

Reporting Adverse Drug Reactions in Product Labels

Therapeutic Innovation
& Regulatory Science
2016, Vol. 50(4) 455-463
© The Author(s) 2016
Reprints and permission:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/2168479016628574
tirs.sagepub.com

**Brenda Crowe, PhD¹, Christy Chuang-Stein, PhD²,
Sally Lettis, PhD³, and Andreas Brueckner, MS⁴**

Abstract

Product labels are intended to provide health care professionals with clear and concise prescribing information that will enhance the safe and effective use of drug products. In this manuscript, we offer suggestions to improve product labels. First, we recommend that product labels that include comparator data be changed to include adjusted incidence proportions (or adjusted incidence rates when needed and appropriate) for adverse drug reactions that are somewhat common. Second, we believe that including comparator incidence in product labels is a good practice, as it gives health care providers and patients appropriate information to put the absolute risks in perspective. Finally, we recommend changing the practice of reporting extremely rare events based on the "Rule of 3" in the Summary of Product Characteristics in Europe. We recommend that these adverse drug reactions be put in a separate table from other adverse drug reactions with a note that it is difficult to reliably estimate their incidences. In exceptional circumstances, it may be possible to present an estimate of their incidence based on postmarketing data. We believe the proposed changes could help product labels to better reflect the risk of a drug relative to a comparator.

Keywords

product label, product safety, harms, adverse drug reactions, adverse events

Introduction

Product labels are intended to provide health care professionals (HCPs) with clear and concise prescribing information that will enhance the safe and effective use of drug products. They are part of marketing authorization requirements and are kept updated throughout the life cycle of a medicine as new efficacy or safety data emerge. Regulatory bodies strive to make product labels appropriate and understandable, but this is no small task because of the complexity of information included.

Both the US Food and Drug Administration (FDA) and the European Medicines Agency have released changes to their product label requirements within the last 10 years.¹⁻³ Clearly both are interested in improving product labels. In that spirit of continual enhancement, we offer suggestions in this manuscript for additional improvement. The issues we highlight are somewhat statistical in nature and, while not complicated, may be new ideas to many readers. In the next sections, we explain some of the problems with current reporting practice, give possible solutions to the problems, offer additional discussion points, and summarize our recommendations.

In this manuscript, we focus on the quantitative summary of AE incidence in the Adverse Reactions (or Undesirable Effects) section of the label, rather than sections such as Warnings and Special Precautions. The latter often present more detailed information about adverse drug reactions (ADRs) of special interest and more comprehensive displays could be

included in those sections than we discuss in this manuscript. We focus on US and European Union (EU) labels, though the principles are broadly applicable to product labels around the world. In addition, we use the terms *drug* or *product* to refer to drugs, biologics, or vaccines.

Current Reporting Practice and Its Limitations

In both the US and EU, all ADRs (adverse events that are plausibly causally related to the drug) should be included in the product label.

A quantitative approach is used in the Adverse Reactions section of the US Product Label. The incidence proportion (number of patients with a new-onset event, or worsening of an event that was present at baseline, divided by number of

¹ Global Statistical Sciences, Eli Lilly and Company, Indianapolis, IN, USA

² Chuang-Stein Consulting LLC, Kalamazoo, MI, USA

³ Clinical Statistics, GlaxoSmithKline, Uxbridge, Middlesex, United Kingdom

⁴ Quantitative Safety & Epidemiology, Novartis, Basel, Switzerland

Submitted 27-Oct-2015; accepted 16-Dec-2015

Corresponding Author:

Brenda Crowe, PhD, Global Statistical Sciences, Eli Lilly and Company, Lilly Corporate Center, Indianapolis, IN 46285, USA.

Email: crowe_brenda_j@lilly.com

participants at risk for the ADR in studies) of “common” ADRs is presented with both drug and comparator information in the table. Here, *common* is defined as an ADR that occurs at or above a specified incidence (where the incidence threshold is judged to be “appropriate to the database”).² For example, adverse reactions with incidence proportion $\geq 3\%$ and more common than placebo could be presented. A comparator must be provided, except in exceptional circumstances. A placebo comparator is preferred, although if unavailable another active control can be used. In general, the studies included in the table all use the common comparator, but this may not always be the case. In the absence of either placebo or active control, good quality database estimates are permitted. Typically, a single ADR table is included; however, more than 1 table can be included when the ADR profile differs substantially from one setting or population to another. According to FDA’s guidance document,

Situations in which there may be important differences in incidence proportions include different indications, formulations, demographic subgroups, study durations, dosing regimens, and types of studies (eg, intensely monitored small studies vs. a large outcome study). In these situations, the content of the additional table or tables should be limited to only those adverse reactions for which there were meaningful differences.¹

