Consistency in the safety labeling of bioequivalent medications

Jon Duke¹,²*, Jeff Friedlin¹ and Xiaochun Li³

¹Regenstrief Institute, Indianapolis, IN, USA
²Department of Medicine, Indiana University School of Medicine, Indianapolis, IN, USA
³Division of Biostatistics, Indiana University School of Medicine, Indianapolis, IN, USA

ABSTRACT

Purpose Bioequivalent medications are required by the Food and Drug Administration to have identical warnings on their labels. This requirement has both clinical and legal importance, yet has never been validated. We sought to determine the real-world consistency of electronic labeling for bioequivalent drugs from different manufacturers.

Methods Using natural language processing, we indexed the adverse drug reactions (ADRs) found in the Adverse Reactions and Post-Marketing sections of 9105 structured product labels. We calculated the standard deviation in ADR labeling for each bioequivalent drug and the percent deviation of each generic label from its corresponding brand. We also analyzed the performance of individual generic manufacturers. For the 25 drugs with the greatest discrepancy in labeled ADRs, we performed manual review to identify causes of inconsistency.

Results 68% of multi-manufacturer drugs had discrepancies in ADR labeling. For a given drug, the mean deviation in number of labeled ADRs was 4.4, and the median was 0.8 (IQR 0 to 3.2). The mean range in number of labeled ADRs was 12 +/− 0.9, and the median was 2 (IQR 0 to 9). Overall, 77.9% of generic manufacturers produced labels differing from brand. Causes of inconsistency included missing tables, outdated post-marketing reports, and formatting issues.

Conclusions Despite FDA mandate, bioequivalent drugs often differ in their safety labeling. Physicians should be aware of such differences and regulators should consider new strategies for harmonizing bioequivalent labels. Copyright © 2012 John Wiley & Sons, Ltd.

key words—drug labeling; generic medications; adverse reactions; drug safety; pharmacoepidemiology

Received 23 January 2012; Revised 22 August 2012; Accepted 31 August 2012

BACKGROUND

Generic drug manufacturers are required by the Food and Drug Administration (FDA) to demonstrate that their medications are ‘bioequivalent’ to reference-branded drugs—meaning that they have identical active ingredients, dose forms, and approximate pharmacokinetics.¹ Similarly, when creating the product labeling (i.e. package insert) for a generic medication, manufacturers are required by the FDA to include safety data identical to that of the brand label.² This requirement is designed to ensure that patients receiving a generic medication will get consistent warnings regardless of which manufacturer’s version of the drug is dispensed. Over the past several years, however, this requirement has become controversial and was at the center of several legal battles including a recent case before the United States Supreme Court.³⁻⁵

The case, PLIVA Inc v Mensing, involved a diabetic patient who developed tardive dyskinesia in 2001 after prolonged use of generic metoclopramide.⁵ The patient sued the manufacturer, PLIVA, stating that its product labeling did not adequately warn of the risk of prolonged use despite emerging evidence in the medical literature. PLIVA’s defense was that it could not be held liable for its label’s safety warnings because as a generic manufacturer, it was simply required to copy the labeling of Reglan, the branded version of metoclopramide. In June 2011, the Supreme Court issued a 5–4 ruling agreeing with the drug makers, noting that “the warning labels of a brand-name drug and its generic copy must always be the same.”⁵ In the court’s view, this “ongoing federal duty of ‘sameness’” superseded state liability laws. The major consequence of this decision is that patients who are harmed by taking a generic drug, even if inadequately labeled, are now unable to sue the drug’s manufacturer for failing to warn them of potential risks.
The PLIVA ruling may have significant long-term implications, yet the very crux of the case, the notion that brand and generic labels for the same drug are in fact the same, has never been tested. Such validation is essential—not just for legal reasons but for clinical ones as well. Physicians seeking to review drug safety information frequently refer to manufacturer labels, either online or in sources such as the Physicians Desk Reference. Yet these sources often present multiple manufacturers’ labels for each drug. For example, Figure 1 shows search results for atenolol labels using the FDA website DailyMed. As shown, a physician must select a particular manufacturer’s label and assume that the same information will be found regardless of which label is chosen. Should the warnings differ from one manufacturer’s label to the next, important safety information could be missed.

