Does the Drug Facts Label for nonprescription drugs meet its design objectives? A new procedure for assessing label effectiveness

Michael P Ryan and Reagan N Costello-White

Abstract

We demonstrate an expanded procedure for assessing drug-label comprehension. Innovations include a pretest of drug preconceptions, verbal ability and label attentiveness measures, a label-scanning task, a free-recall test, category-clustering measures, and preconception-change scores. In total, 55 female and 39 male undergraduates read a facsimile Drug Facts Label for aspirin, a Cohesive-Prose Label, or a Scrambled-Prose Label. The Drug Facts Label outperformed the Scrambled-Prose Label, but not the Cohesive-Prose Label, in scanning effectiveness. The Drug Facts Label was no better than the Cohesive-Prose Label or the Scrambled-Prose Label in promoting attentiveness, recall and organization of drug facts, or misconception refutation. Discussion focuses on the need for refutational labels based on a sequence-of-events text schema.

Introduction

The burden that a nonprescription drug label bears is to inform consumers how to self-medicate safely and effectively without the guidance of a physician or pharmacist (Soller, 1998; Sutherland, 2010). The Food and Drug Administration (FDA, 1999) made a formal effort to lighten that burden by standardizing the content, order, and format of all nonprescription drug labels with the now-familiar Drug Facts Label (DFL). The FDA continued this effort by requiring pharmaceutical drug manufacturers to conduct label comprehension studies for all nonprescription drugs. In their *Guidance to Industry* (FDA, 2010), the FDA suggests that questions about the primary communication objectives of a proposed label be posed to potential consumers in one-to-one interviews using an "open-label" procedure. Respondents are

> … told before the study starts that they can refer to the label during questioning. Questions can begin with the statement "according to the label," however, subjects should not be overly prompted to look at the label during questioning. (FDA, 2010: 9)

The primary objective of this study is to demonstrate an expanded label comprehension procedure that addresses critical deficiencies in the current open-label procedure.

A study designed by Raymond et al. (2002) with guidance from the FDA offers a rare published example of the open-label procedure. Raymond et al. assessed label comprehension of prototype DFL panels on a two-pill package of a nonprescription emergency contraceptive (Plan B Two-Step) and the six-panel patient insert inside the package. The text on the outside of the package conformed to DFL standards; the text insert inside the package provided

The University of Texas at San Antonio, USA

Corresponding author:
Michael P Ryan, Department of Psychology, The University of Texas at San Antonio, San Antonio, TX 78249, USA.
Email: Michael.Ryan@utsa.edu

further information in a question–answer format (e.g. "Who should not use Plan B?"). Raymond et al. made use of the 11 primary communication objectives of the drug to develop 30 interviewer questions for assessing the communication effectiveness of the proposed labeling. In individual interviews, eligible respondents first briefly examined the external DFL panels as if considering their purchase. With the package out of sight, respondents then had to explain the purpose of the drug. The interviewer next asked respondents five more questions (e.g. "According to the label, should Plan B be used as regular birth control?"), allowing them to consult the package label as necessary. Afterward, respondents were asked to open the package and review the six-panel package insert as if they were about to use the product. Finally, respondents answered 24 more questions (e.g. "A woman had unprotected sex a week ago and then used Plan B to prevent pregnancy. Was this a correct use of Plan B?"), referring to the package labeling or the package insert as necessary.

Raymond et al. adhered carefully to the FDA's (2010) *Guidance* for the conduct of a label comprehension study. In addition, they note that their "study design and questionnaire were heavily influenced by the Food and Drug Administration's comments on early drafts of the protocol" (2002, p. 342). The resulting study is a carefully conceived and well-conducted open-label study. However, we believe that there are significant deficiencies in the open-label protocol itself. Our aim in this study is to identify potential improvements in the protocol that can increase the quality of the data obtained from label comprehension studies.

We identify here seven deficiencies in the open-label protocol that may limit its utility for assessing the comprehension of a set of drug facts. First, there is no pretest of respondents' prior knowledge of contraceptive drugs. Second, there is no measure of verbal ability. Third, there is no measure of the care taken by respondents in attending to the text of the DFL or the package insert. Fourth, there is no assessment of the ease with which information could be located on the label in response to interviewer questions. Fifth, there is no measure of respondents' ability to retrieve important label information from long-term memory. Sixth, there is no formal evaluation of the subjective features of the label (e.g. Perceived Organization or Authoritativeness) as a medication guide. And seventh, there is no posttest of respondents' knowledge and beliefs to determine whether the label had confirmed or corrected preconceptions respondents may have had about the drug. We consider these deficiencies of the open-label procedure in the following section. We then describe in our "Methods" section the procedure we have developed to remedy these deficiencies.

Deficiencies in the label comprehension protocol

Assessing medication-specific beliefs. Text comprehension depends on the reader's prior knowledge of the text topic as

well as the text itself (Reed, 2013: 266–292). In fact, a number of studies have demonstrated that clear and well-written science passages can fail to modify a reader's misconceptions about the topic—even when the text refutes those misconceptions (see Tippett, 2010). Therefore, it is imperative that a label comprehension study begin with an assessment of the preconceptions that consumers have about the drug in question. This assessment should not be limited to the primary communication objectives of the label. With respect to an emergency contraceptive pill, for example, some misconceptions about the appropriate use of the product may arise from naïve beliefs about the nature of the reproductive process. Morgan et al. (2002) use a "mental models" approach to identify procedures for eliciting valid and invalid preconceptions that consumers may bring to their reading of health-protective communications. Rating-scale assessments of these beliefs offer a quantitative measure that can serve as a pretest measure or a covariate measure.

Assessing verbal ability. Because medical information seems to pose a challenge to many consumers, label comprehension studies often include a measure of medical literacy. One popular measure is the Rapid Estimate of Adult Literacy in Medicine (REALM: Davis et al., 1993). The REALM assesses health literacy by asking patients to pronounce 66 common medical terms. Rather than attempting to assess general medical literacy, however, it may be more useful to assess the specific knowledge a consumer has about the drug under consideration. The information obtained from a pretest such as we describe above serves that purpose. Although a drug-specific pretest assesses the "health" component of health literacy, it does not assess the "literacy" component. Wide-range vocabulary measures offer a more useful index of general verbal ability. The six-minute, 24-item *Extended Range Vocabulary Test* (ERVT: Ekstrom et al., 1976) is suitable for Grades 7–16, covering a wide range of ability levels.

Label attentiveness. With few exceptions (e.g. Bix et al., 2009), label comprehension studies do not verify that respondents have attended carefully to the text of the label. Instead, respondents read the label as though deciding whether to purchase the drug or how to use the drug. In the absence of any external check on a respondent's attentiveness to the label text, there are likely to be large differences among respondents in what they read and what they do not read. The open-label task itself compounds the problem: a respondent is free to scan the label for an answer if recall fails. In order to control statistically for differences in reading styles and motivation, a measure of label attentiveness is necessary. Ryan (2011) used the tactic of inserting obvious typographical errors throughout a simulated analgesic label, asking participants to circle all errors they found.

Label-scanning. Brass and Weintraub (2003) characterize the open-label task as "one designed to demonstrate the ability of consumers to extract relevant information from the proposed label" (p. 407). Ease of scanning can be assessed by timing how long it takes a respondent to locate target information on a label, but that simple measure fails to take into account the possibility that a respondent may often be recalling the targeted information rather than finding it on the label. A better measure involves giving a respondent a randomized list of target words or phrases to be located and marked on the label. If these targets are drawn from throughout the label text, a count of the number of target phrases found in a fixed amount of time provides a more valid measure of scanning-ease.

Delayed free recall. Because some time may elapse between a respondent's first reading of a drug label and the subsequent use of the drug, it is important to assess the amount and accuracy of the label information that a consumer can retrieve from long-term memory after an initial reading. A consumer may not re-read the label when the time comes to use the drug, relying instead on his or her recollection of what the label had said. Although specific circumstances might dictate another measure, a free-recall task following a distraction task ("recall whatever you can from the label as it comes to mind") has four virtues. First, a several-minute distraction task prior to the recall effort ensures that a respondent is retrieving information from long-term memory rather than from short-term memory. Second, a free-recall task provides a measure of recall that does not depend on question comprehension in the way that multiple-choice or true–false questions do. Third, what consumers recall and do not recall provides useful information about their attention strategies. Finally, adjacent items in a free-recall protocol will frequently belong to whatever subjective category an individual has used to organize label information in long-term memory (Friendly, 1977). Such "category clustering" provides invaluable information about the subjective organization of drug-label information.