Less common ADRs are included separately and are presented without a comparator. ADRs that are identified from spontaneous reports must be separate from the listing of ADRs identified in clinical trials.^{1(III,C)}

In contrast, in the EU Summary of Product Characteristics, ADRs are placed into frequency categories generally determined based on data from randomized controlled trials. The following convention (based on recommendations from the CIOMS III and V working groups⁴) is used for the classification of frequencies: very common ($\geq 1/10$), common ($\geq 1/100$ to $< 1/10$), uncommon ($\geq 1/1,000$ to $< 1/100$), rare ($\geq 1/10,000$ to $< 1/1,000$), and very rare ($< 1/10,000$). A placebo/comparator is typically not included and the frequency category is determined based on the crude incidence on the drug, using data from all relevant studies (eg, those with a follow-up long enough to detect the adverse reaction). The studies included do not need to include a common comparator. The EU guidance specifies that “normally, incidence proportions for the most representative exposure period (eg, 1 week, 3 months, 1 year) should be used to derive the frequency category.”³ While this recommendation makes sense for drugs that are used for a fixed duration, for drugs intended for long-term use, there is no representative duration, and in practice, studies of many differing durations are combined together to derive the incidence estimate. Only in exceptional cases, for example, a drug used for 2 very different indications (ie, oncology and nononcology), is more than 1 table covering frequencies included. If an ADR was not observed in clinical trials but determined to be causally related postauthorization based on spontaneous reports, the “Rule of

3” is required to estimate the frequency category using sample sizes from clinical trials. By this rule, if X is the total sample size of patients exposed to the drug in all relevant clinical trials, then the frequency category would be $3/X$, which is the upper limit of a 95% confidence interval for the true incidence proportion of the event in question.

There are a number of issues and limitations with the current practice.

First, typically data that are presented in comparative tables (as in the US label) are based on crude pooling of data across multiple studies (described more thoroughly in the next section). By crude pooling, we mean pooling data together as if they came from a single study. Crude pooling has the potential to give misleading results in a comparative table because it can result in an overall baseline risk that is different among treatment groups due to differing randomization allocations within a study and different study populations across studies. For example, if one study has 100 placebo and 100 drug patients who come from a population that is at high risk of a particular ADR, while another study has 100 placebo patients and 400 drug patients who are all from a population at low risk for the ADR, overall 50% (100/200) of placebo patients have a low risk for the ADR while 80% (400/500) of drug patients have a low risk for the ADR. Crude pooling will not adjust for this disparity, and the presentation of numbers in the table could look more favorable to a drug than it should. Differences that could affect the ADR incidence proportion are demographic factors (like age, sex, race) or other factors (eg, medical practice, differing time on study). Crude pooling can give misleading results when any factor that impacts the ADR incidence proportion is disproportionally represented in the overall drug and comparator cohorts. Therefore, we offer a solution (see “A Proposed Solution” section below).

Second, incidence proportions do not directly take into account the duration of treatment or the duration of follow-up, though these durations are inherently part of the calculation, because the incidence proportion calculation is specific for that duration of follow-up. Except for events that tend to occur early or not at all (eg, allergic reactions), incidence of an event generally increases with longer duration of treatment and/or follow-up. ADR incidence proportion can be higher in longer-term trials due to increased exposure or follow-up. For example, many drugs to treat diabetes have hypoglycemia as an ADR. With very long-term follow-up, nearly 100% of patients will have at least 1 episode of hypoglycemia. In some situations, it may be more useful for HCPs and patients to know approximately how many patients could expect an event in a year of use rather than just knowing that some percentage of patients experienced an event. A useful metric to report this is incidence rate (IR; number of patients with a new onset event divided by sum of person-time at risk). Another situation where IR is useful is when there is differential dropout in different treatment groups within a study. In this case, IR may give a better within-study estimate of the risk for an AE. IRs assume approximately constant risk over time. If the data do not

support this assumption, IR can give misleading results. Hence, IR should only be reported after verifying that the data appear to support this assumption, for example by showing that incidence proportion in each treatment group is somewhat constant through subintervals of time. Readers may wish to read Kraemer's manuscript,⁵ which discusses how IR can be a misleading statistic.

Third, using the Rule of 3 as a method to estimate incidence proportion for ADRs that were not reported in clinical trials is problematic. This method estimates the incidence proportion by the upper end of a range of plausible values for the incidence proportion. By doing so, ADRs that were *never* reported in clinical trials can be assigned to a frequency category that is higher than the category for ADRs that were reported in clinical trials. In the next section, we describe 2 ways that this situation can be improved.

