplinary approach [†]	436 patients	Acute hospital	83 y	—	+(P)	—	+
Ley ¹³ (1976)	Medication error [‡]	Leaflets (easy, moderate, difficult-to-read, none)	160 patients	Psychiatric outpatient	ND	—	+(P)	—	+
Barker ¹⁴ (1984)	Medication error [§]	Automated bedside dispensing	1,775 OE	Surgical unit	ND	—	+(W)	—	+
McMahon ¹⁵ (1997)	Dose error from parents [¶]	Syringe (marked or not) and demonstrated dose	90 patients	Pediatric outpatient	<4 y	—	+(P)	—	+
Bates ¹⁹ (1998)	Serious medication error (nonintercepted)	Team intervention**	4,220 patients	Acute hospital	53 y	—	+(W)	—	—
Pereles ²⁰ (1996)	Medication error	Self-medication program	107 patients	Inpatient geriatric units	80 y	—	+(P)	—	+
Buchanan ²¹ (1991)	Prescription error (content or labeling error)	Illumination in the workplace (3 levels)	10,888 OE	Army outpatient	ND	—	+(T)	—	+
Leape ²² (1999)	Prescription error leading to preventable ADE	Pharmacist participation in rounds	150 patients	Acute hospital	ND	—	+(W)	—	—

* Failure to order tests or treatments needed to monitor/ameliorate the effects of other tests or treatments.

[†] Choices of medications with potential side effects that would affect patient function, despite availability of better alternatives. The Senior Care Unit intervention included interview of the patient, review of chart, presentation of medication concerns and recommendations during team conference, as well as follow-up of recommendations by clinical pharmacologist.

[‡] Defined as the percentage of pills which were not taken compared to total that should have been taken; leaflets contained information on the name of the drug, what to do if a tablet was forgotten, and an explanation that the medication would take some time to have its effects; leaflets differed in readability.

[§] Defined as any deviation from a physician's order on chart (omission, unordered drug, extra dose, wrong dose >17%, wrong route, wrong time >30 minutes).

[¶] Defined as parent giving >0.2 mL more or less than the desired dose upon asked to demonstrate the dose administration.

** Including the pharmacists being present much more often on the unit and available for questions; distributing a recommended dilutions chart; making available a computerized drip-rate calculation program; standardized labeling of intravenous bags, tubes, and pumps; and implementing a pharmacy communication log.

Quality components include masking, randomization details, allocation concealment, and details on withdrawals and excluded data. The randomization unit could be the patient or parent (P), the staff (S), the ward (W), or time (T). For each quality component, a plus sign signifies its presence, a minus sign denotes its absence. Mask indicates masking (double-blind design); Rand, randomization details supporting the adequacy of the randomization process; AC, allocation concealment; WD, details on withdrawals and excluded data; ND, no data; ADE, adverse drug event; OE, opportunities for error. In some studies, OE equals the number of patients (pts), while in others where several decision/action items may be available per patient, the OE may be substantially larger.

In 9 of the 13 randomized studies, the error rates in the control arms of the studies were remarkably high (10% to 63.3%). Given the high control rate and significant treatment effects, sizable risk differences and attractive estimates of numbers needed to treat were calculated in most of the studies (Table 3).

A total of 24 nonrandomized studies were also retrieved and evaluated.^{14,19,24-45} Their characteristics are summarized in Table 4. They included 18 before-after comparisons (comparisons of different time periods), and 6 studies with concurrent controls, of which 2 were pseudorandomized (patients were rotated in the different arms,³⁸ or a random sample selected for evaluation from the experimental arm with allocation that had been performed sequentially).⁴² Thirteen of the 24 studies had been published before 1990 (vs only 2 of the 13 randomized trials). The studies dealt usually with medication errors, but some had also management errors as well. Interventions and sample sizes varied widely and are summarized in Table 4. Five studies did not even specify their sample size (typically error rates were reported). Five studies had unclear definitions of what counted as an error. Only 4 studies gave some data on clinically serious errors, but even these did not have serious errors as the main study outcome. Finally, the intervention was almost ubiquitously found to be effective. There were