Perceived label characteristics. Label comprehension studies often rely on consumer perceptions in evaluating drug product labels (e.g. preferences for different label formats). Self-reports and subjective ratings are no substitute for behavioral assessment (e.g. label-scanning speed or free recall of drug facts). Nonetheless, a drug product may fail in its mission if it does not have a positive impact on key consumer perceptions. For example, consumers are likely to read a label with greater care if they perceive it to be well organized and easy to read (FDA, 1999: 13255). Standardized rating scales provide an efficient way to collect information about consumer perceptions that are likely to increase their willingness to attend to label directives.

Re-assessing medication-specific beliefs. A posttest using rating-scale assessments of consumers' preconceptions about a medication provides a useful index of the effectiveness of a label in confirming and correcting drug preconceptions (Ryan, 2011). One aspect of "knowledge updating" is the degree to which readers use label information to strengthen pretest beliefs that accord with that information. A second aspect of knowledge updating is the degree to which readers use label information to weaken pretest beliefs that contradict that information. The two measures reflect different ways in which readers modify their beliefs to accord with the authoritative text of the label.

A demonstration study

In order to provide a detailed illustration of our label comprehension procedure, we examined the degree to which the DFL (FDA, 1999) for aspirin achieves its design objectives. We discuss below the four distinguishable objectives identified at different points in the FDA's 1999 *Final Rule*.

Objective 1: Enhancing the appeal of nonprescription drug labels. The DFL was designed to increase reading motivation and reading confidence by reducing the cognitive demands of reading drug facts information. The simplified wording of drug facts and the use of a highly legible visual format were intended to reduce those cognitive demands (FDA, 1999: 13255).

Objective 2: Increasing the ease of finding information on a nonprescription drug label. The organization of the DFL was designed to help consumers quickly locate and read important drug information in order to permit fast and effective product comparisons. An ordered set of standard headings (Active Ingredients, Purposes, Uses, Warnings, Directions, Other Information, Inactive Information, and Questions) is used on the DFL to organize label information to facilitate that scanning process (FDA, 1999: 13254).

Objective 3: Improving the comprehension and recall of drug facts information. The layout, format, and headings of the DFL were designed to improve consumers' ability to read, understand, and recall label information. The DFL format was thought to offer "a more structured, organized, and compact presentation" that would reduce memory demands in promoting a level of comprehension and recall sufficient for the safe and effective use of a nonprescription drug (FDA, 1999: 13254–13255).

Objective 4: Providing a model for the organization of drug facts information. The nature and ordering of headings in the DFL were designed to develop in consumers a decision-making process for selecting and using nonprescription drugs. The standardized ordering of headings in the DFL was meant to help consumers learn and use a decision-making schema for organizing drug facts information (FDA, 1999: 13258).

Label readability and cohesion

The DFL was designed to achieve its objectives through the organization, layout, and simplification of critical drug information. An implicit assumption is that a highly readable prose version of those facts would not serve the purpose as well. We could find no reports of the relative effectiveness of prose versions of drug-label information. Therefore, we seek in this study to use our modified label comprehension procedure to test the hypothesis the DFL is more effective than highly readable prose in achieving the FDA's four design objectives.

We defined a highly readable prose version of drug facts information as one that is both readable and cohesive. *Readability* as it is conventionally measured assesses word difficulty and sentence complexity by computing, respectively, the average number of word strings in a sentence and the average length of word strings in a text. Different readability formulae combine these two values in different ways to obtain an estimate of reading level (Benjamin, 2012). The FDA (2010) recommends that a DFL be written to a readability standard of "no higher than an 8th grade reading level" (p. 5). The Flesch–Kincaid Reading Level for our aspirin Drug Facts Label (DFL) facsimile (see Appendix 1), for example, is just at the 7th grade reading level with a value of 7.2. In contrast, *cohesiveness* is a measure of the degree to which the sentences in a text can be readily linked to each other. Cohesiveness can be roughly estimated as the percentage of sentences in a text that include a word-concept from a prior sentence. Britton and Gülgöz (1991) have demonstrated that reading cohesive text results in the recall of more text propositions per unit reading time than text not so interconnected. Cohesive text is also likely to produce in the reader a mental representation of the relationships among key text concepts that more closely resemble the mental representation of the text author than do less cohesive texts (Britton and Gülgöz, 1991: Exp. 2). Liu and Rawl (2012) have shown that high levels of text cohesion in colorectal cancer screening information are associated with decreased reading time and increased comprehension performance. In addition, Smith et al. (2011) found small but nonsignificant effects of text cohesiveness on the comprehension of research and clinical information about diabetes mellitus effects.

In order to test the hypothesis that the DFL is superior to highly readable prose, we first created a prose equivalent of the DFL for aspirin by expressing the drug facts on that label in simple sentence form. For example, "*Active Ingredient (in each tablet)*: Aspirin 325 mg aspirin NSAID" was re-written as "The active ingredient in each regular-strength tablet is aspirin. Aspirin is a nonsteroidal anti-inflammatory drug (NSAID). There is 325 mg of aspirin in each tablet." Bulleted items were expressed as full sentences. For example, "*Uses*: ■ headache ■ muscle pain ■ toothache …" was re-written as "You can use aspirin to relieve headache. You can use aspirin to relieve muscle pain. You can use aspirin to relieve toothache pain …." The headings used on the DFL

for aspirin were used as paragraph headings in our Cohesive-Prose Label. We ensured that the text of this label was highly cohesive by including a word from a prior sentence in each new sentence (see Appendix 2). We also created a scrambled-prose version of the aspirin drug facts (see Appendix 3) as a comparison label that included all of the drug facts in sentence form, but had little cohesion. Morrow et al. (1996) made use of a similar logic in comparing the drug facts recall of those who read an organized-prose label for a fictional drug with the recall of those who read a scrambled-sentences label for the same drug. Our SPL contained the same sentences as the organized-prose label, but arranged them in one of two random orders. No paragraph headings were included in this scrambled-prose equivalent of the drug facts for aspirin.

Experimental design and hypotheses. Each participant read only one of the three facsimile labels we prepared; all label assessment tasks were administered to every participant. Label condition (DFL, CPL, or SPL) is our only between-subjects treatment variable. The subject variable Verbal Ability is used as a covariate in all of our analyses, and adjusted means are reported whenever that covariate was significantly correlated with a dependent measure. We use throughout a mixed-model analysis of covariance with Label Condition serving as a between-subjects variable, Verbal Ability serving as a covariate, and one or more task measures serving as within-subjects variables.

We used the SPL to create a content-control condition in which participants would be able to read the complete set of aspirin drug facts in the absence of any organizational structure. We hypothesized that a reading of either the Drug Facts Label or the Cohesive-Prose Label would allow participants to perform significantly better on all of our design-objective measures than would a reading of the Scrambled-Prose Label. The question of primary interest, however, is whether the layout and format of the DFL are a more effective way to organize drug facts information than is a cohesive-prose presentation of that information. If that layout and format are uniquely suited to the task of label comprehension then a reading of the Drug Facts Label should allow participants to perform significantly better on all of our design-objective measures than would a reading of the Cohesive-Prose Label. We test that hypothesis by examining measures of (a) the rated appeal of a drug label, (b) the ease of finding information on a drug label, (c) the comprehension and recall of drug facts, and (d) the organization of drug facts information in memory.

Methods

Participants

In total, 55 female and 39 male undergraduates between the ages of 18 and 25 participated in hour-long group sessions in order to satisfy a research requirement; all were

native English speakers. Confirming our expectation that aspirin is not commonly used by undergraduates (see Ellen et al., 1998), only eight students reported having used aspirin in the 30 days prior to their participation in the study. In contrast, 66 reported having used ibuprofen; 51, acetaminophen; 22, naproxen. Participants reported taking some care in reading label information for their preferred analgesics (7 = *extreme care: M* = 3.38, *SEM* = 0.39), but they were confident that they could use aspirin safely and effectively (7 = *extremely confident: M* = 4.78, *SEM* = 0.18). The mean ERVT verbal ability score out of a possible 24 was 8.27 (*SEM* = 0.36), with scores ranging from 2 to 18. These scores did not vary significantly as a function of the label condition to which participants had been randomly assigned, $F(2, 91) = 2.22$, $p > .10$.

