

Provider variation in responses to warnings: do the same providers run stop signs repeatedly?

RECEIVED 2 April 2015

REVISED 17 June 2015

ACCEPTED 30 June 2015

PUBLISHED ONLINE FIRST 24 October 2015



OXFORD
UNIVERSITY PRESS

Patrick E Beeler^{1,2,3}, E John Orav^{1,4}, Diane L Seger⁵, Patricia C Dykes^{1,2}, David W Bates^{1,2,5}

ABSTRACT

Objective Variation in the use of tests and treatments has been demonstrated to be substantial between providers and geographic regions. This study assessed variation between outpatient providers in overriding electronic prescribing warnings.

Methods Responses to warnings were prospectively logged. Random effects models were used to calculate provider-to-provider variation in the rates for the decisions to override warnings in 6 different clinical domains: medication allergies, drug-drug interactions, duplicate drugs, renal recommendations, age-based recommendations, and formulary substitutions.

Results A total of 157 482 responses were logged. Differences between 1717 providers accounted for 11% of the overall variability in override rates, so that while the average override rate was 45.2%, individual provider rates had a wide range with a 95% confidence interval (CI) (13.7%–76.7%). The highest variations between providers were observed in the categories age-based (25.4% of total variability; average override rate 70.2% [95% CI, 29.1%–100%]) and renal recommendations (24.2%; average 70% [95% CI, 29.5%–100%]), and provider responses within these 2 categories were most often clinically inappropriate according to prior work. Among providers who received at least 10 age-based recommendations, 64 of 238 (27%) overrode $\geq 90\%$ of the warnings and 13 of 238 (5%) overrode all of them. Of those who received at least 10 renal recommendations, 36 of 92 (39%) overrode $\geq 90\%$ of the alerts and 9 of 92 (10%) overrode all of them.

Conclusions The decision to override prescribing warnings shows variation between providers, and the magnitude of variation differs among the clinical domains of the warnings; more variation was observed in areas with more inappropriate overrides.

Keywords: clinical practice variation, clinical decision support systems, attitude of health personnel, computer-assisted drug therapy, medical order entry systems

BACKGROUND

Variation in health care across regions and providers has been studied for several decades.¹ Wennberg summarized the main findings in his seminal article of 1973,² stating that (1) they found extensive variation in health care delivery, (2) patients underwent disproportionately more surgical procedures in areas with more surgeons as well as more diagnostic tests in areas with more internists, and (3) health care costs did not correlate with age-adjusted mortality.³

The variation in the incidence of surgical procedures is well known,⁴ especially for some discretionary procedures such as hysterectomies^{5,6} and tonsillectomies.⁷ A recent study presented variations in diagnostic management depending on the intensity of practice between areas.⁸ Further, it has been demonstrated that patients who live in regions with hospitals of higher bed capacity have a substantially higher chance of being hospitalized—unrelated to the disease burden and without benefit for these patients.⁹ In another study, the adherence of heart failure patients to their home medication varied by geographic region.¹⁰

Although there are published data on variations in prescribing behavior,^{11,12} little is known about variations in provider responses to electronic prescribing warnings generated by clinical decision support (CDS) systems.¹³ A recent analysis in our institution showed that electronic warnings were overridden in up to 85% of cases, and chart reviews revealed that the appropriateness of overriding alerts varied dramatically by category of alert type, ranging from 12% to 92%

depending on the clinical domain.¹⁴ However, variations between the providers in terms of how much they differ in their decisions about whether to follow or override electronic prescribing recommendations has not been analyzed. Because of the wide adoption of electronic health records (EHRs), such data are now widely available, and they may have a number of implications on safety improvement and assessment of provider behavior. In addition, overriding important safety warnings may represent unsafe practice, as when drivers repeatedly violate rules for stop signs and red lights, they expose others to greater risk. We undertook this study to examine the degree of variation between providers about the decision to override 6 types of electronic prescribing warnings and also considered whether the likelihood of these warnings were appropriate.