In the current paper, we present a comprehensive analysis of the consistency of electronic safety labeling across bioequivalent generic medications. The goal of this research is to determine whether FDA guidelines have been effective, and whether the underlying assumption of labeling consistency—on the part of doctors, patients, and the Supreme Court—is indeed correct.

METHODS

Data sources

We retrieved 9105 product labels available on DailyMed, an online repository of medication information maintained by the FDA and the National Library of Medicine, as of September 15 2010 (note, this research began after acceptance of the PLIVA case by the Supreme Court but prior to its final ruling.) DailyMed provides public access to drug data in Structured Product Label (SPL) format, an electronic standard designed to facilitate the submission, processing, and utilization of labeling information. SPLs are currently the required form of labeling submission and are available for over 85% of prescription medications. Our search excluded non-human drugs as well as alternative (e.g. herbal) therapies. We used the RxNorm drug terminology to aggregate bioequivalent medications. RxNorm, a terminology maintained by the National Library of Medicine, provides mappings between labels and their associated active ingredient(s) and dosage form (e.g. metronidazole oral tablet). RxNorm also indicates which labels correspond to a brand version of a drug and which to a generic.

Identifying labeled adverse reactions

To extract drug safety data from these labels, we utilized a software application we created known as the Structured Product Label Information Coder and Extractor (SPLICER). SPLICER utilizes natural language processing to identify reactions found within the ‘Adverse Reactions’ and ‘Post-Marketing’ sections of a label. These reactions are then mapped to a standardized drug safety terminology, the Medical Dictionary of Regulatory Activities. The details of the SPLICER application have been described in greater detail elsewhere. A prior evaluation of its performance demonstrated a high degree of accuracy in adverse event extraction with a sensitivity of 92.8% and positive predictive value of 95.1%.

Data analysis

For each clinical drug, we looked at the total number of associated labels, the standard deviation in number of adverse drug reactions (ADRs) found on these labels, and the difference between the maximum and minimum number of labeled ADRs for the drug (range). While the standard deviation provides a good measure of the variability of labeling within an individual drug, the range conveys the worst case scenario—that is, how much might the information presented to a physician differ depending on which manufacturer’s label is selected for review. For each generic drug label, we also calculated its percent deviation from the reference brand. For example, if a generic label listed 90 adverse reactions while its brand counterpart listed 100, the generic’s deviation from brand would be 10%.
drugs with more than one branded product (e.g. lisinopril marketed as both Zestril and Prinivil), we used the Drugs@FDA website to match each generic version to a corresponding brand where available. In order to better understand the source of labeling discrepancies, we manually reviewed the 25 drugs with the greatest deviation and recorded the differences between their generic and brand labels.

In addition to analyzing labeling patterns of individual drugs, we also measured the performance of generic manufacturers in maintaining labels identical to brand. To do so, we aggregated labels by manufacturer using information derived from the SPL. We then calculated the average percent deviation for each manufacturer’s products. We hypothesized that manufacturers producing more drugs would have more streamlined processes and thus greater labeling consistency. To evaluate this, we assessed the relationship between the number of labels produced by a manufacturer and the proportion of those labels that deviated from their brand counterparts.

RESULTS

We extracted 861,271 adverse reactions from 9105 drug labels representing 1540 distinct clinical drugs. The median number of manufacturer labels per drug was 3 (IQR 1 to 5; range 1 to 52). The overall distribution of labels per drug is shown in Figure 2.

Labeling consistency

Of the 1540 drugs, 500 were represented by only a single label and thus by definition had no discrepancies. Of the 1040 drugs with more than one manufacturer’s label, 68% (n = 711) showed some discrepancies within their labeling. The degree of such discrepancy was highly variable. Overall, the mean deviation in labeled ADRs within a single clinical drug was 4.4, and the median was 0.8 (IQR 0 to 3.2). The mean difference between the maximum and minimum number of labeled ADRs per drug was 12 +/− 0.9 and the median was 2 (IQR 0 to 9). These results indicate that the majority of drugs showed relatively small differences across their labels, while a smaller number of medications showed significance variation. This latter group comprises primarily those drugs with >10 generic manufacturers (n = 139), a pattern reflected in Figure 3.

