

Inter-Rater Agreement Among Physicians on the Clinical Significance of Drug-Drug Interactions

Howard R. Strasberg, MD, MS¹, Albert Chan, MD, MS², Stephen J. Sklar, PharmD¹

¹Wolters Kluwer Health, Indianapolis, IN
²Palo Alto Medical Foundation, Palo Alto, CA

ABSTRACT

Background: Alert fatigue could potentially be improved if physicians agreed on which alerts were clinically significant. We conducted a study to determine the extent to which physicians agree on which drug-drug interactions are clinically significant. **Methods:** Two groups of eight generalist physicians reviewed 100 randomly selected drug-drug interactions from the Medi-Span® Drug Therapy Monitoring System™ database and indicated whether they thought each interaction was clinically significant based on the full-text clinical discussion contained within each interaction monograph and their clinical experience. **Results:** The Fleiss Kappa measure of inter-rater agreement was 0.19 (0.12, 0.26) for one group, 0.22 (0.14, 0.29) for the second group and 0.21 (0.15, 0.27) for the combined group. **Conclusion:** We found poor agreement among generalist physicians on which drug-drug interactions are clinically significant. Use of a feature to allow physicians to tailor alerts to their needs may be an important component in reducing alert fatigue.

INTRODUCTION

One of the major challenges in the use of clinical decision support (CDS) software is alert fatigue[1], which is defined as the tendency for clinicians to miss important (“signal”) alerts among a barrage of less important (“noise”) alerts. These less important alerts are generally valid and supported by the medical literature, but they are not necessarily clinically significant enough to warrant a change in the management of a particular patient. For example, a systematic review found that drug safety alerts are ignored in 49% to 96% of cases[2]. Another study found that 9.6% of medication orders resulted in a drug-drug interaction alert[3].

One potential approach to dealing with alert fatigue is to attempt to determine which alerts are the most clinically significant and to configure the CDS software to display only those alerts deemed to be clinically significant. Under this approach, fewer alerts would be displayed, but those that are displayed would potentially be more useful. Therefore, healthcare providers might pay more attention to the alerts and be less likely to ignore them. An example of this approach is the list published by Phansalkar et al[4] containing 15 critical drug-drug interactions that the authors of that study suggest should trigger alerts in all electronic health records (EHRs).

This approach however would only work if physicians agreed on which alerts were clinically significant. Given that physicians have different backgrounds, training, experiences and patient populations, we conducted a study to determine the extent to which a group of generalist physicians agree on which drug-drug interactions are clinically significant.

METHODS

We identified the top 500 drugs prescribed in the United States in 2010 using data from Source Healthcare Analytics, LLC. Using the Medi-Span® Drug Therapy Monitoring System™ (Wolters Kluwer Health, Indianapolis, IN), we screened all 500 drugs against each other to look for potential drug-drug interactions. Because interactions frequently relate to a pair of drug classes rather than to a pair of individual drugs, we removed duplicates from the list such that we were left with 262 unique interaction pairs at the drug class level. We randomly selected 100 interactions from this list for inclusion in the study. Information on each monograph that was presented to each reviewer included the referenced clinical discussion from the Medi-Span database. No information on severity codes assigned to each interaction by Wolters Kluwer Health clinicians was made available to the reviewers.

Two groups of eight physicians were recruited to participate in the study. All were generalist physicians trained in either family medicine or internal medicine. Eight physicians were from the Palo Alto Medical Foundation and eight physicians were from Wolters Kluwer Health. The physicians from Wolters Kluwer Health were not involved in the development of the drug interaction monographs that were reviewed.

For each of the 100 interactions, each physician was asked the following question:

Do you consider this interaction to be clinically significant?

Two choices were given for each answer:

Yes – clinically significant. I would want to see this alert.

No – not clinically significant. I would NOT want to see this alert.

In addition, a comments box was provided.

We utilized the Fleiss Kappa[5] to measure inter-rater agreement for each group of eight physicians independently and for the combined group of sixteen physicians. The Fleiss Kappa compares the observed agreement above chance to the maximum possible agreement above chance. Using the sample size equation proposed by Gwet[6], and assuming arbitrarily that raters would agree 50% of the time, we estimated that we would need a sample size of 100 interactions to achieve a relative error of the inter-rater agreement estimate of +/- 20%. To compute 95% confidence intervals for the Fleiss Kappa estimates, we used Gwet's variance equation[7].

RESULTS

The inter-rater agreement results are presented in Table 1.