METHODS

Design, site, and period of the study

The study was a retrospective, observational cohort study. All types of outpatient prescribing providers were included as participants in the cohort, ie, physicians, nurse practitioners, and physician assistants. Data were collected from the outpatient clinics and ambulatory practices associated with the Massachusetts General Hospital and the Brigham and Women's Hospital, 2 large tertiary care teaching hospitals; most of these practices were located in the community. Overrides of electronic warnings were prospectively logged between January 1,

Correspondence to Dr David W Bates, Chief, Division of General Internal Medicine & Primary Care, Brigham and Women's Hospital, One Brigham Circle, 1620 Tremont St, 3rd Floor, Boston, MA 02120-1613, USA; dbates@partners.org. For numbered affiliations see end of article.

© The Author 2015. Published by Oxford University Press on behalf of the American Medical Informatics Association.

All rights reserved. For Permissions, please email: journals.permissions@oup.com

2009 and December 31, 2011. The institutional ethics committee approved the study and patient consent was waived.

Electronic notifications

The computerized physician order entry (CPOE) system in combination with CDS analyzed in this study was the “Longitudinal Medical Record” (LMR), an internally developed electronic medical record system used by physicians and other clinical staff in the outpatient setting for documentation of medical care. The displayed electronic recommendations for medication allergies, drug-drug interactions (DDIs), and duplicate drugs were initially derived from First DataBank (First Databank, Inc, South San Francisco, California); these were tailored and updated by both review of the literature, and then based on how providers responded to the alerts, with ongoing review by an expert committee. Renal¹⁵ and age-based¹⁶ recommendations and formulary substitution alerts were internally developed for the inpatient setting and were later adapted for outpatients; they have also been modified serially based on user responses. Further details on LMR and implemented CDS notifications are provided elsewhere.¹⁴ Drug safety warnings of 6 clinical domains displayed to the providers were evaluated: (1) medication allergies, (2) DDIs, (3) duplicate drugs, (4) renal recommendations, (5) age-based recommendations, and (6) formulary substitutions. Of note, in each category, various prescribing problems due to various drugs led to electronic warnings. The overall number of alerts was the sum of the numbers derived from the 6 categories.

Statistical analysis

The primary goal was to quantify the variation between the providers in terms of their binary decisions to accept or override electronic prescribing recommendations. This variation was calculated for each clinical domain and for the aggregated total as well. Further analyses included the variations within providers and the override rates.

Histograms were plotted to illustrate the distribution of the crude override rates, which were defined for each provider as the number of overridden warnings divided by the total number of received warnings. However, only providers receiving at least 10 warnings were considered in these analyses in order to avoid distortions resulting from providers with low alert counts that often led to override rates of 0% or 100%. Similarly, providers were considered in the calculations of the crude override rate for each analyzed clinical domain if they responded to at least 10 alerts according to the respective domain. Therefore, the number of excluded providers varied among the analyzed alert domains.

Random effects models (2-level hierarchical, linear models) were used to calculate the variation between and within the providers for the binary decision to override a warning. Providers were eligible for inclusion if they responded to at least 1 alert according to the analyzed clinical domain or 1 alert independent of the clinical domain in the overall analysis. Unlike usual regression models, the random effects model includes a term for each provider on the right-hand side of the regression equation which represents each provider's own unique override rate. These terms are random variables whose distribution is estimated from the observed data; the distribution provides a 95% confidence interval for the provider override rates. The random effects model produces a “shrinkage” estimate of the override rates which remain stable even if a provider has few warnings. Therefore, we were able to produce 95% confidence intervals at both low and high volumes that captured the true override rates of all providers.

Calculations were performed using the software R, version 3.0.2 (R Foundation for Statistical Computing, Vienna, Austria).

RESULTS

A total of 1717 providers received at least 1 prescription warning in 1 or more of the 6 analyzed clinical domains. Among the 1099 providers who received at least 10 alerts, the crude override rate was 53% (82 100 of 155 413). Nevertheless, most of these providers (662 of 1099, 60%) overrode < 50% of the warnings (figure 1). Only 462 of 1717 providers (27%) received 1 or more alerts within each of the 6 domains.