Label facsimiles

Drug Facts Label. Our DFL was a two-page, letter-sized simulation of an aspirin label, with a cover page replicating the Principle Display Panel for a brand name aspirin. We used the exact wording from the then-current aspirin label. The order of information and headings mirrored that DFL order (i.e. Active Ingredient, Uses, Warnings, etc.). Bar-lines, hairlines, indenting, bulleting, and white-space matched those of the actual label. We increased font size to make the DFL more legible, using 16-point, 14-point, and 12-point Arial font for headings, subheadings, and text. The use of boldface and italics also matched the DFL.

Cohesive-Prose and Scrambled-Prose Labels. Our CPL was a three-page, letter-sized prose version of the DFL made by expanding each telegraphic DFL phrase into a complete sentence. The sentence for any bulleted item included an explicit cohesive link to its heading (e.g. the sentence "You can use aspirin to reduce menstrual pain" was included in its appropriate position in the *Uses* paragraph; each bulleted use followed as a separate sentence). Additional cohesive links were added as necessary to ensure that each sentence repeated a word from the previous sentence or from the governing heading. These explicit lexical links aid the reader by eliminating the need to link adjacent sentences by recalling or re-reading earlier sentences or by making language- or knowledge-based inferences (cf. Britton and Gülgöz, 1991; Lesgold et al., 1979). The FDA-mandated headings and subheadings appeared as indented paragraph headings in bold, italicized font. All headings and text were in 12-point Arial font. We created our SPL by eliminating all paragraph headings from our CPL and then preparing two random orderings of those CPL sentences to counterbalance any order effects. We created an artificial paragraph structure with sequences of indented sentence strings that paralleled the length and ordering of paragraphs in the CPL.

Assessing text cohesion. The DFL, CPL, and SPL had comparable Flesch–Kincaid reading grade levels of 7.4, 7.5, and 7.8, respectively. As would be expected, the word count for the CPL and SPL was higher at 1017 and 993, respectively, than the count for the DFL at 613. We computed cohesion values for the CPL and the SPL using the following logic. The CPL is comprised of 94 sentences (counting paragraph headings as sentences). The total number of CPL sentence pairs is then 93 minus the 17 pairs composed of the last sentence on one paragraph and the first sentence of the next—for a final count of 76 CPL pairs that could be lexically linked. We then divided 67—the number of adjacent CPL sentence pairs that were directly linked with a shared word or phrase—by 76 to obtain a lexical cohesion value of 88 percent. Similarly, the total number of SPL sentence pairs that could be lexically linked is then 78 minus the 14 paragraph-transition pairs—for a final count of 64. We found only 12 sentence pairs in the SPL that were directly linked with a shared word, so the value of our lexical cohesion index was 19 percent.

Calculating a cohesion index for the DFL is problematic because it is not true prose. We made two assumptions in order to make an estimate. Because each drug fact in the DFL had given rise to a separate sentence in the CPL and each DFL heading had given rise to a separate paragraph, we again estimated the possible number of lexical links as 76. We obtained an indirect and very liberal estimate of the number of direct lexical links in the DFL by counting the number of bulleted items and subtracting that value from 76. By this logic, each bullet reflects the absence of a cohesive link and calls for a concept in the bulleted heading to be reinstated by a reader (cf. Kieras, 1978). Given a bullet count of 40, we estimated the number of direct cohesive links in the DFL as 36. Dividing that estimate by the total number of idea pairs, we obtained a lexical cohesion index of 47 percent for the DFL. By our estimate, therefore, the cohesiveness value of the DFL is about half that of the CPL.

Label attentiveness. We needed a measure of label attentiveness to control statistically for individual differences in reading style and motivation in our analyses. Following Ryan (2011), we inoculated our DFL, CPL, and SPL texts with obvious spelling errors (e.g. *pan* rather than *pain* or *doctar* rather than *doctor*). Although the spelling errors are shown only for our Drug Facts Label facsimile in Appendix 1, the same spelling errors were included in the Cohesive-Prose Label and the Scrambled-Prose Label actually used by participants in those two label conditions. The spelling errors in all three facsimile labels were distributed at roughly equal intervals throughout the label texts. No error occurred in any word that was central to the meaning of an aspirin drug fact. The DFL, the CPL, and the SPL texts contained the same 26 errors. We told participants that there were at least 20 spelling errors to be found and circled as they read their label. In

addition, we emphasized that it was important that they pay careful attention to all of the information in the label so that they could perform as well as possible on the recall and comprehension tests that would follow their reading of the label. We verified that our attentiveness task was a nonreactive one by examining the correlation between attentiveness scores and the subsequent recall of drug facts from the label. Any shallow processing induced by the error-detection task should impair recall, resulting in a negative correlation between the two measures; in contrast, any deep processing induced by the error-detection task would facilitate recall, resulting in a positive correlation (cf. Craik and Lockhart, 1972). In fact, the overall correlation between error detection and subsequent free recall is close to 0, Pearson's $r(80)=+.09$, $p>.20$, demonstrating that the task neither impaired nor facilitated drug facts recall.

Outcome measures

Objective 1: Enhancing the appeal of nonprescription drug labels. After completing the label-scanning task, participants rated five subjective characteristics of our labels—the labels were not available for inspection as they made these ratings. Four sets of items were rated with a 7-point scale (1=*very much disagree*; 7=*very much agree*). Three items contributed to a Perceived Readability subscale ("The Drug Facts were very clearly written and extremely easy to understand."), Cronbach's $\alpha=.73$. Four items contributed to a Perceived Cognitive Load subscale ("It was hard to put all of the information together"), $\alpha=.76$. Four items contributed to a Perceived Organization subscale ("The label information was very well organized"), $\alpha=.81$. Three items contributed to a Perceived Authoritativeness subscale ("The label information is authoritative with directions that should be carefully followed"), $\alpha=.83$. Finally, four items contributed to a Perceived Recallability subscale. Here, we asked for perceived efficacy ratings (1=*not at all confident*; 7=*extremely confident*) for the recall of drug facts ("I can list from memory all of the reasons one should not use this drug without first consulting a doctor"), $\alpha=.80$.

Objective 2: Increasing the ease of finding information on a nonprescription drug label. We constructed 23 sentence-completion items (e.g. "Vomiting blood is a sign of _____ bleeding") to probe for information from throughout the text of each label. None of these items involved drug facts from which we had drawn the aspirin claims described below. Two different orderings of these items were prepared with the constraint that each missing target word was on a different page than the previous target word. In order to prevent participants from simply recalling the information when they could, participants had to locate and circle each target word on a clean copy of their aspirin label. They also recorded that word or phrase on the response form. We defined "scanning efficiency" as the number of target words correctly circled in 5 minutes.

Objective 3a: Improving the comprehension of drug facts. We used 30 aspirin drug facts from throughout the DFL to construct aspirin claims that were plausibly true or false. Fifteen of these claims were paraphrases of facts directly stated on the label (e.g. "You should drink a full glass of water each time you take aspirin"). We refer to this set of aspirin claims as Label-Congruent (LC) claims because each such claim is explicitly confirmed in each version of the aspirin drug facts (e.g. "drink a full glass of water with each dose"). The remaining 15 aspirin claims were incorrect paraphrases of drug facts directly stated on the label (e.g. "Aspirin should be kept in the refrigerator in order to maintain its potency"). We refer to this second set of aspirin claims as Label-Discrepant (LD) claims because each such claim is directly contradicted in each version of the aspirin drug facts (e.g. "store at room temperature"). We created two different orderings of the 30 claims, with each block of six claims including three LC and three LD claims. In order to encourage participants to rely on their intuitions in rating the perceived validity of each claim, we ask them to use a 6-point Likert scale with anchor-point labels that emphasized intuitive validity (1=*definitely feels false*; 6=*definitely feels true*).

