Understanding the Frequency and Severity of Side Effects: Linguistic, Numeric, and Visual Representations

Rob Hubal

RTI International
PO Box 12194
Research Triangle Park, NC 27709-2194
rhubal@rti.org

Ruth S. Day

Duke University
Box 90085
Durham, NC 27708-0085
ruthday@duke.edu

Abstract

Side effects for prescription drugs vary in their severity and frequency of occurrence. Understanding the status of a given drug on both these dimensions is important for physicians during the prescribing process, for regulators and industry in the approval and safety review process, and for patients in the compliance process. There is a wide variety of terms used to describe severity and frequency information in both professional information sources (such as the approved label) and patient sources (such as pharmacy leaflets). The experiments reported here examine how people understand these terms, whether laypersons interpret them in the same ways as professionals, and the consequences of providing terms in alternative linguistic, numeric, and visual forms. This work holds implications for risk communication for healthcare providers and patients, the needs of low-literacy and low-numeracy audiences, and health literacy in general.

Introduction

Prescription drug information sources such as the Food and Drug Administration (FDA) approved label and patient pharmacy leaflets use a wide variety of terms to describe side effects. For example, the severity of a given side effect might be given as "serious" or "mild", while its frequency of occurrence might be given as "common" or "rare". How healthcare providers interpret these terms can affect their prescribing and monitoring decisions, while patient interpretations can affect their compliance with treatment regimens. A mismatch in interpretation between providers and patients can also cause problems in communication and compliance.

This study examined terms used to describe side effects, (also called adverse events or adverse drug reactions) for prescription drugs. The terms used appear in professional drug information sources such as the FDA-approved label (often reprinted in the *Physicians' Desk Reference* or PDR), the United States Pharmacopeia *Drug Information* (USP-DI), and pharmacy leaflets (also known as consumer medical information or CMI) provided to patients. This work is part of a larger research program on comprehension of drug benefits and risks across a wide range of information sources such as pharmacy leaflets [1], professional labels, Medication Guides, patient package inserts, over-the-counter packages, and direct-to-consumer advertising in TV ads, print ads, and the Internet [2]. The present study examined perception and interpretation of side effect terms, whether laypersons and professionals interpret them in the same ways, and the consequences of providing the terms in alternative linguistic, numeric, and visual forms.

Selection of Terms

We extracted frequency and severity terms from the literature for physicians (PDR), pharmacists (USP-DI), and patient pharmacy leaflets. Initial analyses revealed that the terms were not used consistently within or across sources. Also, there were multiple semantic categories and levels within both sets of terms. For example, Table 1 shows that some frequency terms refer to the number of people who experience a side effect while others refer to incidence (commonality of occurrence). Some frequency terms provide no level information at all (as shown by the "?" in the table), such as "side effects include". Similarly, Table 2 shows two sample semantic categories for severity terms and multiple levels. Additional semantic categories and other details are provided below. We combined terms across information sources and conducted separate studies on frequency and severity terms.

Table 1. Sample Frequency Terms

	Type of Term	
	# of People	Incidence
Highest	*Some people*	*More common*
	A few people	*Less common*
		Rare
Lowest	*Very few people*	*Very rare*
?	*You may experience*	*Include*

Table 2. Sample Severity Terms

	Type of Term	
	Descriptive	Action
Highest	*Life threatening*	*Call doctor immediately*
	Dangerous, serious, severe	*Discuss with doctor before continuing*
	Worrisome	*Tell doctor at next visit*
Lowest	*Bothersome, mild*	*No medical attention*

Frequency Experiment

Participants judged 38 frequency terms from the literature under one of two task conditions. Those in the Numeric Task estimated the number of cases that a frequency term represented, while those in the Visual Line Task placed side effects along a frequency line. Both tasks used the same anchor terms, from "never" to "always".