The distribution of the providers' crude override rates in the clinical domains of the age-based and renal recommendations were skewed to the left (figure 2). Among providers who received at least 10 age-based recommendations, 64 of 238 (27%) overrode $\geq 90\%$ of the warnings and 13 of 238 (5%) overrode all of them. Of those who received at least 10 renal recommendations, 36 of 92 (39%) overrode $\geq 90\%$ of the alerts and 9 of 92 (10%) overrode all of them. Notably, 4 (22%; intersection of the two groups overriding all recommendations) of the mentioned 13+9 providers overrode all received warnings in both clinical domains (figure 3).

Using the random effects model, we were able to include both high and low volume providers, and we found that between-provider differences accounted for 11% of the variability in decisions to override prescribing alerts (table 1), ie, 89% of the variation occurred within a provider. The average override rate measured in the aggregated total of alerts was 45.2%.

Among the 6 clinical domains, the highest magnitudes of between-provider variability were 25% (75% variation within) and 24% (76% variation within) in the categories “age-based recommendations” and “renal recommendations,” respectively. These were also the domains in which provider responses were often clinically inappropriate, according to prior work.¹⁴ The percentages of variation between providers in the remaining 4 clinical domains ranged from 13%

Figure 1: Histogram showing the distribution of the providers' crude override rates of the aggregated total of responses across all alert categories. Only providers who received at least 10 alerts were included (1099 providers).

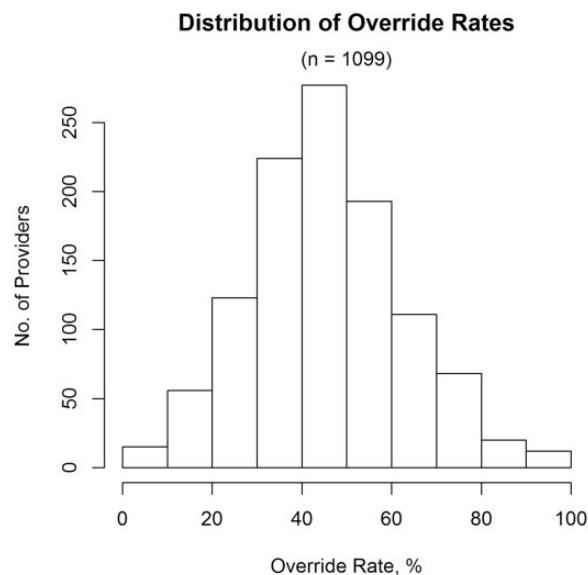


Figure 2: Distributions of the providers' crude override rates per alert category. Only providers who received at least 10 alerts in the respective category were included (number of providers shown in parentheses).