Objective 3b: Improving the recall of drug facts. As their last task in the study, participants had 5 minutes to write down as many facts as they could recall from the aspirin information they had read. We encouraged them to write down any label-related information as it occurred to them—without trying to recall the drug facts in order and without worrying about whether they were completely correct or not. We asked them to write down enough information about each recalled drug fact to allow us to score the recalled item as true or false. When we coded recall, we used gist scoring to give full credit to any fact correctly recalled from the text. We gave partial credit to any fact that referenced specific information from the label even if it was incorrect, distorted, or vague. This scoring scheme allowed us to use a strict or lenient standard for recall in our analyses. Using a strict standard, we counted only drug fact reports that had received full credit; using a lenient standard, we added to our full-credit count any facts that had received partial credit. In order to determine whether our recall measure was biased by the post-label ratings of aspirin claims or by the label-scanning task, we also coded recall reports for any of a new set of 30 aspirin drug facts that were not included among the aspirin claims or among the targets of the scanning probes.

Objective 4: Providing a model for the organization of drug facts information. One virtue of a free-recall measure of memory is that the order in which information is recalled can reveal how that information is organized in memory (Friendly, 1977). If the DFL imposes a "decision-making" category structure upon consumers' organization of drug facts information, then drug facts presented under the same DFL heading would likely be adjacent to each other in our participants'

free-recall protocols. We tested this hypothesis by classifying every label-specific item reported in each recall protocol using a seven-category schema defined by the headings in the DFL. We included only the major categories of aspirin drug facts in the DFL schema: (a) Uses, (b) Reye's Syndrome, (c) Allergy Alert, (d) Stomach Bleeding, (e) Ask Doctor Before Use, (f) Stop Use If, and (g) Directions. We coded all information recalled by our participants, including any incorrect, distorted, or vague information traceable to specific label phrases. We also classified all label-specific items reported on our recall form using a variation of the three-category "naïve medication schema" Morrow et al. (1996) identified for short prescription drug labels. We modified those schema categories to make them more suitable for classifying the kinds of drug facts found on nonprescription drug labels: (a) Purpose, (b) Warnings, and (c) Directions. This dual-coding system allowed us to examine whether the organization of aspirin drug facts in memory is governed by the authorized medication schema embodied in the DFL or by a pre-existing naïve medication schema. We used the adjusted ratio of clustering (ARC: Roenker et al., 1971) to measure the degree to which the formal DFL schema was more successful than Morrow's naïve medication schema in accounting for the order in which drug facts were recalled.

Procedure

After providing informed consent, participants received a folder from which they removed and replaced forms as directed by the experimenter. Because reading the DFL required about 4 minutes and reading the CPL and SPL about 7 minutes, separate group sessions were conducted for each label condition. Although the additional reading time accorded the two prose labels might have advantaged participants in those conditions, we show in our analyses that reading time per se is not critical: the CPL and SPL groups differ in important ways even though participants spent the same amount of time reading each label.

We asked our participants to rate the 30 aspirin claims before they actually read their facsimile label. We told them to read each claim as carefully as necessary to be certain of its meaning and then to take no more than 2 or 3 seconds to circle a number indicating how true it seemed. We explained that we wanted them to base their ratings on their "gut feelings" rather than relying on anything they might have heard or read about aspirin. We also informed them that their ratings would be more reliable if they made "snap judgments" after they had taken the time to understand each claim. Our assumption in providing these instructions was that any preconceptions they had would be rooted in their enduring intuitions rather than in their conscious and deliberate analysis of the claims. We paced them through the claims at a rate of 10 seconds per claim to allow a careful reading of each claim prior to rating its intuitive validity.

We then asked our participants to read their aspirin label as carefully as possible, circling any inserted spelling errors as they found them. After participants had read their labels, they had 6 minutes to complete the 24-item ERVT. Immediately afterward, we asked them to complete the aspirin claims form as a posttest. We again emphasized the importance of rating their "gut feeling" about the truth of a claim after reading it carefully. We also cautioned participants not to try to recall their pretest ratings or specific label information as they made their judgments. We told them that making spontaneous, non-reflective judgments would improve the likelihood that their truth ratings would accurately reflect their enduring intuitions in future days or weeks. As on the pretest, we gave them 10 seconds to read and rate each claim.

Next, we guided participants through the 5-minute scanning task. Afterward, they evaluated the label they had read by completing the 18-item label-evaluation form. The label was not available to them as they made these ratings. Finally, 3 minutes after completing the scanning task, our participants had 5 minutes to recall as many aspirin facts as they could, recording each fact on a different line on the response form. We strongly encouraged them to write down any idea as it occurred, even if they were not certain that it had appeared on the label. After completing this task, each participant received a debriefing form and had the opportunity to ask questions about the study.

Results

Preliminary analyses showed that error-detection scores were not a significant covariate in any analysis. Therefore, we treat our attentiveness scores as an additional outcome measure. We omitted from all analyses data from 12 participants who did not carry out the label-scanning task as instructed.

Aspirin-specific preconceptions. Verbal ability was not a significant covariate in the analyses of truth ratings for LC and LD aspirin claims, $F(1, 78) = 1.03$. Prior to reading any label information, participants rated LC claims as significantly more true ($M = 4.42$, $SEM = 0.05$) than LD claims ($M = 3.60$, $SEM = 0.05$), $F(1, 78) = 20.13$, $MSE = 0.19$, $p < .001$, $\eta^2 = .21$. The mean rating of 4.42 for the 15 LC claims is well above 3.50 (the mathematical midpoint of our 6-point rating scale), $t(81) = 17.34$, $p < .001$, suggesting that our participants were generally certain that our LC claims were valid. The mean rating of 3.60 for the 15 LD claims is just above that scale midpoint, $t(81) = 2.02$, $p < .05$, suggesting that our participants were generally uncertain about whether our LD claims were valid or invalid.

Label attentiveness. Verbal ability did not influence how many of the 26 inserted spelling errors our participants detected and circled, $F(1, 78) = 2.47$, $p > .10$, but the effect of

label version was significant, $F(2, 78)=4.56$, $MSE=18.45$, $p<.05$, $\eta^2=.09$. The number of inserted errors circled by participants as they read the DFL was 15.04 ($SEM=0.74$); by participants as they read the CPL, 18.26 ($SEM=0.80$); by participants as they read the SPL, 17.17 ($SEM=0.78$). Bonferroni comparisons showed that the label version effect was due to a lower error-detection rate in the DFL condition as compared to the CPL condition, $p<.05$. This result may reflect the difficulty of detecting spelling errors in the DFL's telegraphic text.

Objective 1: Enhancing the appeal of nonprescription drug labels

We used a multivariate analysis of variance to analyze the mean ratings on the five label-evaluation scales. Verbal ability did not affect any of the five scale ratings, Wilk's $\Lambda=0.95$, $F(5, 73)<1.00$. The overall effect of label version on sub-scale scores was significant, Wilk's $\Lambda=0.47$, $F(10, 148)=5.78$, $p<.001$, $\eta^2=.31$. Follow-up univariate tests showed that label version significantly affected ratings of Perceived Organization, $F(2, 77)=29.57$, $MSE=1.18$, $p<.001$, $\eta^2=.43$, Perceived Cognitive Load, $F(2, 77)=5.22$, $MSE=1.25$, $p<.007$, $\eta^2=.12$, and Perceived Authoritativeness, $F(2, 77)=5.51$, $MSE=1.30$, $p<.006$, $\eta^2=.13$. However, label version did not affect ratings of Perceived Readability, $F(2, 77)=2.07$, $MSE=1.42$, $p>.10$, or Perceived Recallability, $F(2, 77)=1.32$, $MSE=1.25$, $p>.20$.

As can be seen in Figure 1, the DFL was rated significantly more organized than the SPL, $D_M=+2.04$, $SE_{Diff}=0.30$, $p<.001$, as was the CPL, $D_M=+1.90$, $SE_{Diff}=0.30$, $p<.001$. Both the DFL and the CPL also had significantly lower cognitive-load ratings than the SPL, $D_M=-0.85$, $SE_{Diff}=0.31$, $p<.01$, and $D_M=-0.86$, $SE_{Diff}=0.31$, $p<.01$, respectively. In addition, the DFL and the CPL were both judged as significantly more authoritative than the SPL, $D_M=+0.81$, $SE_{Diff}=0.31$, $p<.01$, and $D_M=+0.96$, $SE_{Diff}=0.31$, $p<.01$, respectively. Bonferroni-adjusted comparisons at .05 level showed that the DFL and the CPL did not differ in Perceived Organization, Cognitive Load, or Authoritativeness ratings.