To better understand the discrepancies found in our analysis, we manually reviewed the 25 clinical drugs with the greatest range in reported adverse events. The
goal of this review was to determine the causes of these differences and whether there were any consistent labeling patterns responsible. We found 1 case in which SPLICER incorrectly skipped a section of ADRs due to formatting variations. For the remaining 24 drugs, we found the following issues in our review set (note that some labels had more than one source of inconsistency): missing data tables \((n=9)\), missing or incomplete post-marketing reports \((n=6)\), indication differences \((n=6)\), missing reactions within tables \((n=5)\), misformatted Adverse Reactions section \((n=3)\), missing pediatric data \((n=2)\), missing all adverse reactions \((n=2)\), adverse reaction data listed in incorrect section \((n=1)\), and listing of data from entirely different drug \((n=1)\). Figure 4 shows examples of our findings.

While many inconsistencies were due to faulty labeling (either editorial error or delays in updating), one source of labeling discrepancy is acceptable under current guidelines. This scenario is the presence of different indications for bioequivalent drugs. This occurs when a generic drug maker, for reasons related to patent protection, seeks approval for only some of the indications of a branded drug. In such cases, the generic labeler will remove the adverse reactions data for the clinical trials of the unapproved indication. For example, the brand drug Coreg is approved for hypertension and congestive heart failure, while the generic equivalent carvedilol is currently approved only for hypertension. In this case, the generic carvedilol label lists 48 fewer adverse reactions than the Coreg label, despite the medications being chemically and pharmokinetically identical. Such differences can also be seen when there is more than one brand drug for a given generic medication. For example, fluoxetine is branded as Prozac for depression and anxiety, and as Sarafem for premenstrual dysphoric disorder. Since different clinical trials were performed to gain these approvals, the safety warnings for these drugs will differ. Subsequent generic versions may be based on one or the other brand and thus, despite having the same label name (e.g. fluoxetine hydrochloride), contain different information. If physicians do not verify the presence of the intended indication when reviewing a generic label, they may inadvertently miss indication-specific safety data that is absent in one version of the label but present in another.

Labeling consistency by manufacturer

We identified 3382 generic product labels produced by 239 manufacturers. The number of adverse reactions on these labels differed from their brand counterparts by a median of 6.0% (IQR 0.6 to 12.7). 1977 labels (58.4%) showed some degree of deviation, with 70 labels having a deviation greater than 100% indicating that the generic label contained more adverse reactions than the corresponding brand. This latter phenomenon occurred due to either the removal of ADRs from the brand label without subsequent changes in the generic label, or to erroneous identification of the reference listed drug (e.g. generic fluoxetine compared with Sarafem instead of Prozac). In total, 186 manufacturers (77.9%) produced labels with some deviation from brand. Using logistic regression, we assessed the hypothesis that drug makers who produce more drugs will have a lower proportion of discrepant labels. Our analysis did not support this hypothesis, instead finding a positive correlation between number of drugs produced and the proportion of discrepant
ADVERSE REACTIONS

Clinical Trial Experience

Adverse events reported in >1% of patients receiving AccuNeb and more frequently than in patients receiving placebo in a four-week double-blind study are listed in the following table.