Proposals to Improve the Adverse Reaction Section

How to Report Results Based on Data From Multiple Studies

Consider a hypothetical development program that investigated a new product for the treatment of some disease. For simplicity, we assume that the program consists of 3 trials, a phase 2 dose-response trial and two phase 3 trials. The dose-response study investigated 3 doses (low, medium, and high) against placebo in treatment-naïve patients over a 6-month period. This phase 2 study showed that efficacy increased as dose increased. Furthermore, all 3 doses offer therapeutic advantage over the placebo. The first phase 3 study investigated the new product in treatment-naïve patients, but including only the low and the medium doses plus a placebo. Treatment duration was 12 months. The second phase 3 study compared the high dose against a placebo in treatment-refractory patients, again over a 12-month period. Because it is generally harder to induce a response in a treatment-refractory population, the phase 3 study in this population included only the high dose which was anticipated to have the highest efficacy in the treatment-refractory population based on the observed effect in the treatment-naïve population in the phase 2 trial.

We consider a particular adverse event (AE), though the discussion is broadly applicable to any AE. This specific AE had higher incidence proportion in some studies than others due to differing baseline characteristics, differing treatment durations, and other factors. These other factors could be, for example, environmental factors such as seasonal and annual changes in incidence (which occur, eg, with influenza).

For now, we assume that the new product does not increase the risk of this AE. In other words, the risk for the AE is the same for patients receiving placebo and any of the 3 doses over a given period. Because of this, we will combine patients on different active doses within a study into 1 single group.

For ease of illustration, we assume that there were 100 patients in each of the 4 treatment groups in the phase 2 study, 350 patients in each of the 3 treatment groups in the treatment-naïve phase 3 study, and 500 patients in each of the 2 treatment groups in the treatment-refractory phase 3 study. These 3 studies contributed 1500 patients exposed to therapeutic doses of the new treatment to the safety database. Suppose that (1) the observed incidence proportion of patients with the AE in the shorter duration phase 2 study is 10% in each group, (2) the observed incidence proportion of patients with the AE in the treatment-naïve phase 3 study is 19% in each group, and (3) the observed incidence proportion in the treatment-refractory phase 3 study is 40% in each group. These incidence proportions are summarized in Table 1. By construction, the observed incidence proportions are the same for the new treatment and placebo within each study.

As noted in the previous section, IR may be a better way to summarize the AEs in a study in some situations. An example is when exposure differs substantially between treatment groups within some or all of the studies that will be combined. This could happen in cancer trials where longer exposure is often a sign of treatment effectiveness. In Table 2, we present the IR for each study in our hypothetical example. For illustration purpose, we assume that all patients stayed in the trials for the entire study duration; that is, there was no dropout in the study even though this is not the typical situation where incidence ratio shows the most value.

In the second-to-last row in Table 1, we include incidence proportions for each treatment group obtained from crude pooling. By crude pooling, we mean adding the number of patients receiving the same treatment that experienced the AE across the 3 studies and dividing it by the total number of patients on the same treatment. This leads to a crude incidence proportion of 24% for the new treatment group. The corresponding incidence proportion for placebo is 29%. So, while the observed incidence proportion is the same in the 2 treatment groups in each of the 3 studies, the crude incidence proportions show a different picture. The new treatment has a smaller crude incidence proportion! Without looking at the original data, some may even conjecture that the new treatment has a protective effect against the AE, especially considering that a chi-square test comparing 24% against 29% yields a 2-sided *P* value of .006. The conjecture of a protective effect is clearly erroneous. This paradoxical conclusion based on the crude incidence proportions arose from inappropriate collapsing of studies when an important variable ("study," in this case) is not included in the analysis. This phenomenon is called "Simpson's Paradox."⁶

The root of the problem is an inappropriate collapsing of studies, resulting in the pooled populations no longer being comparable between the 2 treatment groups. For example, in the pooled placebo group, treatment-refractory patients (whose AE incidence was very high) make up 53% (= 500/950) of the patients. The corresponding percentage in the new treatment group is 33% (= 500/1500). Since Study 3 patients report the

Table 1. Observed Incidence Proportions of Patients With a Particular Adverse Event in the New Treatment and Placebo Group in 3 Trials.

	New Treatment, n/N (%)	Placebo, n/N (%)	Total Patients in the Study
Phase 2 study	30/300 (10.0%)	10/100 (10.0%)	400
Phase 3 study	133/700 (19.0%)	67/350 (19.1%)	1050
Phase 3 study in refractory patients	200/500 (40.0%)	200/500 (40.0%)	1000
Incidence proportion from crude pooling	363/1500 (24.0%)	277/950 (29.0%)	2450
Study size-adjusted incidence proportion	$\frac{400}{2450} \times \frac{30}{300} + \frac{1050}{2450} \times \frac{133}{700}$ $+ \frac{1000}{2450} \times \frac{200}{500} = 26.1\%$	$\frac{400}{2450} \times \frac{10}{100} + \frac{1050}{2450} \times \frac{67}{350}$ $+ \frac{1000}{2450} \times \frac{200}{500} = 26.2\%$	

N, the total number of patients in the group; n, the number in the group that experienced the event.