Objective 2: Increasing the ease of finding information on a nonprescription drug label

Scanning ease was measured as the number of target drug-facts phrases found and circled on a facsimile label within a fixed amount of time. These scores were adjusted for verbal ability effects because that covariate had a significant effect on scanning performance, $F(1, 78)=11.88$, $MSE=12.64$, $p<.001$, $\eta^2=.13$. Label version significantly affected scanning scores, $F(2, 78)=15.31$, $MSE=12.64$, $p<.001$, $\eta^2=.28$. Of the 23 label targets, those reading the DFL found and circled 10.83 ($SEM=0.67$) within the allotted 5-minute interval; those reading the CPL found and circled 9.97
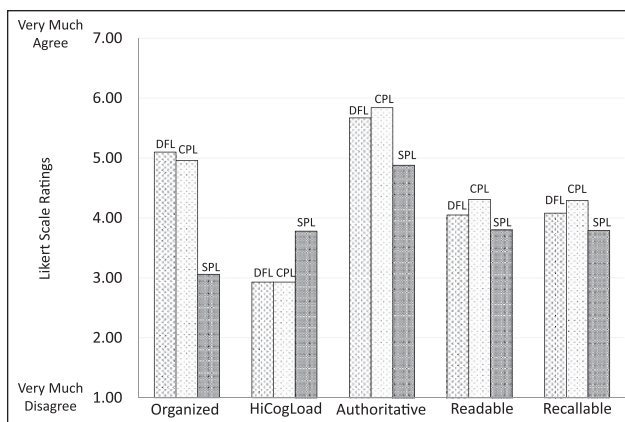

Figure 1. Ratings of label perceptions for Drug Facts Label (DFL), Cohesive-Prose Label (CPL), and Scrambled-Prose Label (SPL).

($SEM=0.70$); those reading the SPL found and circled 5.85 ($SEM=0.68$). Scanning success was significantly greater for those reading the DFL than for those reading the SPL, $D_M=4.98$, $SE_{Diff}=0.96$, $p<.001$. Those reading the CPL were also significantly more successful than those reading the SPL, $D_M=4.12$, $SE_{Diff}=0.98$, $p<.001$. However, a Bonferroni comparison showed that those reading the DFL did not locate significantly more targets than those reading the CPL, $p>.10$.

Objective 3a: Improve the comprehension of drug facts information

Because pretest ratings of LC and LD claims did not differ significantly across label conditions, we subtracted the initial rating of each claim from its final rating to obtain an index of the degree to which reading the label appropriately updated claim ratings. A positive value of this knowledge-updating index for LC claims reflects confirmatory updating; a negative value of the index for LD claims reflects corrective updating.

Participant-by-participant knowledge updating. We entered claim type (LC and LD) as a repeated-measures variable and label version (DFL, CPL, or SPL) as a between-subjects variable. We did not adjust this index for verbal ability because it was not a significant covariate in this analysis, $F(1, 76)>1.00$. The mean updating scores in each label condition are shown in Figure 2. Label version had no main effect on updating scores, $F(2, 79)=1.44$, $MSE=0.13$, nor did it interact with claim type, $F(1, 78)<1.00$. However, the value of the knowledge-updating index was significantly greater for LC claims ($M=+0.62$, $SEM=0.06$) than for LD claims ($M=+0.04$, $SEM=0.06$), $F(1, 79)=58.96$, $MSE=0.23$, $p<.001$, $\eta^2=.43$. Because both updating indices should have changed as a result of reading the aspirin drug
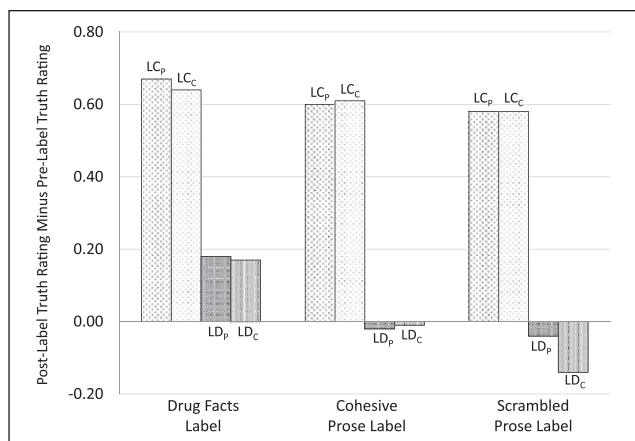
Figure 2. Knowledge updating values for Label-Congruent (LC) and Label-Discrepant (LD) beliefs by participants (P) and by individual claims (C).

Figure 3. Total number of aspirin drug facts recalled from each label and breakdown by number of purpose, warnings, and directions drug facts.

facts, we also tested whether the two indices were significantly different from the no-change value of 0. While the knowledge-updating index for ratings of LC claims was significantly different from 0, $t(80) = 11.24$, $p < .001$, the updating index for ratings of LD claims was not, $t(80) < 1.00$.

Claim-by-claim knowledge updating. We also conducted a claim-by-claim analysis. The mean updating indices for each of the 15 LC claims and the 15 LD claims were treated as two levels of a between-claims factor, with label version treated as three levels of a within-claims factor. There was no main effect of label version, $F(2, 56) = 2.56$, $p < .10$, and no label version × claim type interaction, $F(2, 56) = 1.32$, $MSE = 0.10$, $p > .20$. The effect of claim type on the knowledge-updating index was highly significant, $F(1, 28) = 15.85$, $MSE = 0.17$, $p < .001$, $\eta^2 = .36$. As was the case in the individual-subject analysis, a significant increase in the updating index was found for LC claims ($M = +0.61$, $SEM = 0.11$), but no significant decrease was found for LD claims ($M = +0.01$, $SEM = 0.11$). Here too, the updating index for ratings of LC claims was significantly different from 0, $t(80) = 5.55$, $p < .001$, and the index for ratings of LD claims was not, $t(80) < 1.00$. As can be seen in Figure 2, the knowledge-updating effect sizes are nearly identical for both the individual-subject and the individual-claim analyses. In both analyses, ratings for LD claims were unchanged by any of the three labels.

Objective 3b: Improving the recall of drug facts

In our scoring of participants' free-recall responses, we gave full credit to any response that correctly paraphrased label information (e.g. "you should see a doctor if you have trouble hearing"). We gave partial credit to any response that included specific information from the label—even if it was incorrect (e.g. "don't take more than 3 drinks a day"),
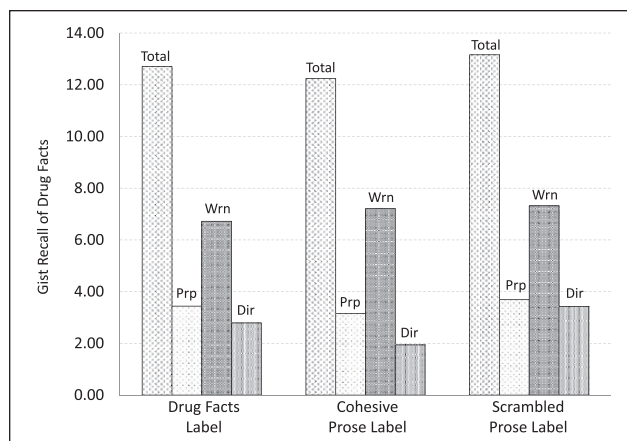
distorted (e.g. "you should not take aspirin if you have Reye's syndrome"), or vague (e.g. "stomach ulcers or bleeding problems"). This scoring scheme allowed us to use a strict or lenient standard for recall in our analyses. Using a strict standard, we counted only drug facts reports that had received full credit. Using a lenient standard, we added to our full-credit count any fact that had received partial credit. The DFL, CPL, and SPL did not differ significantly in their ERVT-adjusted recall scores using either a strict criterion, $F(2, 78) < 1.00$, or a lenient criterion, $F(2, 78) < 1.00$. We report lenient recall scores as Total Recall in Figure 3 because we wanted a sensitive measure of the impact of the label on memory. Our reasoning was that any item of information that bore the imprint of specific wording in the label text demonstrated the impact of the label on memory—even if it was not strictly correct. This lenient standard also ensured that we had a large enough pool of responses to use in our category-clustering analyses.