Tasks. Participants in the Numeric Task were told this was a survey on perception of medication information. The focus was on information about side effects and how often they occur. Thus one drug might "always" produce a given side effect, while another might "never" produce it. Participants heard sentences with frequency terms embedded in the sentences and were to try to determine about how often they thought the side effects would occur. (*Note*: Specific side effects were not mentioned, so they played no role in the task.) Participants might not know the precise number of cases for a given term but were to just give a 'ballpark' number, their best guess about the approximate number of people out of 100 who would experience the side effect. They answered using a scale from 0 ("none") to 100 ("all") in steps of 10, with the addition of 0-1 at the low end and 99-100 at the high end. They had five seconds to provide their answer for each item.

Participants in the Visual Task received the same instructions except they were given each frequency term on a slip of paper and had to place it somewhere along a line on a bulletin board to indicate the percentage of cases it represented. One end of the line was defined as "always" and the other "never". The distance from the "never" anchor was then measured.

Frequency terms. The 38 items were presented in random order embedded in the following sentence frames:

- This side effect is [*frequency term*].

  common, frequent, infrequent, likely, possible, rare, uncommon, unlikely, usual (9 total)

- This side effect [*occurrence term*].

  is manifested, occurs (2 total)

- This side effect has been [*observation term*].

  noted, observed, reported, shown (4 total)

- You may [*experience verb*] this side effect.

  develop, experience, feel, have (4 total)

- [*number term*] people experience this side effect.

  a minority of, a signification proportion of, few, many, most, several, some (7 total)

- This drug [*causation verb*] this side effect.

  causes, is associated with, produces (3 total)

- Side effects include X, Y, Z. (1 total)

- This side effect is [*degree term*] common.

  less, more, somewhat, very (4 total)

- This side effect is very rare. (1 total)

- You [*probability term*] experience this side effect.

  can, may, might (3 total)

These frames represent multiple semantic categories for the frequency terms, including:

- Frequency (common, frequent, infrequent, likely, possible, rare, uncommon, unlikely, usual) (9 total)
- Probability (can, may, might) (3 total)
- Occurrence (is manifested, occurs) (2 total)
- Observation (noted, observed, reported, shown) (4 total)
- Experience (develop, experience, feel, have) (4 total)
- Number (a minority, a significant proportion, few, many, most, several, some) (7 total)
- Degree (less, more, somewhat, very) (4 total)
- Causation (causes, is associated with, produces) (3 total)
- Inclusion (includes) (1 total)

Participants. A total of 222 Duke undergraduate students participated in these studies as part of a course requirement. These individuals could be expected to have only a layperson's (as opposed to a medical professional's) understanding of frequency and severity terms.

Each participant was assigned to either the Numeric (N=206) or Visual (N=16) Task. Those in the Numeric Task were tested in a large group while those in the Visual Task were tested individually (owing to the use of the physical apparatus). Each judged both frequency and severity terms (as described below) using the same task condition (Numeric or Visual).

Overall results. Overall frequency estimates were very high. The mean frequency (i.e., number of cases out of 100) was 49.1 (SD = 20.1), while the mean distance from the "never" anchor, converted to a percentage distance to the "always" anchor, was 48% (SD = 16%). Table 3 illustrates how the frequency terms were arrayed, showing the average estimate and spacing given for each frequency term. The Pearson correlation between the Numeric and Visual tasks (a between-subjects comparison) is a highly significant $r = 0.94$ ($p < 0.01$).

The frequency terms clustered, that is, participants perceived multiple terms as signifying similar occurrences. Table 4 shows results from a simple pairwise clustering analysis of Numeric Task estimates using significance set at $p < 0.005$ to correct for Type I errors.