Table 2. Observed Incidence Rates for a Particular Adverse Event in the New Treatment and Placebo Group in 3 Trials.

	New Treatment, n/PY (IR per 100 PY)	Placebo, n/PY (IR per 100 PY)	Total PY
Phase 2 study	30/150 (20.0)	10/50 (20.0)	200
Phase 3 study	133/700 (19.0)	67/350 (19.1)	1050
Phase 3 study in refractory patients	200/500 (40.0)	200/500 (40.0)	1000
Crude pooling	363/1350 (26.9)	277/900 (30.8)	2250
Study size-adjusted incidence rate (per 100 PY)	$(\frac{200}{2250} \times \frac{30}{150} + \frac{1050}{2250} \times \frac{133}{700})$ $+ \frac{1000}{2250} \times \frac{200}{500} \times 100 = 28.4$	$(\frac{200}{2250} \times \frac{10}{50} + \frac{1050}{2250} \times \frac{67}{350})$ $+ \frac{1000}{2250} \times \frac{200}{500} \times 100 = 28.5$	

IR, number of patients with the event per 100 person-years; n, the number in the group that experienced the event; PY, person years, total time at risk in years.

AE with a higher incidence proportion, the placebo group is expected to report a higher AE incidence proportion because it has a proportionally higher percentage of Study 3 patients. Another contributing factor is that proportionally, the new treatment group has a higher proportion of patients who were in the study for only 6 months (20% in the new treatment group versus 11% in the placebo group). These 2 factors together suggest that the new treatment group has a higher percentage of patients who, by the year they happened to participate and the treatment duration, reported a lower incidence proportion of the AE.

Table 2, which focuses on IR, shows the same phenomenon. Crude pooling gives an IR of 26.9 per 100 person-years (PY) for the new treatment versus 30.8 per 100 PY for placebo, though the new treatment IR is virtually identical to the placebo IR for each study.

A proposed solution

How can we report the combined experience of patients in studies such as these 3 studies and avoid the aforementioned paradox? (Although this event would not qualify as an ADR, we use it to illustrate the paradox. Furthermore, appropriate representation of the data can assist in the determination of which AEs are indeed ADRs.) One may argue that it is not appropriate to combine experience of patients from the 6-month study with that of patients from a 12-month study, or to combine certain studies, because of differences in the study populations. One may also argue that if there are differences in the reported incidence proportion (or rate) between studies, it is not appropriate to combine them either. For these reasons,

ADRs have been presented for different populations in some product labels. For example, adverse events are presented separately for patients with osteoarthritis and rheumatoid arthritis and pediatric patients with juvenile rheumatoid arthritis in the US Celebrex label (http://www.accessdata.fda.gov/drugsatfda_docs/label/2011/020998s033,021156s0031bl.pdf). However, to give useful data to HCPs and patients, one must strike a balance between the overwhelming complexity of presenting many study results separately (which may have the added disadvantage of low sample sizes and hence low precision in incidence estimates) and the desire to present groups of patients with important differences in separate tables.

Assuming a decision has been made to combine data from several studies and present the pooled results in a single table, a relevant question then is, if we are to combine data from multiple studies, how can we do it in such a way that we compare incidence proportions in patients who are comparable in all aspects of their conditions except for the treatment they received?

To answer this question, we propose to use the concept of standardized incidence proportions (or standardized IRs), discussed by Chuang-Stein and Beltangady.⁷ Their idea was to calculate the AE incidence proportion in a “common” population. One logical way to define this common population is to construct a population that is composed of patients from the 3 studies. The percentage of patients from each study is determined by the total number of patients in that study relative to the total number of patients from the 3 studies combined. In Table 1, the total number of patients is 2450, the percentage of patients from the phase 2 study, the phase 3 study in treatment-

naïve population, and the phase 3 study in treatment-refractory population is 16% (400/2450), 43% (1050/2450), and 41% (1000/2450), respectively. We will apply the idea proposed by Chuang-Stein and Beltangady to different measures of risk in this paper. We propose to standardize according to the study sizes, and hence refer to this particular kind of standardization as “study size–adjusted” or simply “adjusted.” Note that another option would be to use Mantel-Haenszel (MH) weights, which were also discussed by Chuang-Stein and Beltangady. Though we are not against using the MH weights, we feel that the study-size weights are easier to interpret and communicate. In addition, study size–adjusted results were often similar to those using the MH weights.

One can then calculate the AE incidence proportion for each treatment group in this common population. Since the observed incidence proportions are the same for the 2 treatment groups within each study, the study size–adjusted incidence proportion is the same for the 2 treatment groups and is obtained as 26.1% (for the new treatment group), as shown by the formula given in the last row of Table 1. By the same method, the study size–adjusted incidence proportion for placebo is 26.2%.