A further analysis showed that label version did not differentially affect the amount of Purpose, Directions, or Warnings drug-facts recall. These three categories reflect those in the naïve medication schema identified by Morrow et al. (1996). Here too, verbal ability was a significant covariate, $F(1, 78) = 6.19$, $MSE = 4.32$, $p = .015$, $\eta^2 = .07$. Neither label version, $F(2, 78) < 1.00$, nor its interaction with drug-fact category, $F(4, 156) = 1.02$, was significant. The significant main effect of drug-fact category, $F(2, 156) = 13.19$, $MSE = 4.25$, $p < .001$, $\eta^2 = .15$, is due to differences among the categories in the frequency with which they appear among the drug facts on the aspirin label. The relative differences in the ERVT-adjusted means in each category approximate the relative frequencies of each kind of fact in the label itself. Figure 3 displays the ERVT-adjusted recall scores for each category of label content.

We also conducted an analysis of the free-recall data in which we counted only correctly recalled aspirin facts that

had not been among those we included in our set of aspirin claims nor in our set of label-scanning targets. This scoring procedure allowed us to assess whether either task had biased our measure of recall performance. However, we still found that the labels did not differ in their recall effectiveness, $F(2, 78) < 1.00$. Recall of this restricted set of 30 drug facts was marginally greater for the DFL ($M = 4.16$, $SEM = 0.39$) than for the CPL ($M = 3.64$, $SEM = 0.47$) or the SPL ($M = 3.43$, $SEM = 0.45$).

Objective 4: Providing a model for the organization of drug facts information

Using the order in which each participant had recorded aspirin drug facts on the response sheet, we counted the number of pairs of adjacent items that belonged to the same schema category. We did separate counts for the DFL seven-category schema and Morrow's three-category schema. Roenker et al.'s (1971) ARC index computes the ratio of same-heading pairs observed to the total number of possible pairs; it also corrects for the number of same-heading pairs that would occur by chance. In addition, it includes a k value that corrects for individual differences in the number of categories from which two of more items are reported. For coding using Morrow et al.'s naïve medication schema, our participants' k values were 2 or 3; for coding using the authorized DFL schema, their k values were 2, 3, or 4. Computed ARC values can range from -1.00 to $+1.00$, but negative scores are not interpretable (Roenker et al., 1971). We followed the common practice of replacing all negative ARC values with a zero to indicate that a given recall protocol involved only a chance level of category clustering. An ARC score of .30 represents a moderate degree of clustering, but the best use of the index is to determine which of two or more classification schemes is associated with the highest ARC score. For the DFL, CPL, and SPL aspirin texts, 26, 19, and 23 participants, respectively, recalled enough aspirin information to permit the computation of an ARC score for both schemas and the use of a repeated-measures analysis.

We analyzed the two ARC scores for each participant as a within-subjects medication-schema factor and label version as a between-subjects factor, including verbal ability as a covariate. Verbal ability significantly influenced ARC scores, $F(1, 64) = 3.90$, $MSE = 0.06$, $p = .053$, $\eta^2 = .06$, but label version did not, $F(2, 64) < 1.00$. Medication schema had a significant impact on ARC scores, $F(1, 64) = 8.10$, $MSE = 0.03$, $p = .006$, $\eta^2 = .11$, but the medication schema × label version interaction did not, $F(2, 64) < 1.00$. Overall, the degree of category clustering was almost twice as high for Morrow et al.'s naïve medication schema ($M = 0.30$, $SEM = 0.03$) as for the DFL's authorized medication schema ($M = 0.17$, $SEM = 0.02$). Bonferroni comparisons showed the ARC values for Morrow et al.'s naïve schema to be significantly greater than those for the
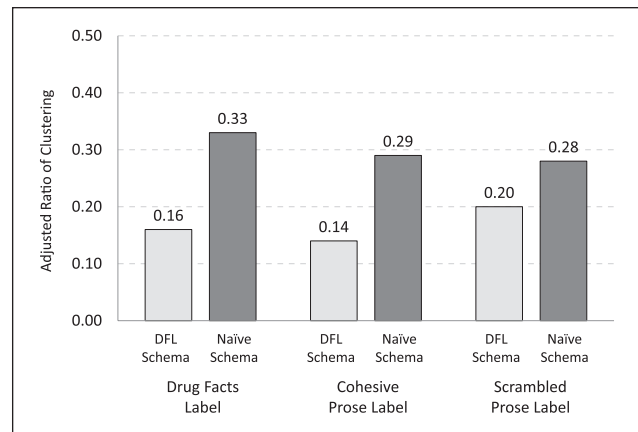

Figure 4. Degree to which drug facts are clustered in recall by drug-facts-label categories and and by Morrow's naïve-medication-schema categories.

authorized schema for each version of the aspirin drug facts, ps < .05 (see Figure 4 for mean ARC values for each label version and each category scheme). However, it is also the case the ARC values for the authorized medication schema are significantly greater than zero for the DFL, the CPL, and the SPL label texts, $t(24) = 5.33$, $p < .001$, $t(17) = 3.50$, $p < .005$, and $t(21) = 5.00$, $p < .001$, respectively. The pattern of results suggests that both medication schemas play a significant role in organizing aspirin information in memory, but the naïve medication schema is the more influential.

Discussion

Summary of findings

The Drug Facts Label performs and fails to perform much as its Cohesive-Prose counterpart on every measure we used. Participants rated aspirin drug facts as significantly more organized, less cognitively demanding, and more authoritative when presented in a Drug Facts Label or in a Cohesive-Prose Label than in a Scrambled-Prose Label. Perceptions of readability and recallability did not differ across the three label conditions. Label-scanning performance was significantly and comparably better for both the DFL and the CPL as compared to the SPL. Unexpectedly, the free recall of drug facts was no greater for the DFL than for our CPL or our SPL. Morrow et al.'s (1996) categories of Purpose, Warnings, and Directions accounted for the organization of drug facts in recall to a significantly greater extent than did the seven major headings of the DFL in all three label conditions, but this effect did not vary as a function of label condition. Most importantly, all three labels were equally effective in promoting the confirmatory updating of LC claims. In contrast, all three labels were equally ineffective in promoting the corrective updating of LD claims.

Interpretation

We interpret our findings by evaluating the degree to which the DFL meets its design objectives (FDA, 1999).

Objective 1: Enhancing the appeal of nonprescription drug labels. Participants reading the DFL for aspirin did perceive the reading process to be less demanding cognitively than reading the SPL. They also perceived the text to be more organized and the information to be more authoritative than did those reading the SPL. However, the CPL was equally superior to the SPL on those three dimensions. In contrast, participants did not judge the aspirin information on the DFL or the CPL to be significantly more understandable or more memorable than that on the SPL. We believe that this latter finding reflects the challenge participants face in monitoring comprehension or predicting recall. Perceptions of comprehension were not correlated with either confirmatory updating, Pearson's $r(80)=-.15$, or corrective updating, $r(80)=+.00$, respectively, nor were perceptions of recallability correlated with lenient gist recall, $r(80)=+.18$. In summary, the DFL meets the objective of motivating label-reading efforts by reducing Perceived Cognitive Load, increasing Perceived Organization, and increasing Perceived Authoritativeness. However, it is no more effective in doing so than the CPL (see Figure 1).

Objective 2: Increasing the ease of finding information on a nonprescription drug label. Those scanning the DFL for target words or phrases found significantly more in a 5-minute period than did those scanning the SPL. However, those scanning the CPL found as many targets as those scanning the DFL. Kools et al. (2008) varied the number of headings included in three different versions of a five-page health education text and asked their undergraduate participants to search for 10 information targets in the text. They found that all versions of the text that included headings were searched more quickly than a control text version that had no headings. Our CPL used as paragraph headings the same headings as the DFL, and our SPL had no headings. Although the DFL does achieve the objective of helping readers quickly locate information, the scanning advantage of the DFL over the SPL may arise simply from the use of headings rather than from any other feature of the label.