Table 1 – Inter-rater agreement results

Physician Group	Fleiss Kappa (95% confidence interval)
Wolters Kluwer Health (n=8)	0.19 (0.12, 0.26)
Palo Alto Medical Foundation (n=8)	0.22 (0.14, 0.29)
Combined (n=16)	0.21 (0.15, 0.27)

For a randomly selected interaction reviewed by the combined group of sixteen physicians, there was a 62.12% chance that two physicians would agree on whether or not the interaction was clinically significant. However, there was also a 51.95% chance that two physicians would agree just by chance. Therefore, the actual agreement above chance (10.17%) expressed as a fraction of the maximum possible agreement above chance (48.05%) yielded a Fleiss Kappa of 0.21.

Figure 1 provides the number of interactions each number of physicians rated as clinically significant. The two interactions that all sixteen physicians agreed were clinically significant were: (i) Clopidogrel with Fluconazole; and (ii) Oxybutynin with Potassium Chloride. The single interaction that all sixteen physicians judged as not clinically significant was Naproxen with Atenolol.

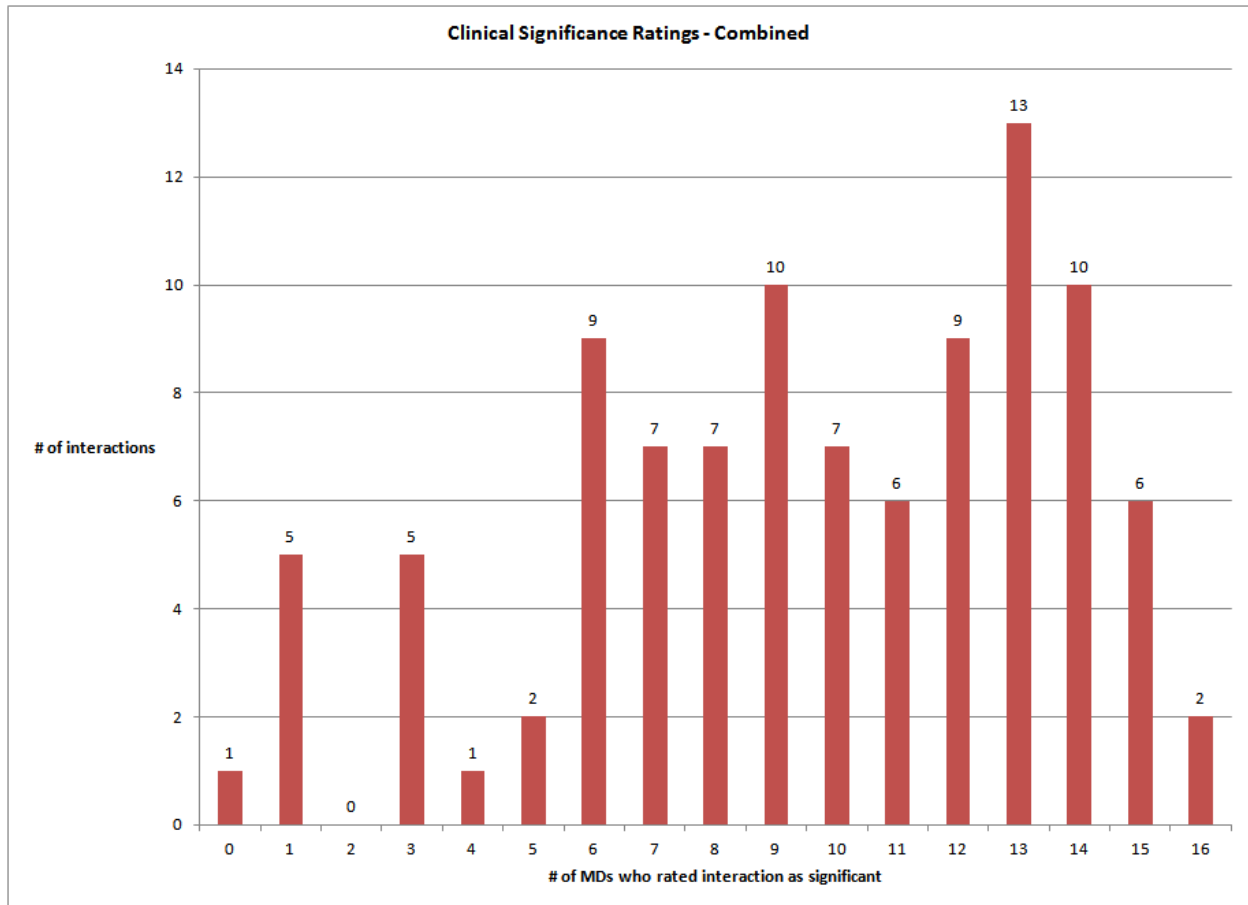


Figure 1 – Number of interactions each number of physicians rated as clinically significant. For example, there were 10 interactions that 9 (out of 16) physicians rated as clinically significant.