This 26% can be interpreted as the incidence proportion of AEs expected if the treatment (or placebo) is given to a population with a patient mix like the “common” population defined above. This number is higher than the crude incidence proportion in the new treatment group and is lower than the crude incidence proportion in the placebo group. Applying the same methodology to the IRs of Table 2, the study size–adjusted IRs were found to be 28.4 per 100 PY and 28.5 per 100 PY, respectively.

The concept of standardization is not new. It is used in estimating, for example, mortality from coronary heart disease (CHD) in a population. Since the rate changes with age, a researcher will typically investigate the mortality rate within each predefined age category in a stratified sample drawn from the population. It is not unusual to draw proportionally more subjects from certain age categories to get a better estimate because of low occurrences of CHD deaths in those categories. When this happens, the age mix in the sample will not reflect the age mix in the population, and hence the overall mortality rate from the sample, if unadjusted, will underestimate the mortality rate in the population of interest. A general practice is to present a “standardized mortality rate.”⁸ It is a weighted average of age-specific mortality rates, using the percentages of age categories in the overall population as weights.

In our example, we proposed to set the population to be patients included in all 3 trials. This is a convenient choice even though the resulting population may not truly reflect the patient population or the manner by which the new treatment will be used in the target population. However, the approach provides a meaningful way to compare the AE incidence proportions between the 2 treatment groups. It allows users of the labels to assess the additional risk attributable to the new treatment based on clinical trial experience.

Choice of Risk Measure to Combine Comparative Risk Across Studies

It is also possible and reasonable to first estimate the within-study metric (such as risk ratio, or risk difference), then use a weighting scheme to combine these. This is how most meta-analysis methods work in combining results from multiple studies. Table 3 gives an example where the relative risk is constant (doubling the risk of the new treatment compared with placebo).

For the example in Table 3, one can apply the study size–adjusting method to the relative risk in individual studies to obtain an adjusted relative risk. One can then use the latter and the adjusted incidence proportion in the control group to determine the adjusted incidence proportion in the new treatment group. For data in Table 3, the study size–based weights are 0.16, 0.43, and 0.41 (the same as shown in “A Proposed Solution” above). The adjusted relative risk for new treatment relative to placebo and adjusted incidence proportion in the placebo group are 2% and 13%, respectively, leading to an adjusted incidence proportion of 26% for the new treatment (ie, $0.13 \times 2.0 = 0.26$).

$$\text{Adjusted relative risk} = 0.16 \times 2.0 + 0.43 \times 2.0 + 0.41 \times 2.0 = 2.0$$

$$\text{Incidence in the placebo group} = 0.16 \times 0.05 + 0.43 \times 0.094 + 0.41 \times 0.20 = 0.13$$

If we apply the study size–adjusting method to the incidence proportion in the new treatment group across the 3 studies directly, we will also obtain an adjusted incidence proportion of 26% ($0.16 \times 0.10 + 0.43 \times 0.19 + 0.41 \times 0.40 = 0.26$), which is the same as the adjusted incidence proportion using relative risk as a means to combining risk across studies.

It turns out that the above property holds for all examples where the relative risk is constant across studies. In other words, when relative risk is constant across studies, applying the study size–adjusting method to the incidence proportions directly will preserve the constancy in the adjusted incidence proportions between a drug and the comparator. It can also be shown that the adjusting method preserves constant risk differences. As a result, we believe the study size–adjusting method is suitable for combining incidences (or by the same logic, IR) from multiple studies, and is therefore appropriate for presenting the overall incidence proportions (or IR) in the product label.

Updating the Rule-of-3 Estimate

As noted in the “Current Reporting Practice” section above, the EU requires that the “Rule of 3” be used to assign a frequency category for events that were not observed in clinical trials. Problems with this method are given in the “Current Reporting Practice” section above and principally relate to the fact that events that were *never* observed in clinical trials are assigned a frequency that is greater than some events that were actually observed.

Table 3. Observed Adverse Event Incidence Proportion in the New Treatment and Placebo Group in 3 Trials: New Data to Illustrate What Happens With Relative Risk of 2.0 in Each Trial.

	New Treatment, n/N (%)	Placebo, n/N (%)	Study Size	Risk Difference (New – Placebo)	Relative Risk
Phase 2 study	30/300 (10%)	5/100 (5%)	400	5%	2.0
Phase 3 study	133/700 (19%)	33/350 (9.4%)	1050	9.6%	2.0
Phase 3 study in refractory patients	200/500 (40%)	100/500 (20%)	1000	20%	2.0
Study size–adjusted incidence proportion	26%	13%		13%	2.0

N, the total number of patients in the group; n, the number in the group that experienced the event.