Table 3. Frequency Estimation

Term	Numeric Task Value (SD)	Visual Task % Distance (SD)
Always (ANCHOR)		
Produces	84.7 (23.8)	76 (24)
Causes	82.4 (25.3)	82 (23)
Most	79.4 (18.2)	77 (16)
Very	79.0 (20.4)	79 (18)
Frequent	73.8 (17.5)	80 (15)
Likely	73.2 (17.4)	69 (17)
Usual	72.0 (21.3)	71 (14)
A significant proportion	70.7 (20.3)	74 (16)
Common	70.5 (17.9)	66 (15)
Many	69.9 (17.3)	71 (17)
Is associated with	69.4 (24.3)	52 (21)
Include	67.0 (31.1)	50 (16)
Manifested	65.5 (27.5)	60 (26)
Occurs	65.4 (29.6)	71 (25)
More	63.8 (16.6)	68 (15)
Develop	53.4 (19.3)	64 (23)
Somewhat	52.2 (18.0)	35 (13)
Have	51.6 (21.5)	55 (25)
Experience	50.6 (22.4)	61 (25)
Feel	48.6 (20.2)	47 (18)
Can	48.5 (21.4)	38 (17)
May	47.2 (21.7)	35 (19)
Several	44.6 (25.2)	57 (17)
Might	43.1 (20.5)	31 (15)
Possible	38.0 (21.7)	38 (16)
Shown	37.9 (23.5)	45 (18)
Reported	36.7 (24.1)	41 (20)
Noted	35.0 (23.1)	42 (21)
Some	33.0 (14.6)	36 (12)
Less	31.8 (15.7)	20 (9)
Observed	29.4 (22.6)	46 (18)
A minority of	20.5 (13.0)	18 (10)
Infrequent	16.2 (15.9)	15 (7)
Few	15.8 (11.2)	18 (9)
Uncommon	13.3 (16.1)	14 (8)
Unlikely	13.3 (14.0)	13 (7)
Rare	9.7 (13.6)	10 (6)
Very rare	9.6 (15.2)	5 (2)
Never (ANCHOR)		

Table 4. Frequency Term Clustering

Always (ANCHOR)
Causes, Most, Produces, Very
A significant proportion, Common, Frequent, Include, Is associated with, Likely, Manifested, Many, More, Occurs, Usual
Can, Develop, Experience, Feel, Have, May, Might, Several, Somewhat
Less, Noted, Observed, Possible, Reported, Shown, Some
A minority of
A few, Infrequent
Uncommon, Unlikely
Rare, Very rare
Never (ANCHOR)

Severity Experiment

Tasks. The severity experiment was performed in the same manner as the frequency experiment. Participants in the Numeric Task were told that this was a study on perception of medication information, and the focus was on information about side effects, but this time on how severe they were. Thus, a given side effect might have "no" effect on a patient, while another might have "maximum" effect. Participants heard sentences about side effects and were to try to determine how severe they thought each side effect would be. To do so, they were to select a number from 0 ("none") to 100 ("maximum"), using the same steps as in the frequency experiment. They were to give a 'ballpark' number to indicate the approximate severity, even if they did not know the precise degree of severity.

Participants in the Visual Task were given the same instructions except that they were to place each term along a line on a bulletin board to indicate severity, with anchor points "maximum" and "none". The distance from "none" was then measured.

Severity terms. We derived 19 severity terms from the same literature as for frequency terms. There were two semantic categories, descriptive terms (e.g., annoying) and action terms (what they would do if they experienced the side effect while taking the medication). The two sentence frames were of the form:

- This side effect is [*descriptive term*].

  annoying, bothersome, dangerous, fatal, inconsequential, life-threatening, mild, serious, severe, troublesome, worrisome (11 total)

- If you experience this side effect, [*action term*].

  call your doctor, call your doctor immediately, continue to monitor symptoms, get emergency help, ignore symptoms, it usually requires no medical attention, rush to the ER, tell your doctor at the next visit (8 total)

Overall results. Overall severity estimates were very high, with a mean severity of 60.8 (SD = 28.3) for the Numeric Task and a mean distance to the "maximum" anchor of 58% (SD = 13%) for the Visual Task. Table 5 illustrates how the severity terms were arrayed, showing the average estimate and spacing given for each severity term. The Pearson correlation between the Numeric and Visual tasks was a highly significant $r = 0.85$ ($p < 0.01$).