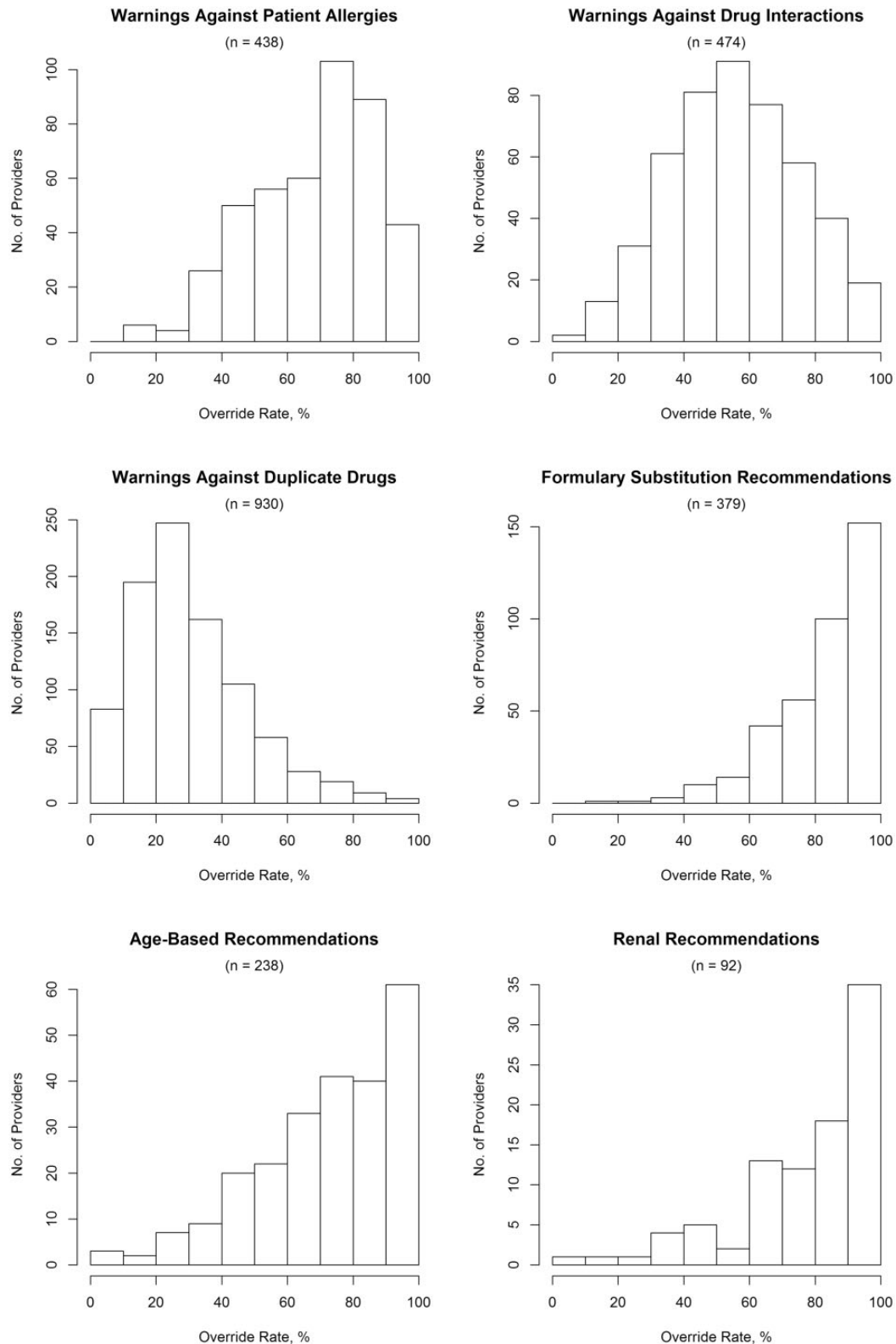


Figure 3: Two-dimensional plots of the crude override rates of the 2 clinical domains with the highest variation and also low appropriateness of overrides.¹⁴ One dot represents 1 provider. Left: Only providers who received at least 10 alerts in each of the 2 clinical domains were included. Right: For comparison, all providers who received at least 1 alert in each of the 2 clinical domains were considered (number of providers shown in parentheses).