only 3 exceptions: in one study,³⁴ the authors described 14 different interventions, but claimed that only 3 were actually implemented—the failure of implementation was felt responsible for the lack of effect; one other exception occurred in a pseudorandomized study⁴² which did not demonstrate any improvement in errors reflecting serious drug interactions by using a computerized medication profile; the third exception occurred in a study evaluating an automated medication cart-filing system.⁴³

In the comparison of randomized versus nonrandomized studies, nonrandomized reports were more likely to have been published before 1990 (OR, 6.5; 95% confidence intervals [95% CI], 1.2 to 36; $P = .035$). As compared to nonrandomized reports, those of randomized design were more likely to report data on clinically serious errors (OR, 3.1; 95% CI, 0.7 to 14.7; $P = .229$), to find that the intervention was not effective in reducing errors (OR, 3.1; 95% CI, 0.7 to 15; $P = .213$), and to specify the sample size (OR, undefined; $P = .140$), but these associations were far from being significant.

DISCUSSION

Medical errors recently have drawn large publicity and the identification of simple measures to reduce their impact

Table 2. Definitions of Errors and Methods of Data Collection

Study (Year)	Error Definition		Clinically Serious Errors	Recorded Clinical Harm to Patient	Deaths Recorded	Data Collection
	Clarity	Type of error				
Klassen ¹⁶ (1993)	Yes	D	2/16	No	No	Active data entry per protocol
McCarthy ¹⁸ (1990)	Yes	D	Yes	No	No	Active data entry per protocol
Attard ¹⁷ (1992)	Yes	D	Yes	Yes	No	Chart review?
Sakr ²³ (1999)	Yes	D, Ma	Yes	No	No	Active data entry per protocol
Overhage ¹² (1997)	Yes	M, Ma	Not necessarily, most probably not	No	No	Electronic medical records
Owens ¹¹ (1990)	Yes	M, Ma	Not necessarily, most probably not	No	No	Chart review
Ley ¹³ (1976)	Yes	M	Not necessarily, most probably not	No	No	Pill count
Barker ¹⁴ (1984)	Yes	M	Not necessarily, most probably not	No	No	Direct observation
McMahon ¹⁵ (1997)	Yes	M	Not necessarily, most probably not	No	No	Direct observation
Bates ¹⁹ (1998)	Yes	M	17/134 were life-threatening; another 65/134 were serious	Yes	No	Chart review + active data entry
Pereles ²⁰ (1996)	No	M	Not necessarily, most probably not	No	No	Home visit 40 d postdischarge
Buchanan ²¹ (1991)	Yes	M	Not necessarily, most probably not	No	No	Direct observation
Leape ²² (1999)	Yes	M	Mentions severity grading in the protocol, but no data are provided	No	No	Chart review

D indicates diagnosis error; M, medication error (including prescription, dosing, dispensing, omission errors); Ma, other management errors.