The severity terms clustered, that is, participants perceived multiple terms as signifying similar severities. Table 6 shows results from a pairwise clustering analysis of Numeric Task estimates.

Discussion and Implications

The findings reported here hold several implications for effective risk communication.

Level effects. Overall estimates of both frequency and severity terms were very high. For example, "common" side effects were perceived to occur in about 70% of cases, about the same level as "likely". Although drug information

Table 5. Severity Estimation

Term	Numeric Task	Visual Task
	Value (SD)	% Distance (SD)
Maximum (ANCHOR)		
Fatal	92.9 (24.2)	93 (4)
Get emergency help	86.9 (21.6)	90 (6)
Serious	79.6 (17.7)	77 (12)
Life threatening	79.4 (26.9)	92 (4)
Call doctor immediately	77.0 (30.4)	82 (11)
Dangerous	75.3 (29.0)	82 (9)
Severe	71.0 (35.1)	83 (10)
Rush to ER	65.5 (38.7)	94 (3)
Troublesome	64.6 (27.3)	50 (19)
Worrisome	63.5 (21.2)	47 (17)
Annoying	63.0 (29.2)	36 (16)
Continue to monitor symptoms	59.6 (25.5)	44 (15)
Usually requires no medical attention	50.7 (36.4)	23 (23)
Call doctor	50.5 (30.9)	69 (15)
Tell doctor at next visit	48.6 (31.4)	55 (18)
Mild	40.1 (25.3)	21 (14)
Inconsequential	30.7 (38.8)	18 (19)
Bothersome	30.4 (24.2)	41 (21)
Ignore symptoms	25.0 (24.7)	8 (7)
None (ANCHOR)		

Table 6. Severity Term Clustering

Maximum (ANCHOR)
Fatal
Get emergency help
Life threatening, Serious
Call doctor immediately, Dangerous, Severe
Annoying, Rush to ER, Troublesome, Worrisome
Continue to monitor symptoms
Call doctor, Tell doctor at next visit, Usually requires no medical attention
Mild
Bothersome, Ignore symptoms, Inconsequential
None (ANCHOR)

sources vary in whether they provide any estimates for such terms and what those estimates are, their levels are generally lower than the perception results reported here. For example, the PDR uses the terms "reported" and "observed" for rates between 0-5% for ten target drugs, while laypersons in the Numeric Task estimated these same terms as 37% and 29%, respectively.

These experiments show that laypersons substantially overestimate both the probability and severity of side effects across a wide range of descriptor terms. Therefore, when healthcare providers and patients discuss the likelihood or severity of a given side effect, they may be interpreting the terms very differently.

Term equivalence. The finding of clusters within a set of terms suggests that laypersons perceive some terms to be equivalent in degree of frequency or severity. Selecting appropriate equivalents (within clusters) versus contrasts (across cluster boundaries) is important, both for written and oral risk communication. The possibility that the clusters are different for professionals may also limit effective communication between healthcare providers and patients.

Meaning of terms. There are few resources that define what frequency and severity terms mean. This may be a function of the wide variety of terms used to signify the frequency of occurrence and severity of side effects, as well as the inconsistencies found both within and across sources. Indeed, a recent FDA guidance document to the pharmaceutical industry [3] recommends "nonspecific terms that lack a commonly understood or precise meaning" be avoided, and implies that terms demanding "vague and subjective judgment" do not provide meaningful information.

For frequency, a side effects probability scale has been developed [4], though specific to a given drug and patient. Ten questions asked the patient for information such as "Did the adverse drug reaction improve when the drug was discontinued?" and "Was the reaction more severe when the dose was increased, or less severe when the dose was decreased?". A total score of 9 indicated a "highly probable" side effect, a score between 5 and 8 indicated a "probable" side effect, a score of 1 to 4 indicated a "possible" side effect, and a score of 0 indicated a "doubtful" probability.

For severity, one portrayal categorizes side effects into three levels [5]: mild (bothersome but requiring no change in therapy); moderate (requiring change in therapy, additional treatment, or hospitalization); and severe (disabling or life-threatening).