<table>
<thead>
<tr>
<th>Adverse Reaction</th>
<th>1.25 mg AccuNeb (N=115)</th>
<th>0.63 mg AccuNeb (N=117)</th>
<th>Placebo (N=117)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asthma Exacerbation</td>
<td>13</td>
<td>11.1</td>
<td>8.5</td>
</tr>
<tr>
<td>Atrial Fibrillation</td>
<td>0.9</td>
<td>0.9</td>
<td>0</td>
</tr>
<tr>
<td>Gastroenteritis</td>
<td>0.9</td>
<td>3.4</td>
<td>0.9</td>
</tr>
<tr>
<td>Cold Symptoms</td>
<td>0</td>
<td>3.4</td>
<td>0.9</td>
</tr>
<tr>
<td>Flushing</td>
<td>2.6</td>
<td>2.6</td>
<td>1.7</td>
</tr>
<tr>
<td>Lymphadenopathy</td>
<td>2.6</td>
<td>2.6</td>
<td>2.6</td>
</tr>
<tr>
<td>Sinusitis/Parotid Infection</td>
<td>0.9</td>
<td>0.9</td>
<td>0.9</td>
</tr>
<tr>
<td>Urticaria</td>
<td>0.9</td>
<td>1.7</td>
<td>1.7</td>
</tr>
<tr>
<td>Nausea</td>
<td>1.7</td>
<td>1.7</td>
<td>1.7</td>
</tr>
<tr>
<td>Abdominal Pain</td>
<td>4.4</td>
<td>4.4</td>
<td>4.4</td>
</tr>
<tr>
<td>Back Pain</td>
<td>3.4</td>
<td>2.5</td>
<td>2.5</td>
</tr>
<tr>
<td>Headache</td>
<td>3.2</td>
<td>2.1</td>
<td>2.1</td>
</tr>
<tr>
<td>Abnormal Liver Function Tests</td>
<td>7.5%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jaundice</td>
<td>2.5%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constipation</td>
<td>2.1%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

There was one case of ST segment depression in the 1.25 mg AccuNeb treatment group.

No clinically relevant laboratory abnormalities related to AccuNeb administration were seen in this study.

Figure 4.  a. The Adverse Reactions sections from two different manufacturer’s labels for albuterol sulfate solution. In the upper generic label, a table has been omitted. It can be seen in the lower brand label. b. The Adverse Reactions sections from two different manufacturer’s labels for fenofibrate. In the upper generic label, several ADRs have been omitted from the table including asthenia, flu syndrome, and diarrhea. c. The Adverse Reactions sections from two different manufacturer’s labels for dicyclomine hydrochloride tablets. In the upper generic label, the section is erroneously formatted and lists no adverse reactions. In the lower brand label, the adverse reactions are correctly listed. d. A generic manufacturer’s labeling for atenolol in which the safety data for another drug (the ACE-inhibitor Altace) has been inadvertently inserted.
labels ($p < 0.001$). Looking at the actual degree of discrepancy, we found no significant correlation between the number of drugs produced and a manufacturer’s average percent deviation from brand (see Figure 5). However, we identified a group of concerning manufacturers (right upper quadrant of the figure) who produced a large number of labels with a high degree of deviation from their branded counterparts.

**DISCUSSION**

Despite existing FDA requirement, bioequivalent medications frequently differ in their safety labeling. In some cases, the inconsistencies are minor and of limited clinical significance, in other cases they are more egregious and of genuine concern. However, with the majority of drugs showing some variation within their labeling, and the majority of generic manufacturers having produced labels discrepant from the reference brand, systemic solutions should be considered. The presence of such labeling differences may complicate physician practice, raising the possibility that important safety information may be missed even when a current label is fully reviewed. In the following discussion, we will briefly consider the clinical and regulatory implications of our findings.

Physicians frequently use labeling information, either directly or indirectly, to make prescribing decisions.
decisions. Indirect use includes referring to drug databases (e.g. First Databank) or online drug information resources (e.g. Epocrates) that utilize labeling to produce digested drug safety information. Our findings suggest that the information obtained about a drug may differ depending on which version of a label is selected. The clinical consequence of these differences will vary based on the adverse reactions themselves and their prominence within the labeling. However, physicians should be aware that branded drug labels generally contain the most comprehensive safety information. The brand version of a drug is the first to receive labeling updates and thus is not subject to the delays found in its generic counterparts. Brands also tend to have the widest range of approved clinical indications, and thus their labels will contain adverse reaction data from the widest range of clinical trials. As such, we advocate review of brand labeling by providers even if a patient is taking a generic version of a drug. We further extend our recommendation for brand label review to drug knowledge vendors.