Table 3. Summarized Results of Randomized Studies

Study (Year)	Error Definition	Bases of Error Count	Errors Made (%)		Comparison Statistics			
			Intervention	Control	RR	RD	NNT	P Value
Klassen ¹⁶ (1993)	Missed serious injury	Patients	16/493 (3)*	0/498 (0)	—	−3	harm	<.001
McCarthy ¹⁸ (1990)	Parent not recognizing illness as being serious	Patients	2/10 (20)	1/10 (10)	2.00	−10	harm	NS
Attard ¹⁷ (1992)	Wrong management (operate or not)	Patients	2/50 (4)	9/50 (18)	0.22	14	7	.051
Sakr ²³ (1999)	Significant clinical error	Patients	65/704 (9)	80/749 (11)	0.86	1	69	.2
Overhage ¹² (1997)	Error of omission (immediate)	OE	ND (22)	ND (46)	0.47	24	4	<.001
	Error of omission (24 h)	OE	ND (29)	ND (50)	0.58	21	5	<.001
	Error of omission (hospital-stay)	OE	ND (37)	ND (56)	0.66	19	5	<.001
Owens ¹¹ (1990)	Inappropriate drug choice	Patients	40/221 (18)	73/215 (34)	0.53	16	6	<.001
Ley ¹³ (1976)	Medication error (antidepressants)	OE	ND (3) [easy leaflet]	ND (16)	0.17	13	8	<.001
			ND (8) [moderate leaflet]		0.50	8	12	
			ND (15) [difficult leaflet]		0.90	2	61	
	Medication error (tranquillizers)	OE	ND (6) [easy leaflet]	ND (15)	0.39	9	11	<.001
			ND (8) [moderate leaflet]		0.53	7	14	
			ND (15) [difficult leaflet]		1.01	0	harm	
Barker ¹⁴ (1984)	Medication error	OE	96/902 (11)	139/873 (16)	0.67	5	19	<.001
McMahon ¹⁵ (1997)	Dose error from parents	Patients	5/30 (17) [syringe]	19/30 (63)	0.26	47	2	<.001
			0/30 (0) [marked syringe]		0.00	63	2	
			80/2173 (4)	54/2047 (3)	1.40	−1	harm	NS
Bates ¹⁹ (1998)	Serious medication error	Patients [†]	14/51 (27)	25/56 (45)	0.61	17	harm	<.001
Pereles ²⁰ (1996)	Medication error	Patients [†]	106/4038 (3)	137/3602 (4)	0.69	1	85	<.001
Buchanan ²¹ (1991)	Prescription error	OE	[high-illumination]		1.01	−0	harm	NS
			126/3249 (4) [mid-illumination]					
			3/75 (4)	8/75 (11)	0.38	7	15	<.001

* Two of the 16 injuries were clinically important; the intervention group was superior in terms of shortening waiting time in the emergency room (average difference 0.3 h).

[†] Data in the table show medication errors per number of patients, but it is conceivable that in some cases more than one error could have occurred in the same patient (not reported). The results were similar when analyzed as errors per 1000 patient-days in Bates et al.¹⁹ (6.0 vs 4.8) and as errors per doses of medications given in Pereles et al.²⁰ (0.045 vs 0.086).

Comparison statistics include the risk ratio (RR), risk difference (RD), number needed to treat (NNT), and P value: the value derived from χ^2 or Fisher's exact test. Opportunities for error (OE) is equal to the number of patients in some studies, while in others where several decision/action items may be available per patient, the OE may be substantially larger. ND indicates no data. NS, not significant ($P > .2$).

has been considered a public priority.⁴ This overview shows the complexity involved in studying interventions aimed to minimize errors in health care delivery. The definition of medical errors is difficult to set within clear boundaries. Medical errors are multifaceted and may involve the diagnosis, management, ordering, carrying out of orders, prescriptions and other medical care by physicians,

nurses, pharmacists, other staff, the patients themselves, or the parents and custodians responsible for their care. The heterogeneity of the studies retrieved in this overview shows these diverse components in the fabric of health care where errors may arise.

In this overview, we used a very strict definition of errors with many exclusion criteria in an effort to bring the