An Institute of Medicine report notes that the FDA describes "serious and complex" reactions in this way [6]: they result in death, are life-threatening, require hospitalization, prolong hospitalization, cause disability, cause congenital anomalies, or require intervention to prevent permanent injury.

A questionnaire for side effects, in lay terms, was developed to be used in computer-assisted interviews [7]. However, frequency and severity terms were not addressed.

Representation, literacy, and numeracy. Results from the Numerical and Visual Tasks were highly correlated – in fact, nearly identical. This type of equivalence has not always been found (e.g., [8]), but may be due to the use here of a linear visual scale and/or to the absence of specific side effects in sentence frames. Nevertheless, the present results suggest different strategies for communicating with individuals with various levels of literacy and numeracy. Alternative representations [9] of the same side effect information for a given drug can be used for different populations. For example, percentages might be provided for those who understand numbers, or for those situations where it is likely that numeric versus linguistic representation will lead to differences in estimation of risk [10,11]. For other individuals, a more spatial display can be used, such as a line with end-points and individual side effects placed either closer to an end or in the middle. Simple end-point descriptors would be used for all representations and could be either linguistic or pictorial. Although highly literate individuals might not need spatial layouts to understand the information, such representations might still enable them to grasp the information more quickly and remember it better.

References

[1] Day, R.S. 1998. Optimizing Patient Comprehension through Medicine Leaflets: Cognitive Experiments. In Hartzema, A.G., Tolleson-Rinehart, S., Sleath, B.L., Day, R.S., and Bush, P.J. eds. *Optimizing Patient Comprehension through Medicine Information Leaflets*. Rockville, MD: United States Pharmacopeia Drug Quality and Information Program.

[2] Day, R.S. 2006. Comprehension of Prescription Drug Information: An Empirical & Theoretical Research Program. In Proceedings of the AAAI Spring Symposium, Workshop on Argumentation for Consumers of Healthcare. Menlo Park, CA: AAAI Press.

[3] Food and Drug Administration. 2006. Guidance for Industry: Adverse Reactions Section of Labeling for Human Prescription Drug and Biological Products – Content and Format. Rockville, MD: U.S. Department of Health and Human Services.

[4] Naranjo, C.A., Busto, U., Sellers, E.M., Sandor, P., Ruiz, I., Roberts, E.A., Janecek, E., Domecq, C., and Greenblatt, D.J. 1981. A Method for Estimating the Probability of Adverse Drug Reactions. *Clinical Pharmacology and Therapeutics* 30:239-245.

[5] Calis, K.A. 2004. Clinical Analysis of Adverse Drug Reactions: A Primer for Clinicians. *Hospital Pharmacy* 39(7):697-712.

[6] Chrvala, C.A., and Sharfstein, S. eds. 1999. *Definition of Serious and Complex Medical Conditions*. Washington, DC: Institute of Medicine (IOM).

[7] Corso, D.M., Pucino, F., DeLeo, J.M., Calis, K.A., and Gallelli, J.F. 1992. Development of a Questionnaire for Detecting Potential Adverse Drug Reactions. *The Annals of Pharmacotherapy* 26(7-8):890-896.

[8] Svensson, E. 2000. Comparison of the Quality of Assessments Using Continuous and Discrete Ordinal Rating Scales. *Biometrical Journal* 42:417-434.

[9] Day, R.S. 1988. Alternative Representations. In Bower, G.H. ed. *The Psychology of Learning and Motivation*. New York: Academic Press, 22, 261-305.

[10] Gurmankin, A.D., Baron, J., and Armstrong, K. 2004. The Effect of Numerical Statements of Risk on Trust and Comfort with Hypothetical Physician Risk Communication. *Medical Decision Making* 24(3):265-271.

[11] Knapp, P., Raynor, D.K., and Berry, D.C. 2004. Comparison of Two Methods of Presenting Risk Information to Patients about the Side Effects of Medicines. *Quality & Safety in Health Care* 13:176-180.