We propose that the Rule of 3 should not be used to estimate incidence proportion. Instead, ADRs that were identified during postmarketing experience that were never observed during clinical trials, should be reported separately from the table of ADR incidences. A note can be made to the effect that because these reactions were reported voluntarily from a population of uncertain size, it is difficult to reliably estimate their frequency. For an example of such language, see Section 6.3 of the US Prozac label highlights (<http://pi.lilly.com/us/prozac.pdf>).

In situations where patients are being monitored for a rare adverse reaction and there is a good way to track the number of events and the number of exposed patients, postmarketing data could be used to estimate incidence proportion. Tysabri and progressive multifocal leukoencephalopathy is an example of this. See Section 5 (Warnings and Precautions) of the US package insert of Tysabri (<http://www.tysabri.com/prescribingInfo>).

Discussion

Various authors (eg, Berry et al, 2002, 2004,^{9,10} Peters et al¹¹) have published research showing that the qualitative descriptions of event frequency (such as rare, common) are not terribly helpful because people interpret these terms very differently from how they were intended. In fact, Berry et al⁹ concluded, “Our results showed that qualitative descriptions led to gross overestimation of risk. Until further work is done on how patients taking the drugs interpret these terms, the terms should not be used in drug information leaflets.” This research backs up our recommendations to present quantitative information in product labels. In addition, we believe that there are various advantages to comparative tables. As noted by Crowe et al,¹² many things impact the estimate of absolute incidence proportion, such as the method of elicitation, the duration of treatment, the preferred terms that were grouped together to form the AE, the definition of “treatment emergence,” duration of studies and differences in populations and disease severity. A table that includes comparator incidence proportions (or IRs) goes a long way to helping HCPs and patients understand what part of the increase in risk is attributable to drug. Comparative tables have been recommended by Schwartz et al¹³ within a “Drug Facts Table” that includes incidence proportion on drug, comparator and the attributable risk (ie, difference in

incidence proportions). We believe that such a table is a great idea, but it should present adjusted incidences (per “A Proposed Solution” above) when needed (see the “Recommendations” section below for when adjustment is not necessary). The tables must only include studies for which the treatments were directly compared; hence, they are more appropriate for common ADRs. The tables also should make it clear what was the duration of therapy.

In the case of rare ADRs, we would want to use all relevant clinical trial information, rather than using only studies for which the treatments were directly compared (as we recommend for common events). In general, we suggest presenting this information in frequency classes without comparator information. One reason for this is that the comparator may have zero events and this may give an unrealistic comparison. If there are very few events, the point estimate may have a very large margin of error, which might not make it very useful for HCPs or patients. But the margin of error may be decreased by using all relevant studies.

There are many outstanding problems that still need to be addressed. We briefly note some of them here, though we consider detailed discussion of their solutions out of scope for this paper.

First, we focus on common ADRs whose incidences appear in the Adverse Reactions Section of a product label in the US. There may be situations when neither incidence proportions nor incidence rates are adequate because the drug effect changes over time, or recurrence of an ADR represents an important treatment consideration. When an ADR is an AE of special interest, more attention is often needed to understand the nature and the time course of the event. In both cases, this could include examining incidence over different exposure durations. For an example of this, readers can view the Tysabri US package insert (<http://www.tysabri.com/prescribingInfo>). In Table 1 of the Tysabri label, estimated incidence of progressive multifocal leukoencephalopathy over different exposure durations (1-24 months, 25-48 months, and 49-72 months) is given for patient subgroups defined by 3 risk factors in the US. Statistical methods could include the piecewise exponential model.^{14,15} Additionally, O'Neill^{16,17} discusses methods for analyzing and summarizing AEs from clinical trials with a special emphasis on their time of occurrence and the impact of differential subject exposure and pretreatment status.

In this manuscript, we do not go into detail on the graphical and tabular methods described in O'Neill^{16,17} because our focus is on the Adverse Reactions (or Undesirable Effects) section of the label, rather than sections such as Warnings and Special Precautions. The latter often present more detailed information about ADRs of special interest, and more comprehensive displays could be included in this section. We further caution that the interpretation of proportions in separate periods of time could be influenced by having a “survivor” population. That is, those who can tolerate the drug do not drop out early. Incidences in later periods can look lower because of this.

For some events (such as hypoglycemia with insulin use), understanding AE recurrence may be very important. There is a rich literature on the analysis of recurrent events. The analysis can be quite complex and is typically reserved for events of special interest.