Table 4. Summary of Nonrandomized Controlled Studies

Study (Year-Design)	Error Definition	Intervention	Sample Size	Clear Definition	Data on Serious Errors	Intervention Effective
Barker ²⁸ (1969-BA)	Medication errors	Unit dose distribution system	14,058 OE	Yes	No	Yes
Barker ¹⁴ (1984-BA)	Medication errors	14 interventions	ND	Yes	No	No
Bates ¹⁹ (1998-BA)	Serious medication errors	Computer physician order entry	4,538 patients	Yes	Yes	Yes
Bates ²⁴ (1999-BA)	Medication errors (excluding missed dose)	Computerized physician order entry with decision support	1,817 patients	Yes	Yes	Yes
Borel ²⁹ (1995-BA)	Drug-dispensing system errors	Automated drug dispensing	1,802 OE	Yes	No	Yes
Crooks ⁴⁵ (1967-BA)	Wrong prescription or administration	Single prescription sheet	14,949 doses	No	No	Yes
Evans ³⁷ (1994-BA)	Allergic/ idiosyncratic reactions and severe ADEs	Computerized alerts, standard antibiotic administration, prompt physician notification of ADEs	79,719 patients	Yes	Yes	Yes
Evans ²⁵ (1998-BA)	Wrong orders for antibiotics	Computerized antibiotic management program	1,681 patients	No	No	Yes
Gard ⁴⁴ (1995-BA)	Dosing errors (antibiotics)	Computer-generated protocol	ND	No	No	Yes
Hayman ³² (1989-BA)	Medication dispensing errors	Computer monitoring + disciplinary actions	ND	No	No	Yes
Hill ²⁶ (1967-BA)	Wrong prescription or administration	Redesigned drug sheet	1,738 doses	Yes	No	Yes
Hynniman ⁴⁰ (1970-CC)	Medication errors	Unit dose distribution system	11,481 doses	Yes	No	Yes
Klein ⁴³ (1994-BA)	Medication errors	Automated dispensing system for cart filling	7,842 doses	Yes	No	No
Koepsell ⁴² (1983-CC)	Drug interaction prescription errors	Computerized medication profile	6,186 patients	Yes	No	No
Leape ²² (1999-BA)	Ordering errors resulting in preventable ADEs	Pharmacist participation in rounds	150 patients	Yes	Yes	Yes
McDonald ³⁸ (1977-CC)	Medication errors	Counselling elderly before discharge/ memory aids	165 patients	Yes	No	Yes
Means ⁴¹ (1975-CC)	Medication errors (commission)	Computerized unit dose system	2,671 OE	Yes	No	Yes
Milliken ³⁰ (1990-BA)	Medication discrepancies	Computerized unit dose system	7,802 doses	Yes	No	Yes
Morris ³⁵ (1989-BA)	Dispensing errors	Participation in clinical trials	50 patients	Yes	No	Yes
Raetz ³³ (1988-BA)	Calculations and labeling errors	Computer-assisted calculations	509 OE	Yes	No	Yes
Ray ³⁶ (1995-CC)	Dispensing errors	Automated point-of-use unit drug distribution system	ND	No	No	Yes
Schnell ²⁷ (1976-BA)	Medication errors	Unit dose distribution system	21,856 OE	Yes	No	Yes
Simborg ³⁹ (1975-CC)	Medication errors	Automated medication system	2,662 doses	Yes	No	Yes
Wilson ³¹ (1997-BA)	Medication occurrences	Computerized records shared by pharmacy and nursing	ND	Yes	No	Yes

ADEs indicates adverse drug events; BA, before-after comparison; CC, concurrent controls; ND, no exact data (neither on number of patients nor on opportunities for error); OE, opportunities for error.

topic into sharp focus. We acknowledge that we may have missed some studies which are sometimes difficult to even code for database purposes. Other studies such as those assessing interventions to reduce the omission of preventive measures or even those assessing interventions to reduce redundant test ordering or improve the behavior and reaction time of physicians may be perceived as interventions aiming to reduce errors in a broader sense. There is an even more extensive literature on these topics, some of which has been already reviewed previously.⁸