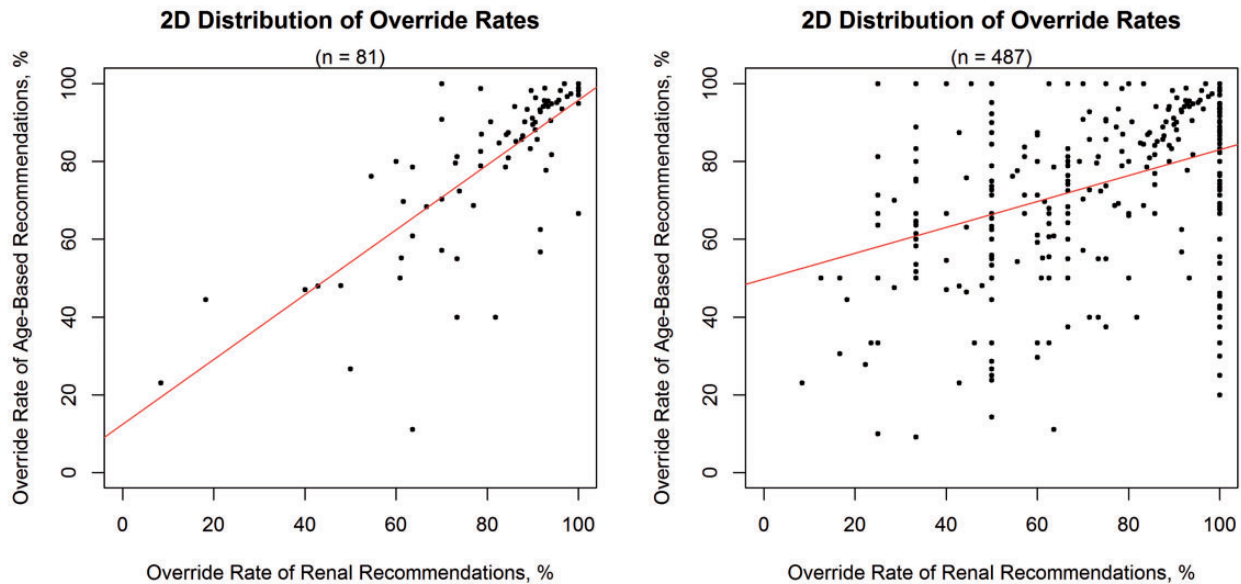


Table 1: Variation between and within providers as well as the true average override rate

	No. of Displayed Alerts	No. of Providers Receiving At Least 1 Alert	Variation Between Providers	Variation Within Providers	Percentage of Variation Between Providers	Percentage of Variation Within Providers	True Override Rate (95% CI)
Total of warnings	157 482	1717	0.02583	0.21002	11.0	89.0	45.2 (13.7%-76.7%)
Warnings against patient allergies	26 408	1160	0.0311	0.1461	17.6	82.4	68.6 (34.0%-100%)
Warnings against drug-drug interactions	24 849	1177	0.03338	0.20258	14.1	85.9	55.4 (19.6%-91.2%)
Warnings against duplicate drugs	75 889	1606	0.02683	0.17498	13.3	86.7	29.2 (0%-61.3%)
Formulary substitution recommendations	15 945	1067	0.02042	0.11245	15.4	84.6	79.5 (51.5%-100%)
Age-based recommendations	10 501	827	0.04402	0.12938	25.4	74.6	70.2 (29.1%-100%)
Renal recommendations	3890	584	0.04279	0.1337	24.2	75.8	70.0 (29.5%-100%)

(duplicates) to 18% (allergies); in addition, the lowest average override rate was measured within the domain of the warnings against duplicate drugs (29%). The highest average override rate was measured for formulary substitution recommendations (79%).

In particular, the age-based and renal recommendations had the highest magnitudes of variation between providers; however, the numbers of displayed alerts in these categories accounted for the 2 smallest proportions among the clinical domains. This was true independent of the number of received alerts per provider. If a provider received on average 1.9 alerts (mean number of the lowest quartile), then renal recommendations represented the smallest proportion

(1.1%) followed by age-based recommendations (4.1%), formulary substitution recommendations (8.2%), warnings against DDIs (12.9%), warnings against patient allergies (12.9%), and warnings against duplicate drugs (60.8%). If a provider received 307.9 alerts (mean number of the highest quartile), then the proportions were 2.6%, 7.2%, 10.2%, 16.2%, 17.7%, and 46.1%, respectively.

DISCUSSION

We found that decisions to override electronic prescribing warnings showed variation between providers, and the magnitudes of variation differed among the clinical domains of the warnings. Regarding the

overall analysis, the provider responses varied moderately with a percentage of variation of 11%. However, the percentage of variation in responses to electronic recommendations for age-based and renal prescribing was 25% and 24%, respectively, and thus substantially higher than the variation in responses to warnings of the other domains, ranging from 13% to 18%. This is important because we had found in prior work that overridden age-based and renal recommendations are likely to be inappropriate.¹⁴ The high proportion of variation measured *within* providers may be explained in part by changes in providers' behavior around whether or not they are familiar with a particular patient. For instance, a provider would (and should) override allergy warnings to a medication they know the patient is already tolerating.¹¹ However, a few providers overrode all or nearly all the warnings they saw, even in important categories. This is almost certainly unsafe.