Second, methods for analyzing safety data typically do not take into account missing data/withdrawal information. While there is considerable guidance and debate recently regarding handling missing data for efficacy data,¹⁸⁻²⁰ guidance for how to handle missing data for safety analysis is paltry. These issues have come up at a recent FDA advisory committee meeting in which there was concern that “differences in patient characteristics between dropouts on the placebo and active arms induce bias in safety comparisons.”^{21(p. 253)} Finally, there is no consistent methodology for selection of ADRs. All ADRs should be included in the product label if they are at least possibly causally related, based on, for example, their comparative incidence in clinical trials, or on findings from epidemiological studies and/or on an evaluation of causality from individual reports. Crowe et al¹² presented one way of using a systematic approach, but currently there is no industry-wide solution to this problem. Some of the problems noted in this manuscript extend to establishing causality. For example, if crude pooling is used to assess clinical-trial data for causality, mistakes could occur.

An additional challenge is how to compare across numerous available drugs. The US FDA guidance recommends that the following statement, or an appropriate modification, should precede the presentation of adverse reactions from clinical trials:

Because clinical trials are conducted under widely varying conditions, adverse reaction rates observed in the clinical trials of a drug cannot be directly compared to rates in the clinical trials of another drug and may not reflect the rates observed in practice.¹

As noted in the FDA recommendation, there are many things that impact the actual incidence estimate of an AE. For example, the collection and coding of AEs could impact the estimate. Furthermore, culture often plays an important role in AE reporting. While we believe that warnings not to use product labels to directly compare across compounds should be heeded, we also think that including numerical comparative tables will help put incidence rates or proportions associated with different compounds in perspective. For example, if the labels of two

drugs for the same indication include quantitative information on a placebo or a common active control, users of the labels could examine responses on the comparator to gauge the level of additional risk associated with the drugs.

On December 18, 2014, FDA proposed to amend its prescription drug and biological product labeling regulations to require electronic distribution of the prescribing information intended for HCPs.²² Prescribing information is currently distributed in paper form on or within the package from which a prescription drug or biological product is dispensed. Not only was FDA proposing that prescribing information intended for HCPs be distributed electronically, but it also proposed that paper form not be permitted, except as provided by regulations. The rationale for the change is to help ensure that the most current prescribing information is publicly accessible for the safe and effective use of human prescription drugs.

We welcome the proposed change because it opens up the possibility of a more comprehensive coverage of both the benefit of and adverse reactions to a pharmaceutical product. It allows for update of the information when data from additional randomized trials and other data sources become available. By virtue of the distribution route, the electronic file of the prescribing information could contain links to other important documents such as FDA communications to HCPs on emerging information about a product. In addition, since space is no longer a limiting factor, future prescribing information (after the implementation of the proposed change) could include anticipated timing of important adverse reactions as well as the management of these reactions. Differences in responses among subgroups with good biologic explanation could also be covered in more detail. These additions will make the prescribing information more useful to HCPs. The latter represents an important step in the right direction as treatment decisions are becoming more personalized.

Recommendations

In summary, we have 3 main recommendations. First, we recommend that product labels be changed to include adjusted incidence proportions (or incidence rates if needed and appropriate) for ADRs that are candidates for inclusion in the Adverse Reactions section of a product label. As illustrated in the “Proposals to Improve the Adverse Reaction Section” above, the proposed adjustment method is easy to implement when study-level data are available. For older products, the implementation could be challenging if study-level data are no longer available. We therefore recommend that our adjustment method be applied to recent labels and certainly to new labels.

There are a couple of occasions when continuing with the current practice in data pooling will not distort the comparison of ADR incidences between a new drug and a comparator. They occur when the ratio of patients receiving the new drug to that receiving the comparator is approximately the same across all the studies included in the integrated summary, or when the incidences of AEs in the comparator group are nearly

the same across the same set of studies. On such occasions, naïve pooling will produce similar results to the proposed adjustment method.

During the label negotiation between a sponsor and regulators, a sponsor will submit a proposal on whether to present ADRs in one table or in multiple tables for different patient populations. In addition, a sponsor may propose to present incidence rates instead of incidence proportions, or both, in the ADR tables. We advocate applying the proposed adjustment method to each set of the ADR tables as long as the conditions negating the needs for adjustment stated above are not met.

Second, for common events, we believe that including comparator incidence in product labels is a good practice that should be more widely implemented because it gives HCPs and patients appropriate information to put the absolute risks in perspective.

Finally, we recommend changing the EU practice of reporting extremely rare events based on the “Rule of 3.” These ADRs could be put in a separate table from other ADRs with a note that it is difficult to reliably estimate their frequency. In exceptional circumstances, it may be possible to present an estimate for their incidence based on postmarketing data.

In conclusion, it is important to put enough information in the product label to make it useful and understandable but not so much that the information is overwhelming. We advocate implementation of these recommendations as they can help make product labels better reflect the risk of a drug relative to a comparator.