Given the observed high frequency of errors, there is a dearth of randomized trial data about ways to reduce medical errors. Even with a less comprehensive search, we were able to identify a larger amount of nonrandomized evidence. Compared with the randomized trials, the observational studies were older, although several recent ones were also identified; and they had an equally high, if not even higher, rate of positive findings, raising concerns about publication bias and selection biases in targeting study settings and patient populations. In many cases, it is conceivable that when large improvements are seen in nonrandomized studies, randomized trials may never be performed, such as in the case of unit dose distribution systems or, more recently, several computerized interventions. Randomized trials may not always be applicable to all settings and questions⁴⁶ and observational studies may not always square with the results of randomized trials.⁴⁷ Observational studies have the advantage of studying effectiveness in real practice settings, while randomized trials may be prone to study efficacy under more controlled circumstances with less generalizability. Observational evidence may yield useful insights and should not be discarded.

However, there are several cases where randomization would have been easy to implement without affecting the generalizability or the difficulty of conducting a study. Studies using concurrent controls usually are easy to transform into randomized designs. Finally, before-after comparisons are easier to implement than randomized trials, but one should caution that the use of historical controls may sometimes inflate the magnitude of the treatment effect. Before-after comparisons may be particularly biased, if they extend over a long period of time during which several other changes and interventions that may affect the error rate are implemented (either systematically or sporadically) in the health system where the study is conducted. Such serendipitous changes may result in large reductions of the error rate and they may often be unknown or difficult to model and adjust appropriately. For example, a Hawthorne effect may be observed where physicians make fewer errors over time, especially after a computerized system has been installed, regardless of the merits of the system itself.

Many other important studies in the field were excluded, because they had no control comparison at all. For example, one otherwise well-designed study implemented a computer-alert system to prevent injury from adverse

drug events.⁴⁸ The outcome was the number of true-positive alerts. Measuring such parameters in uncontrolled studies is useful for assessing the performance of a system. However, comparative outcomes (vs no intervention or a standard intervention) on hard endpoints, e.g., deaths or serious morbidity due to iatrogenic errors, would prove indispensable. Such outcomes need to be studied with the gold standard of randomized trials, especially if the proposed interventions are not devoid of cost.

The heterogeneity of the retrieved studies precludes any attempt at their quantitative synthesis. However, it is interesting that the majority of both randomized and nonrandomized studies give positive results for reducing medical errors, and practically almost all of them show that some or all of the interventions under study are worthwhile adopting, based on the results for one or more outcomes. The magnitude of the reported treatment effects is often very large. This probably means that there is large room for improvements in controlling medical errors, often with relatively simple interventions. Regardless of the magnitude of the treatment effects, the magnitude of the error rates in the control groups is also remarkably high, ranging from 10% to 63% in 9 of the 13 studies. These high rates are probably strong evidence for the high incidence of medical errors.

Alternately, it is also possible that biases may be operating in this field of research. Such biases could stem either from publication lag and publication bias,^{49,50} or from the difficulty of assessing outcomes objectively in studies where masking and allocation concealment are difficult to achieve.⁹ Theoretically, this is a field of research where a study with negative results (e.g., when a behavioral or computerized intervention does not affect the error rate) may be difficult to publish. However, given that the large majority of studies, both randomized and nonrandomized, yield positive results, the evidence published to date gives limited insight on why error prevention interventions may sometimes fail rather than succeed. Studies explicitly designed to study this aspect should be encouraged. The available data suggest that even with the large error reductions achieved, there are still large rates of errors even in the interventional groups. Reasons for these failures require better study.

Some other methodologic limitations of the available randomized evidence on medical errors need to be discussed. First, many of the tested interventions may be difficult to generalize and/or their effectiveness may depend on familiarity of and training on their principles. Studies in acute hospitals were sometimes conducted in renowned tertiary care, university-affiliated medical centers in the United States; it is unknown whether the same favorable results would be obtained in other hospitals or in other parts of the world. Furthermore, many of the prior trials have dealt with specialized populations such as children and their parents, the elderly, acute care, or specific outpatient settings. Their findings may not be generalizable to different populations. Furthermore, many

involved cognitive interventions (reminders, educational tools, etc.) and their effectiveness may vary depending on the rigor and intensity of their application.