Researchers have observed medical practice variation for many conditions and procedures.¹⁷ Nevertheless, relatively little is known about variation in the providers' responses to electronic prescribing recommendations.¹³ Articles have been published on variation in ordering behavior, integrating costs as a proxy.^{11,12} For example, Schroeder et al¹¹ found a 4.4-fold variation in drug use between 33 internists, and also that laboratory costs (17-fold variation) were significantly correlated with the drug costs per provider. Zhang et al¹² observed only a 1.6-fold variation in the drug costs among Medicare beneficiaries between geographic regions; however, the use of high-risk medication varied 3.9-fold, and the awareness of drug-disease interactions varied 4.1-, 8.6-, and 7.8-fold for dementia, hip or pelvic fracture, and chronic renal failure, respectively.

We used random effects models to characterize the mean variations between health care providers in their decisions to override or follow electronic recommendations for 6 clinical domains as well as for the aggregated total of these alerts. A recent study calculated the median absolute deviation for a number of defined management decisions to describe the variation between 18 cancer centers.¹⁸ The authors based their choice of the median absolute deviation on the robustness to outliers. However, outliers were not a specific concern in the present study because the variation was calculated between a total of 1717 providers considering 157 482 warning responses.

High variation—especially if in an area in which there is also considerable inappropriateness—can indicate a potential to improve care. One research group implemented interventions into their CPOE system to reduce unnecessary laboratory tests and, notably, measured a significant reduction of variation in the ordering of bundle tests and electrocardiograms.¹⁹ After the dissemination of data on geographic variation of tonsillectomy rates among 13 areas in Vermont, another group found a 46% reduction in overall tonsillectomy rates.⁷

In this study, the variations between providers for accepting or overriding age-based and renal recommendations were much higher than other categories of warnings, ie, different providers varied more in how they responded to these categories, whereas override rates in the other 4 alert categories were more similar from one provider to another. This was the case even though we had also previously found substantial clinical benefits from introduction of age-based and renal recommendations.^{15,16} However, in these initial studies, we did not follow up on the overrides. Since high proportions of inappropriate overrides had been observed among age-based recommendations and renal recommendations,¹⁴ targeted feedback to select providers might increase adherence to guidelines and improve quality of care.²⁰ These data could also be used in credentialing, as they are likely to be much more objective and clinically meaningful than, for example, the clinical assessment of a peer, which is what is mainly used today. However, anything being used for credentialing should be validated

carefully before it is used. Data like this represent one more stream of the “big data” that will be available from electronic records in the future.²¹ While much of the focus on big data has been on patient data, multiple streams of data will also be available about providers.

On the one hand, many warnings about issues such as DDIs may be unnecessary and can contribute to alert fatigue. On the other hand, some DDIs clearly cause substantial harm, eg, Juurlink et al²² have shown at the population level that when some medications are started together the readmission goes up substantially. Along with reducing the burden of clinically insignificant alerts, patient-specific services have been shown to achieve high acceptance rates among clinicians.^{23–25} However, a study by Nanji et al¹⁴ also underscored that not all warnings should be accepted, eg, more than three-fourths of the patient allergy warnings had been overridden, but 92% of these overrides were appropriate. One implication is that it would be much better to use domains with high rates of inappropriately overridden warnings as a main target for interventions or credentialing. In this context, when assessing the appropriateness of overrides of prescribing warnings, it is critical to consider various reasons for overriding, since, for example, the benefit from pain medication might exceed the risk for an adverse drug event. However, not only patient-specific risks can influence the appropriateness of overrides, but also CPOE and alert-related factors such as the human factors characteristics of how the alerts are displayed need to be considered,²⁶ which depend on the system that has been implemented. Finally, a recent study investigated the influence of provider characteristics on responses to electronic prescribing warnings.²⁷

Some limitations need to be taken into account in interpreting this study. We did not assess how often overrides of prescribing warnings actually resulted in adverse drug events, which was beyond the scope of the present investigation. Further, the CPOE system was developed in-house and displayed CDS alerts that have been tailored and updated by expert groups over time. However, some alert categories were derived from a proprietary database widely used in the United States, with the main modification being that we turned off many unnecessary alerts. Most systems currently being deployed in the United States are using commercial EHRs installed recently, and thus the results with these may differ.