Author Note

The views expressed in this article are those of the authors and not necessarily of the companies involved.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: B.C., S.L., and A.B. are full-time employees of Eli Lilly and company, GlaxoSmithKline, and Novartis, respectively, and conducted this study as part of their employment duties.

References

1. US Food and Drug Administration. Adverse Reactions Section of labeling for human prescription drug and biological products—content and format. <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm075057.pdf>. Published 2006. Accessed October 3, 2015.
2. US Food and Drug Administration. Physician Labeling Rule Requirements for Prescribing Information. <http://www.fda.gov/drugs/guidancecomplianceregulatoryinformation/lawsactsandrules/ucm084159.htm>. Published 2006. Accessed October 3, 2015.
3. European Commission; Enterprise and Industry Directorate-General. A guideline on summary of product characteristics (SmPC). http://ec.europa.eu/health/files/eudralex/vol-2/c/smpc_guideline_rev2_en.pdf. Published 2009. Accessed October 3, 2015.
4. CIOMS Working Groups III and V. *Guidelines for Preparing Core Clinical-Safety Information on Drugs*. Geneva: Council for International Organizations of Medical Sciences; 1999.
5. Kraemer HC. Events per person-time (incidence rate): a misleading statistic? *Stat Med*. 2009;28:1028-1039.
6. Rucker G, Schumacher M. Simpson’s paradox visualized: the example of the rosiglitazone meta-analysis. *BMC Med Res Methodol*. 2008;8.
7. Chuang-Stein C, Beltangady M. Reporting cumulative proportion of subjects with an adverse event based on data from multiple studies. *Pharm Stat*. 2011;10:3-7.
8. Everitt B, Skrondal A. *The Cambridge Dictionary of Statistics*. New York, NY: Cambridge University Press; 2010.
9. Berry DC, Knapp P, Raynor DK. Provision of information about drug side-effects to patients. *Lancet*. 2002;359:853-854.
10. Berry D, Raynor T, Knapp P, Bersellini E. Over the counter medicines and the need for immediate action: a further evaluation of European Commission recommended wordings for communicating risk. *Patient Educ Couns*. 2004;53:129-134.
11. Peters E, Hart PS, Tusler M, Fraenkel L. Numbers matter to informed patient choices: a randomized design across age and numeracy levels. *Med Decis Making*. 2014;34:430-442.
12. Crowe B, Brueckner A, Beasley C, Kulkarni P. Current practices, challenges, and statistical issues with product safety labeling. *Stat Biopharm Res*. 2013;5:180-193.
13. Schwartz LM, Woloshin S, Welch HG. Using a drug facts box to communicate drug benefits and harms: two randomized trials. *Ann Intern Med*. 2009;150:516-527.
14. Lambert PC, Smith LK, Jones DR, Botha JL. Additive and multiplicative covariate regression models for relative survival incorporating fractional polynomials for time-dependent effects. *Stat Med*. 2005;24:3871-3885.
15. Crowther MJ, Riley RD, Staessen JA, Wang J, Gueyffier F, Lambert PC. Individual patient data meta-analysis of survival data using Poisson regression models. *BMC Med Res Methodol*. 2012;12:34.
16. O’Neill RT. Statistical analyses of adverse event data from clinical trials. Special emphasis on serious events. *Drug Inf J*. 1987;21:9-20.
17. O’Neill RT. Assessment of safety. In: Peace KE, ed. *Biopharmaceutical Statistics for Drug Development*. New York, NY: Marcel Dekker; 1988.
18. O’Neill RT, Temple R. The prevention and treatment of missing data in clinical trials: an FDA perspective on the importance of dealing with it. *Clin Pharmacol Ther*. 2012;91:550-554.
19. Little RJ, D’Agostino R, Cohen ML, et al. The prevention and treatment of missing data in clinical trials. *N Engl J Med*. 2012;367:1355-1360.

20. Panel on Handling Missing Data in Clinical Trials and Committee on National Statistics, Committee on National Statistics, Division of Behavioral and Social Sciences and Education, National Research Council. *The Prevention and Treatment of Missing Data in Clinical Trials*. Washington, DC: National Academies Press; 2010.
21. US Food and Drug Administration. Pulmonary Allergy Drugs Advisory Committee Meeting: FDA briefing document. <http://www.fda.gov/downloads/AdvisoryCommittees/CommitteesMeetingMaterials/Drugs/Pulmonary-AllergyDrugsAdvisoryCommittee/UCM367411.pdf>. Published 2013. Accessed October 3, 2015.
22. US Food and Drug Administration. Electronic distribution of prescribing information for human prescription drugs, including biological products. <https://www.federalregister.gov/articles/2014/12/18/2014-29522/electronic-distribution-of-prescribing-information-for-human-prescription-drugs-including-biological>. Published 2014. Accessed October 3, 2015.