Some studies used less than optimal randomization methods, in ways that the allocation of the patients to each of the compared arms could have been manipulated. Almost all of the included studies were unmasked, although this was usually fully justified given the nature of the studied observations. The unit of randomization in these studies must be carefully selected. In some cases, using the patient as the unit of randomization may be inappropriate: skills acquired while managing intervention patients may be applied to the care of control patients.

These limitations being acknowledged, clearly all the identified pertinent studies required substantial ingenuity for their design, since studying medical errors requires innovation both in the way interventions are applied and outcomes are assessed. Several of the studies included more than 1,000 patients, and the results are probably valid. However, before the proposed interventions are widely adopted, the cost-effectiveness and effects on quality of life should also be considered alongside with their effects on error counts. Variability across the trials may make generalization difficult. While it is easy to agree that preventing errors is desirable, specific interventions must be evaluated according to their own merits with respect to effectiveness and cost-effectiveness in specific health care settings.

Finally, all errors are not equal, and their impact may vary tremendously. For example, although medication errors are strongly linked to adverse drug events,⁵¹ most medication or prescription errors will not result in any harm. Many studies on medical errors have given limited information of the seriousness of the captured errors. Even when outcomes were defined to include only "serious" errors, the seriousness may have varied substantially from study to study and even within the same study different "serious" errors may have had different consequences for patients. We found a dearth of data pertaining to actual clinical harm to the patient and no study focused on patient deaths. Many studies assessed parameters that almost certainly would not have resulted in death. Although less serious errors that affect many patients and that may expend extensive resources are definitely worth studying, we believe that studies incorporating hard patient outcomes, such as iatrogenic mortality and serious iatrogenic illness, should be encouraged.

In order to perform such studies, large numbers of patients would have to be included, since the event rates are likely to be low. Given the fact that errors occur at various steps in the sequence of medical care and could involve various care providers (e.g., physicians, pharmacists, nurses) or even the patients themselves, observational studies may be used to provide insight on which steps are more error-prone. These steps could then be the target for probing the efficacy of interventions by randomized trials. Still, in order to design such studies, we have to first recognize that medical errors, including serious ones,

should be subjected to study from the perspective of a systematic approach and become the target of large experimental investigation, rather than be silenced for legal purposes.

We thank Ms. Priscilla Chew for her invaluable help in the literature search and Ms. Marian Perez for the retrieval of articles.