We conclude that the largest part of variation was found *within* the individual provider; however, variation *between* the providers was substantial in responses to alerts for clinical domains around which a high proportion of overrides was inappropriate. It may be possible to use override rates to target these providers in order to improve the safety of care—as interventions may be warranted for drivers with defined patterns of risky behavior—and they could also be used for purposes such as credentialing. Override data should now be routinely available. Organizations should track their override rates overall and also analyze overrides by provider and alert type. For types of alerts that are frequently overridden appropriately, however, intervening at the individual level could be counterproductive.

CONTRIBUTORS

PEB designed and performed the research, analyzed and interpreted data, and wrote the manuscript. EJO analyzed and interpreted data, and reviewed the manuscript. DLS and PCD collected data and reviewed the manuscript. DWB designed the research, analyzed and interpreted data, and cowrote the manuscript. All authors approved the final submitted version of the manuscript.

COMPETING INTERESTS

PEB, EJO, DLS, PCD: None.

DWB: Dr Bates is a coinventor on Patent No. 6029138 held by Brigham and Women's Hospital on the use of decision support software for radiology medical management, licensed to the Medicalis Corporation. He holds a minority equity position in the privately held company Medicalis. He receives equity and cash compensation from QPID, Inc, a company focused on intelligence systems for electronic health records.

FUNDING

This study was funded by grant U19 HS021094 from the Agency for Healthcare Research and Quality. Dr Beeler was supported by the Swiss National Science Foundation. The funding sources played no role in the design and conduct of this study; the collection, management, analysis of the data, or the interpretation of the results; the review and approval of the manuscript.