From a practical perspective, achieving true harmonization across all versions of a drug is a tremendous challenge. With dozens of labeling changes every month (e.g. 47 revisions in November 2011\textsuperscript{13}), there will inevitably be a delay between brand label changes and updates to their generic counterparts. Under current guidelines, the FDA simply asks generic manufacturers to update labels “at the very earliest time possible” after changes to a brand label.\textsuperscript{14} In a recent draft guidance, the FDA has been more specific, requiring manufacturers to update their labels within 30 days.\textsuperscript{15} However, the question remains whether manufacturers can comply with this shortened timeframe and whether it can be practically enforced. Additionally, whereas prior to 2000 the FDA provided notifications to generic manufacturers of changes to brand labels, such notifications have now been discontinued.\textsuperscript{14} Thus, the timeliness of updates is dependent on the internal processes of individual manufacturers. Given these challenges, changes in the labeling cascade may be necessary to ensure ongoing synchronization of drug safety warnings.

### Clinical relevance of labeling discrepancies

Overall the clinical relevance of the discrepancies found in this study is unclear. Of the variations seen, 72% occurred in the clinical trials section of the label, while 28% occurred in post-marketing reports. This pattern differs from the overall frequency of ADRs found across all labels (90% in clinical trials vs. 10% in post-marketing reports). This suggests that many discrepancies may reflect simple delays in propagating labeling changes after post-marketing updates. To perform a more thorough assessment of the clinical relevance of labeling discrepancies, detailed severity and frequency data on each of the missing reactions would be necessary. At present, however, no such severity indices exist, and SPLICER does not provide consistent frequency data given the wide variation in frequency descriptors across labels (e.g. frequent, > 2%, between 3% and 10%, seen rarely in the same class, etc). We are currently at work developing a validated severity index for labeled reactions, which should support more robust characterization of such differences in the future.

### Study strengths and limitations

Our study is the first to explore labeling consistency on a large-scale basis. Prior research using manual methods has shown similar inconsistencies in the use of black box warnings.\textsuperscript{16} By developing an automated, reproducible mechanism for identifying labeling variation, we have not only expanded on previous research but created a system capable of monitoring such inconsistencies on an ongoing basis. Furthermore, by looking at variation by individual manufacturer, we have shown that while the problem is systemic, there are better and worse performers.

Our study has limitations as well. The most significant limitation is our focus on electronic SPLs as opposed to the original paper labeling. Any errors found in our study may reflect problems with formatting and translation of SPLs rather than true labeling flaws. However, technically aggregating and processing concurrently approved paper-based labels on this scale is not practically achievable. It is also worth noting that we focused on the Adverse Reactions and Post-Marketing sections of the label and did not analyze the Precautions and Boxed Warning data. These latter sections are important yet are more technically complex to achieve accurate natural language processing, and thus we sought to avoid degradation in overall data quality. Finally, SPLICER’s performance in detecting labeled adverse reactions is good, with a 93% sensitivity and 95% PPV, but these small error rates may translate into a large number of absolute errors when dealing with a large dataset. Indeed, the greater number of manufacturer’s labels per drug, the more likely...
an inconsistency will be found. In our manual review, however, we found the patterns detected by SPLICER to be correct. We are currently working on improvements to SPLICER that will enhance both its accuracy and support inclusion of additional labeling sections.

CONCLUSION

Despite existing mandate, bioequivalent medications from different manufacturers often differ in their safety labeling. This variation stands in contrast to the expectations of providers, the FDA, and, more recently, the United States Supreme Court. While the clinical significance of such labeling discrepancies remains unclear, we suggest for now that physicians review branded drug labeling even when a patient is taking a generic version of a medication. Further research will be necessary to determine the clinical importance of these labeling discrepancies as well as to identify optimal solutions for ensuring ongoing harmonization of safety data across bioequivalent drugs.

ACKNOWLEDGEMENTS

Dr. Duke had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. Funding was provided by the Regenstrief Institute. Dr. Duke is also supported by a Young Investigator’s award from the Indiana Clinical Translational Sciences Institute (KL2RR025760). Special thanks to Dr. Patrick Ryan for his thoughts on this research.

CONFLICT OF INTEREST

The authors report no conflicts of interest in the performing this research.

KEY POINTS

- FDA mandates identical labeling between bioequivalent brand and generic medications, but this consistency has never been validated
- Analyzing structured product labels, we found the majority of bioequivalent drugs show differences in safety warnings between manufacturers
- Reasons for these differences included missing tables, outdated post-marketing reports, and formatting issues
- New strategies should be considered for harmonizing bioequivalent labels

REFERENCES