REFERENCES

1. Brennan TA, Leape LL, Laird NM, et al. Incidence of adverse events and negligence in hospitalized patients: Results of the Harvard Medical Practice Study I. *N Engl J Med*. 1991;324:370-6.
2. Bates DW, Cullen DJ, Laird N, et al. Incidence of adverse drug events and potential adverse drug events. Implications for prevention. ADE Prevention Study Group. *JAMA*. 1995;274:29-34.
3. Donchin Y, Gopher D, Olin M, et al. A look into the nature and causes of human errors in the intensive care unit. *Crit Care Med*. 1995;23:294-300.
4. Kohn LT, Corrigan JM, Donaldson MS, eds. *To Err Is Human. Building a Safer Health System*. Washington, DC: National Academy Press; 1999.
5. Leape LL, Berwick DM. Safe health care: are we up to it? *BMJ*. 2000;320:725-6.
6. Weingart SN, Wilson RM, Gibbert RW, Harrison B. Epidemiology of medical error. *BMJ*. 2000;320:774-7.
7. Bates DW. Using information technology to reduce rates of medication errors in hospitals. *BMJ*. 2000;320:788-91.
8. Johnston ME, Langton KB, Haynes RB, Mathieu A. Effects of computer-based clinical decision support systems on clinician performance and patient outcome. A critical appraisal of research. *Ann Intern Med*. 1994;120:135-42.
9. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled clinical trials. *JAMA*. 1995;273:408-12.
10. Laupacis A, Sackett DL, Roberts RS. An assessment of clinically useful measures of the consequences of treatment. *N Engl J Med*. 1988;318:1728-33.
11. Owens NJ, Sherburne NJ, Silliman RA, Fretwell MD. The Senior Care Study. The optimal use of medications in acutely ill older patients. *J Am Geriatr Soc*. 1990;38:1082-7.
12. Overhage JM, Tierney WM, Zhou X-H, McDonald CJ. A randomized trial of "corollary orders" to prevent errors of omission. *J Am Med Inform Assoc*. 1997;4:364-74.
13. Ley P, Jain VK, Skilbeck CE. A method for decreasing patients' medication errors. *Psychol Med*. 1976;6:599-601.
14. Barker KN, Pearson RE, Hepler CD, Smith WE, Pappas CA. Effect of an automated bedside dispensing machine on medication errors. *Am J Hosp Pharm*. 1984;41:1352-8.
15. McMahon SR, Rimsza ME, Bay RC. Parents can dose liquid medication accurately. *Pediatrics*. 1997;100:330-3.
16. Klassen TP, Ropp LJ, Sutcliffe T, et al. A randomized, controlled trial of radiograph ordering for extremity trauma in a pediatric emergency department. *Ann Emerg Med*. 1993;22:1524-9.
17. Attard AR, Corlett MJ, Kidner NJ, Leslie AP, Fraser IA. Safety of early pain relief for acute abdominal pain. *BMJ*. 1992;305:554-6.
18. McCarthy PL, Sznajderman SD, Lustman-Findling K, et al. Mothers' clinical judgment: a randomized trial of the Acute Illness Observation Scales. *J Pediatr*. 1990;116:200-6.
19. Bates DW, Leape LL, Cullen DJ, et al. Effect of computerized physician order entry and a team intervention on prevention of serious medication errors. *JAMA*. 1998;280:1311-6.
20. Pereles L, Romonko L, Murzyn T, et al. Evaluation of a self-medication program. *J Am Geriatr Soc*. 1996;44:161-5.