Objective 3: Improving the comprehension and recall of drug facts information. Our comprehension measure focuses on the degree to which reading a set of drug facts modifies one's preconceptions about aspirin—increasing truth values for facts confirmed by a label and decreasing values for facts refuted by a label. Neither the DFL nor the CPL produced significantly greater confirmational change or refutational change than did the SPL (see Figure 2). We used recall as a surrogate measure of a reader's ability to apply label information because drug facts that are not recalled

cannot influence the safe and effective use of a nonprescription drug. Neither the DFL nor the CPL produced significantly more recall on any of our four indices than did the SPL (see Figure 3). These findings indicate that the DFL may not fully achieve its comprehension and application objectives.

Objective 4: Providing a model for the organization of drug facts information. Our analysis of category clustering indicates that neither the DFL nor the CPL has any greater effect on the mental organization of drug facts than does the SPL. Although DFL categories significantly influenced recall clustering for all three labels, a simpler naïve schema using only the categories of *Purpose, Warnings*, and *Directions* (Morrow et al., 1996) was significantly more effective in accounting for the order in which drug facts were recalled (see Figure 4). These findings strongly suggest that the DFL does not achieve its objective of influencing the way in which consumers organize drug facts information.

Limitations

Our data indicate that the DFL was no more effective than its cohesive-prose equivalent in achieving any of its design objectives. Our findings are limited in the following ways.

Cohesiveness manipulation. Although we were able to establish a high degree of lexical cohesiveness in our cohesive-prose version of the drug facts, free recall was no greater than in the scrambled-prose version. We believe the failure of our cohesiveness manipulation is due to the specialized nature of the drug-label discourse itself. Text comprehension research has shown that causal connections are particularly effective in improving the comprehension and recall of narrative texts (cf. Trabasso and Van den Broek, 1985). Because our CPL, like the SPL and the DFL, is simply a categorized listing of aspirin drug facts, the cohesive links we created established lexical links between adjacent sentence pairs rather than the causal links that Britton and Gülgöz created for their narrative text. The recall of drug facts information in both the DFL and our CPL might be improved by inserting causal links that explicitly organize the drug facts as steps in a decision-making process.

Choice of nonprescription drug. Our findings are limited to the DFL for aspirin. Because aspirin is not commonly used by undergraduates (Ellen et al., 1998), they are unlikely to have read the label and will be biased largely by their preconceptions about the drug rather than by their actual experience in using the drug. Because the aspirin drug facts are FDA-approved and appear on all product labels for aspirin, our facsimile label is authoritative in a way that a fictional drug label (cf. Morrow et al., 1996) would not be. The fact that our young adult college students are not likely to be familiar with aspirin as an analgesic nor with the DFL for

aspirin does not threaten the internal validity of our study, but it does mean that the effects we report here may be qualitatively different for other more commonly used nonprescription analgesics or for other classes of nonprescription drugs (e.g. antacids, antihistamines, and anti-diarrheals).

Nature of the label-processing task. Our label-reading task is not designed to simulate the label inspection process that occurs when a consumer is comparing one product with another in the same class at the point-of-purchase or when a consumer subsequently self-medicates with that drug. In both situations, the reading process is likely to be less systematic and thoughtful than the task we posed our participants. Prior to reading their assigned label, our participants were asked to rate the perceived truth of 30 different claims about the information to be found on an aspirin drug label. They were then told that they would have a chance to read the actual drug facts for aspirin before rating these claims a second time and recalling what they could from the label. The label-reading instructions they were given encouraged participants to read all of the label information very carefully. As a check on the care they took in inspecting the label information, we asked them to circle any of the typographical errors we had inserted into the drug facts text. In summary, we have every reason to believe that our participants read the aspirin drug facts much more closely than the average consumer would in making his or her drug purchase or in referring to the label for guidance in using the drug for symptomatic relief. Our intention was to ensure that participants performed as well as possible on our dependent measures. For that reason, we believe that under everyday reading conditions, consumers would not perform nearly as well as our participants on the tasks we posed them. Future studies are needed to determine the generality of our findings across a range of more ecologically valid conditions.

Participant sample. We must also limit our conclusions about label comprehension to young adult college students. Different consumer populations might vary in their preconceptions about aspirin, their familiarity with the DFL, and their general knowledge of nonprescription drugs. In addition, our procedure focuses on a knowledge-updating measure of comprehension. This measure is a critical one for assessing label effectiveness, but there are many other ways to define and measure comprehension (White, 2011). Clearly, conceptual replications of our findings using other dependent measures should be conducted.

Carry-over effects. Given the number of tasks in our procedure, it is possible that carry-over effects bias some of our measures. The drug-facts recall task, for example, was preceded not only by the label-reading task but also by the intervening vocabulary test, the post-label aspirin belief ratings, the label-scanning task, and label-evaluation ratings. The vocabulary test does not share any content with the DFL, and the label-evaluation ratings were based on participants' recollection of the label, rather than on a re-inspection of the label. However, the aspirin belief statements may have reminded participants of specific drug facts on the label, and the label-scanning task necessarily involved a selective inspection of the label. We corrected for potential contamination effects by scoring recall protocols only for drug facts that were not included in the aspirin beliefs scale and not included among the label-scanning targets. Scores on this ancillary measure showed the same effect as the other measures of recall. In addition, any contamination effects on drug facts recall would have affected each label condition in the same way. It is also possible that pre-label belief ratings influenced the reading of the DFLs and the post-label belief ratings. Once again, any contamination effects would have been constant across the three label conditions.

Implications

Although further refinements of our procedure may be necessary, two important insights about label comprehension emerge in this study. First, the impact of consumers' domain-specific drug preconceptions and their domain-general medication schemas on the label comprehension process deserves much more examination. Second, nonprescription drug labels may require major modifications in order to ensure that they correct consumers' drug preconceptions and to facilitate the development of valid medication schemas.

Prior-knowledge effects. In this study, we define label comprehension as the degree to which consumers can use a nonprescription drug label to modify their knowledge of a drug so that they can use that drug safely and effectively. This definition assumes that consumers have their own preconceptions—however vague and ill informed—about how they can make best use of a given drug to treat the symptoms that concern them. Our definition also assumes that label comprehension is a mental representation that arises from the interaction of readers' preconceptions about a drug with the text of the drug label (cf. Jungermann et al., 1988; Morris and Aikin, 2001). We use changes in truth ratings for a set of aspirin drug-facts claims to measure the degree to which a label is effective in modifying drug-specific preconceptions about aspirin. Existing measures of health literacy (Davis et al., 1993; Parker et al., 1995) do not serve this purpose. Because such measures seek to assess a consumer's general health and medical literacy, they provide no specific information about consumers' knowledge of a given drug. Measures of verbal ability or text readability are useful for determining how challenging a given consumer will find a given drug label, but they too provide no information about consumers' knowledge of a given drug. Measures of the recall of label information provide an index of what drug facts were read and encoded, but in and of themselves, they offer no information about how consumers understand the drug facts they recall.

Furthermore, our category-clustering analyses indicate that participants appeared to be using Morrow et al.'s (1996) naïve medication schema—Purpose, Warnings, and Directions—rather than the schema dictated by the major headings of the DFL—Uses, Reye's Syndrome, Allergy Alert, Stomach Bleeding, Ask Doctor Before Use, Stop Use If, and Directions. Morris and Aikin (2001: 515) argue that a medication schema for nonprescription analgesics, for example, may serve as a template to guide a consumer's reading of the DFL for a new analgesic. However, the categories of "Ask Doctor Before Use …" and "Stop Use If …" call for very different consumer actions, and those who classify them only under the broad category of "Warnings" may fail to make distinctions that are critical for the safe use of aspirin. Further research on the role played by naïve medication schemas in the processing and application of drug facts information is clearly necessary.