REFERENCES

- Paul-Shaheen P, Clark JD, Williams D. Small area analysis: a review and analysis of the North American literature. *J Health Polit Policy Law* 1987;12(4):741–809.
- Wennberg JE, Gittelsohn AM. Small area variations in health care delivery. *Science* 1973;182(4117):1102–08.
- Wennberg JE. Forty years of unwarranted variation—and still counting. *Health Policy Amst Neth* 2014;114(1):1–2. doi:10.1016/j.healthpol.2013.11.010.
- Birkmeyer JD, Reames BN, McCulloch P, et al. Understanding of regional variation in the use of surgery. *Lancet* 2013;382(9898):1121–29. doi:10.1016/S0140-6736(13)61215-5.
- Wennberg JE, Barnes BA, Zubkoff M. Professional uncertainty and the problem of supplier-induced demand. *Soc Sci Med* 1982;16(7):811–24.
- Roos NP. Hysterectomy: variations in rates across small areas and across physicians' practices. *Am J Public Health* 1984;74(4):327–35.
- Wennberg JE, Blowers L, Parker R, et al. Changes in tonsillectomy rates associated with feedback and review. *Pediatrics* 1977;59(6):821–26.
- Song Y, Skinner J, Bynum J, et al. Regional variations in diagnostic practices. *N Engl J Med* 2010;363(1):45–53. doi:10.1056/NEJMsa0910881.
- Fisher ES, Wennberg JE, Stukel TA, et al. Associations among hospital capacity, utilization, and mortality of US Medicare beneficiaries, controlling for sociodemographic factors. *Health Serv Res* 2000;34(6):1351–62.
- Zhang Y, Wu S-H, Fendrick AM, et al. Variation in medication adherence in heart failure. *JAMA Intern Med* 2013;173(6):468–70. doi:10.1001/jamainternmed.2013.2509.
- Schroeder SA, Kenders K, Cooper JK, et al. Use of laboratory tests and pharmaceuticals. Variation among physicians and effect of cost audit on subsequent use. *JAMA* 1973;225(8):969–73.
- Zhang Y, Baicker K, Newhouse JP. Geographic variation in the quality of prescribing. *N Engl J Med* 2010;363(21):1985–88. doi:10.1056/NEJMp1010220.
- Van der Sijs H, Aarts J, Vulto A, et al. Overriding of drug safety alerts in computerized physician order entry. *J Am Med Inform Assoc* 2006;13(2):138–47. doi:10.1197/jamia.M1809.
- Nanji KC, Slight SP, Seger DL, et al. Overrides of medication-related clinical decision support alerts in outpatients. *J Am Med Inform Assoc* 2014;21(3):487–91. doi:10.1136/amiajnl-2013-001813.
- Chertow GM, Lee J, Kuperman GJ, et al. Guided medication dosing for inpatients with renal insufficiency. *JAMA* 2001;286(22):2839–44.
- Peterson JF, Kuperman GJ, Shek C, et al. Guided prescription of psychotropic medications for geriatric inpatients. *Arch Intern Med* 2005;165(7):802–07. doi:10.1001/archinte.165.7.802.
- Corallo AN, Croxford R, Goodman DC, et al. A systematic review of medical practice variation in OECD countries. *Health Policy Amst Neth* 2014;114(1):5–14. doi:10.1016/j.healthpol.2013.08.002.
- Weeks JC, Uno H, Taback N, et al. Interinstitutional variation in management decisions for treatment of 4 common types of cancer: a multi-institutional cohort study. *Ann Intern Med* 2014;161(1):20–30. doi:10.7326/M13-2231.
- Neilson EG, Johnson KB, Rosenbloom ST, et al. The impact of peer management on test-ordering behavior. *Ann Intern Med* 2004;141(3):196–204.
- Bates DW, Kuperman GJ, Wang S, et al. Ten commandments for effective clinical decision support: making the practice of evidence-based medicine a reality. *J Am Med Inform Assoc* 2003;10(6):523–30. doi:10.1197/jamia.M1370.
- Bates DW, Saria S, Ohno-Machado L, et al. Big data in health care: using analytics to identify and manage high-risk and high-cost patients. *Health Aff Proj Hope* 2014;33(7):1123–31. doi:10.1377/hlthaff.2014.0041.
- Juurink DN, Mamdani M, Kopp A, et al. Drug-drug interactions among elderly patients hospitalized for drug toxicity. *JAMA* 2003;289(13):1652–58. doi:10.1001/jama.289.13.1652.
- Fraser GL, Stogsdill P, Dickens JD, et al. Antibiotic optimization. An evaluation of patient safety and economic outcomes. *Arch Intern Med* 1997;157(15):1689–94.
- Lesprit P, Duong T, Girou E, et al. Impact of a computer-generated alert system prompting review of antibiotic use in hospitals. *J Antimicrob Chemother* 2009;63(5):1058–63. doi:10.1093/jac/dkp062.
- Beeler PE, Eschmann E, Rosen C, et al. Use of an on-demand drug-drug interaction checker by prescribers and consultants: a retrospective analysis in a Swiss teaching hospital. *Drug Saf Int J Med Toxicol Drug Exp* 2013;36(6):427–34. doi:10.1007/s40264-013-0022-1.
- Seidling HM, Phansalkar S, Seger DL, et al. Factors influencing alert acceptance: a novel approach for predicting the success of clinical decision support. *J Am Med Inform Assoc* 2011;18(4):479–84. doi:10.1136/amiajnl-2010-000039.
- Cho I, Slight SP, Nanji KC, et al. The effect of provider characteristics on the responses to medication-related decision support alerts. *J Am Med Inform Assoc* 2014;21(3):487–91. doi:10.1016/j.jimmedinf.2015.04.006.

AUTHOR AFFILIATIONS

¹Division of General Internal Medicine & Primary Care, Brigham and Women's Hospital, Boston, Massachusetts, USA

²Harvard Medical School, Boston, Massachusetts, USA

³Research Center for Medical Informatics, University Hospital Zurich, Switzerland

⁴Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts, USA

⁵Partners Healthcare Systems, Inc, Wellesley, Massachusetts, USA