21. Buchanan TL, Barker KN, Gibson T, Jiang BC, Pearson RE. Illumination and errors in dispensing. *Am J Hosp Pharm.* 1991;48:2137-45.
22. Leape LL, Cullen DJ, Clapp MD, et al. Pharmacist participation on physician rounds and adverse drug events in the intensive care unit. *JAMA.* 1999;282:267-70.
23. Sakr M, Angus J, Perrin J, Nixon C, Nicholl J, Wardrope J. Care of minor injuries by emergency nurse practitioners or junior doctors: a randomized controlled trial. *Lancet.* 1999;354: 1321-6.
24. Bates DW, Teich JM, Lee J, et al. The impact of computerized physician order entry on medication error prevention. *J Am Med Inform Assoc.* 1999;6:313-21.
25. Evans RS, Pestotnik SL, Classen DC, et al. A computer-assisted management program for antibiotics and other antiinfective agents. *N Engl J Med.* 1998;338:232-8.
26. Hill PA, Wigmore HM. Measurement and control of drug-administration incidents. *Lancet.* 1967;1:671-4.
27. Schnell BR. A study of unit-dose drug distribution in four Canadian hospitals. *Can J Hosp Pharm.* 1976;29:85-90.
28. Barker KN. The effects of an experimental medication system on medication errors and costs. Part I. Introduction and errors study. *Am J Hosp Pharm.* 1969;26:324-33.
29. Borel JM, Rascati KL. Effect of an automated, nursing unit-based drug-dispensing device on medication errors. *Am J Health Syst Pharm.* 1995;52:1875-9.
30. Milliken DL, Gervais AA. The impact of computerization on medication discrepancies in a centralized unit dose drug utilization system. *Hosp Pharm.* 1990;25:650-2.
31. Wilson AL, Hill JJ, Wilson RG, Nipper K, Kwon IW. Computerized medication administration records decrease medication occurrences. *Pharm Pract Manag Q.* 1997;17:17-29.
32. Hayman JN. Microcomputer-assisted monitoring of medication-dispensing errors. *Am J Hosp Pharm.* 1989;46:1806-9.
33. Raetz KG, Walker PC. Computer-assisted TPN calculations: time savings and improved accuracy associated with use of a micro-computer. *Hosp Pharm.* 1988;23:800-4.
34. Barker KN, Harris JA, Webster DB, et al. Consultant evaluation of a hospital medication system: implementation and evaluation of the new system. *Am J Hosp Pharm.* 1984;41:2022-9.
35. Morris WT, Lamb AM. Clinical trials are usually beneficial to patients and hospitals. *N Z Med J.* 1989;102:472-4.
36. Ray MD, Aldrich LT, Lew PJ. Experience with an automated point-of-use unit-dose drug distribution system. *Hosp Pharm.* 1995;30:18, 20-3, 27-30.
37. Evans RS, Pestotnik SL, Classen DC, Horn SD, Bass SB, Burke JP. Preventing adverse drug events in hospitalized patients. *Ann Pharmacother.* 1994;28:523-7.
38. MacDonald ET, MacDonald JB, Phoenix M. Improving drug compliance after hospital discharge. *Br Med J.* 1977;2:618-21.
39. Simborg DW, Derewicz HJ. A highly automated hospital medication system. Five years' experience and evaluation. *Ann Intern Med.* 1975;83:342-6.
40. Hynnmnan CE, Conrad WF, Urch WA, Rudnick BR, Parker PF. A comparison of medication errors under the University of Kentucky unit dose system and traditional drug distribution systems in four hospitals. *Am J Hospital Pharm.* 1970;27: 802-14.
41. Means BJ, Derewicz HJ, Lamy PP. Medication errors in a multidose and a computer-based unit dose drug distribution system. *Am J Hosp Pharm.* 1975;32:186-91.
42. Koepsell TD, Gurtel AL, Diehr PH, et al. The Seattle evaluation of computerized drug profiles: effects on prescribing practices and resource use. *Am J Public Health.* 1983;73:850-5.
43. Klein EG, Santora JA, Pascale PM, Kiteros JG. Medication cart-filling time, accuracy, and cost with an automated dispensing system. *Am J Hosp Pharm.* 1994;51:1193-6.
44. Gard JW, Starnes HM, Morrow EL, Sanchez PJ, Perlman JM. Reducing antimicrobial dosing errors in a neonatal intensive care unit. *Am J Health Syst Pharm.* 1995;52:1508-13.
45. Crooks J, Weir RD, Coull DC, et al. Evaluation of a method of prescribing drugs in hospital and a new method of recording their administration. *Lancet.* 1967;1:668-71.
46. Stampfer M. Observational epidemiology is the preferred means of evaluating effects of behavioral and lifestyle modification. *Control Clin Trials.* 1997;18:494-9.
47. Kunz R, Oxman AD. The unpredictability paradox: review of empirical comparisons of randomized and non-randomised clinical trials. *BMJ.* 1998;317:1185-90.
48. Raschke RA, Gollihare B, Wunderlich TA, et al. A computer alert system to prevent injury from adverse drug events: development and evaluation in a community teaching hospital. *JAMA.* 1998;280:1317-20.
49. Ioannidis JP. Effect of the statistical significance of results on the time to completion and publication of randomized efficacy trials. *JAMA.* 1998;279:281-6.
50. Simes RJ. Publication bias: the case for an international registry of clinical trials. *J Clin Oncol.* 1986;4:1529-41.
51. Bates DW, Boyle DL, Vander Vliet M, Schneider J, Leape L. Relationship between medication errors and adverse drug events. *J Gen Intern Med.* 1995;10:199-205.