DISCUSSION

We found poor, but statistically significant, agreement among generalist physicians regarding which of a defined group of drug-drug interactions are clinically significant. This low inter-rater agreement was remarkably consistent across two different samples of generalist physicians. In 97 out of 100 interactions there was at least some disagreement on whether the interaction was clinically significant.

Phansalkar et al[4] have published a list of 15 critical drug-drug interactions that they suggest should trigger alerts in all EHRs. Neither of the two interactions in our study that 16 of 16 physicians indicated was clinically significant appeared in Phansalkar's list. Also, if Phansalkar's list had been used as the basis for determining which interactions to display to physicians, and if an arbitrary threshold of 12 out of 16 physicians rating an interaction as clinically significant in our study had been used to determine clinical significance, use of Phansalkar's list would have achieved a precision of 100% and a recall of 2.5%. Use of Phansalkar's list as an institution's sole group of interactions to be screened would therefore be expected to miss numerous clinically significant interactions.

We attempted to minimize the variability among physicians by limiting the study to generalists. Even so, we still found poor agreement. Based on some of the qualitative comments made in the comments box of the study tool, there may have been two schools of thought on some of these interactions. Under one school of thought, certain interactions should be part of the working knowledge base of any practicing physician, and therefore, it would be a nuisance to display these alerts to physicians. Under the other school of thought, physicians still benefit from reminders even for items in their own knowledge base. For example, reminders for influenza vaccine have been shown to increase compliance rates with influenza guidelines[8], notwithstanding the fact that physicians are generally aware of these guidelines. In addition, physicians, who are generally pressed for time, may not have the

time to review each patient's medication list thoroughly before ordering a new medication. The CDS software, on the other hand, can conduct such a review very efficiently.

Nevertheless, with the low level of agreement that was demonstrated, there are clearly differences among generalist physicians on which interactions they would like to see as alerts. Therefore, offering users the option to tailor alerts to their specific needs may be an important component in reducing alert fatigue. For example, physician users may be presented with options such as "Do not display this alert again" alongside each alert. This stored preference could have a time limit, allowing users to reaffirm their decisions at periodic intervals. This alert customization could potentially be overridden by the data provider in cases where important new clinical information became available about the drug-drug interaction in question.

One limitation of this study is that we did not present each interaction in the context of a particular patient, and assessment of clinical significance might depend on the context. For example, some interactions may only be significant at certain doses, only for elderly patients or children, only for patients with poor renal function, only for patients with diabetes mellitus, or only for patients with certain genetic polymorphisms. It is possible that if we had presented these interactions along with detailed information about a fictional patient, we would have found more agreement among physicians.

Another limitation of this study is that by design, our sample only included generalist physicians. Future work could look at whether inter-rater agreement is higher among specialist physicians of the same specialty. Specialists may have specific knowledge related to: (a) the subtleties of when it is clinically appropriate to use two particular drugs together; (b) how to monitor for possible problems; and (c) how to manage any problems that may arise.

CONCLUSION

We found poor agreement among generalist physicians regarding which of a defined group of drug-drug interactions are clinically significant. Use of a feature to allow physicians to tailor alerts to their needs will be an important component in reducing alert fatigue.

REFERENCES

- 1 Cash JJ. Alert fatigue. *Am J HealthSyst Pharm*. 2009; 66(23): 2098-2101.
- 2 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;13:5-11.
- 3 Zwart-van Rijkom JE, Uijtendaal EV, ten Berg MJ, van Solinge WW, Egberts AC. Frequency and nature of drug-drug interactions in a Dutch university hospital. *Br J Clin Pharmacol*. 2009;68(2):187-93.
- 4 Phansalkar S, Desai A, Bell D, et al. High-priority drug-drug interactions for use in electronic health records. *J Am Med Inform Assoc*. 2012;19(5):735-43.
- 5 Fleiss JL. Measuring nominal scale agreement among many raters. *Psychol Bull*. 1976; 76(5): 378-382.
- 6 Gwet KL. Inter-rater reliability discussion corner. Available at http://agreestat.com/blog_irr/sample_size_determination.html. Accessed on 11-26-2012.
- 7 Gwet KL. Computing inter-rater reliability and its variance in the presence of high agreement. *Br J Math Stat Psychol*. 2008 May;61(Pt 1):29-48.
- 8 Tang PC, LaRosa MP, Newcomb C, Gorden SM. Measuring the effects of reminders for outpatient influenza immunizations at the point of clinical opportunity. *J Am Med Inform Assoc*. 1999;6:115-21.