Label redesign. Our findings also have implications for the design and development of drug warning labels. Although the FDA promulgates Primary Communication Objectives for each nonprescription drug, it is notable in the present context that the FDA sets forth no Primary *Refutation* Objectives. Morgan et al. (2002) have identified detailed procedures for eliciting misconceptions that individuals may have about the safe and effective use of a nonprescription drug. Conceptually based misconceptions are not easily changed, but refutation texts have been shown to be effective in producing those changes (Kendeou, Walsh, Smith et al., 2014; Tippett, 2010; Van Den Broek and Kendeou, 2008). In general, refutation texts identify a misconception, label it as incorrect, and explain why it is incorrect. Modifying the DFL to make it serve as a refutation text for significant misconceptions about nonprescription drugs would require a dramatic change in the product labels for those drugs. However, a label that fails to correct significant misconceptions about the safe and effective use of a drug most certainly should be viewed as a defective label. The need for refutational nonprescription drug labels calls for a fundamental reconsideration of the primary communication objectives of a drug label: those objectives should include *ideas to be dispelled* as well as *ideas to be instilled*.

In addition to developing drug labels that are more effective in refuting misconceptions, it is also important to modify the text structure of those labels so that they more explicitly reflect the "decision-making process" originally envisioned by the FDA (1999: 13258–13259). Cook and Mayer (1988) report that only 22 percent of their undergraduates could pick out and label a set of four short sequence-of-events passages from among sets of four other kinds of text passages. Their study indicates that young adults will not likely be able to discern the decision-making sequence implicit in a Drug Facts Label. However, if the headings of the DFL are explicitly identified as steps in a process, and the causal links among those steps are clearly stated (see Kendou, Smith and O'Brien, 2013), consumers may be better able to build a mental model of how to use nonprescription drugs safely and effectively (cf. Bostrom et al., 1994; Britton and Gülgöz, 1991).

References

Benjamin RG (2012) Reconstructing readability: Recent developments and recommendation in the analysis of text difficulty. *Educational Psychology Review* 24(1): 63–88.

Bix L, Bello NM, Auras R, et al. (2009) Examining the conspicuousness and prominence of two required warning on OTC pain relievers. *Proceedings of the National Academy of Sciences* 106(16): 6550–6555.

Bostrom A, Atman CJ, Fischhoff B, et al. (1994) Evaluating risk communications: Completing and correcting mental models of hazardous processes, part II. *Risk Analysis* 14(5): 789–798.

Brass EP and Weintraub M (2003) Label development and the label comprehension study for over-the-counter drugs. *Clinical Pharmacology and Therapeutics*. 64(5): 406–412.

Britton BK and Gülgöz S (1991) Using Kintsch's computational model to improve instructional text: Effects of repairing inference calls on recall and cognitive structures. *Journal of Educational Psychology* 83(3): 329–345.

Cook LK and Mayer RE (1988) Teaching readers about the structure of scientific text. *Journal of Educational Psychology* 80(4): 448–456.

Craik FIM and Lockhart RS (1972) Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior* 11(6): 671–683.

Davis TC, Long SW, Jackson RH, et al. (1993) Rapid Estimate of Adult Literacy in Medicine (REALM): A shortened screening instrument. *Family Medicine* 25(6): 391–395.

Ekstrom RB, French JW, Harman HH, et al. (1976) *Manual for Kit of Factor-Referenced Cognitive Tests*. Princeton, NJ: The Educational Testing Service.

Ellen PS, Bone PF and Stuart EW (1998) How well do young people follow the label? *Journal of Public Policy & Marketing* 17(1): 70–85.

Food and Drug Administration (FDA) (1999) Over-the-counter human drugs: Labeling requirements. *Federal Register* 62(51): 13254–13303.

Food and Drug Administration (FDA) (2010) *Guidance for Industry: Label Comprehension Studies for Nonprescription Drug Products*. Rockville, MD: US Department of Health and Human Services, Center for Drug Evaluation and Research.

Friendly ML (1977) In search of the M-Gram: The structure of organization in free recall. *Cognitive Psychology* 9(2): 188–249.

Jungermann H, Schürz H and Thüring M (1988) Mental models in risk assessment: Informing people about drugs. *Risk Analysis* 8(1): 147–155.

Kendeou P, Smith ER and O'Brien EJ (2013) Updating during reading comprehension: Why causality matters. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 39(3): 854-865.

Kendeou P, Walsh EK, Smith ER, et al. (2014) Knowledge revision processes in refutation text. *Discourse Processes* 51(5–6): 374–397.

Kieras DE (1978) Good and bad structure in simple paragraphs: Effects on apparent theme, reading time, and recall. *Journal of Verbal Learning and Verbal Behavior* 17(1): 13–28.

Kools M, Ruiter RAC, Van De Wiel MWJ et al. (2008)The effects of headings in information mapping on search speed and evaluation of a brief health education text. *Journal of Information Science* 34(6): 833–844.

Lesgold AM, Roth SF and Curtis ME (1979) Foregrounding effects in discourse comprehension. *Journal of Verbal Learning and Verbal Behavior* 18(3): 291–308.

Liu CJ and Rawl SM (2012) Effects of text cohesion on comprehension and retention of colorectal cancer screening information: A preliminary study. *Journal of Health Communication* 17: 222–240.

Morgan MG, Fischhoff B, Bostrom A, et al. (2002) *Risk Communication: A Mental Models Approach*. Cambridge: Cambridge University Press.

Morris LA and Aikin KJ (2001) The "pharmacokinetics" of patient communications. *Drug Information Journal* 35(2): 509–527.

Morrow DG, Leirer VO, Andrassy JM, et al. (1996) Age differences in schemes for taking medication. *Human Factors* 38(4): 556–573.

Parker RM, Baker DW, Williams MV, et al. (1995) The Test of Functional Health Literacy in Adults (TOFHLA): A new instrument for measuring patients' literacy skills. *Journal of General Internal Medicine* 10(10): 537–541.

Raymond EG, Dalebout SM and Camp SI (2002) Comprehension of a prototype over-the-counter label for an emergency contraceptive pill product. *Obstetrics & Gynecology* 100(2): 342–349.

Reed SK (2013) *Cognitive Psychology: Theory and Applications* (9th edn). Belmont, CA: Wadsworth.

Roenker DL, Thompson CP and Brown SC (1971) Comparison of measures for the estimation of clustering in free recall. *Psychological Bulletin* 76(1): 45–48.

Ryan MP (2011) Modifying preconceptions about the safe and effective use of aspirin: Product label comprehension as knowledge updating. *Applied Psychology: Health and Well-Being* 3(3): 322-349.

Smith CA, Hetzel S, Dalrymple P, et al. (2011) Beyond readability: Investigating coherence of clinical text for consumers. *Journal of Medical Internet Research* 13(4): e104

Soller RW (1998) Evolution of self-care with over-the-counter medications. *Clinical Therapeutics* 20(Suppl. C): C134–C140.

Sutherland K (2010) The quest for the Holy Grail: The fully informative, yet fully understandable, OTC label. *Pro Te: Solutio* 3(3): 13–19.

Tippett CD (2010) Refutation text in science education: A review of two decades of research. *International Journal of Science and Mathematics Education* 8(6): 951–970.

Trabasso T and Van den Broek P (1985) Causal thinking and the representation of narrative events. *Journal of Memory and Language* 24(5): 612–630.

Van den Broek P and Kendeou P (2008) Cognitive processes in comprehension of science texts: The role of co-activation in confronting misconceptions. *Applied Cognitive Psychology* 22(3): 335–351.

White S (2011) The cognitive and linguistic demands of everyday, functional literacy tasks: With application to over-the-counter drug label. *Written Language & Literacy* 14(2): 224–250.

Appendix 1. Drug Facts Label

See new warnings information

G E N U I N E

ASPIRIN

Aspirin (NSAID) Pain Reliever/Fevr Reducer

THE MIRACLE DRUG

Fast, Safe Pain Relief
PLUS the Lifesaving Bennefits of Aspirin
325 mg

24
COATED
TABLETS

Aspirin is not appropriate forr everyone, so be sure to talk to your doctor before you begin an aspirin regimen.

Expires 4/13 209560D 3 12843 10111   1

Do not use this product if printed safty seal bearing aspirin brand name under cap is torn or missing

---

Drug Facts

Active Ingredient	*Purposes*
(in each tablet)	
Aspirin 325 mg (NSAID)*…………………………………..	Pain reliever/
	fever reduser

*nonsteroidal anti-inflammatory drug

Uses temporarily relieves
■ headache ■ muscle pan ■toothache
■ menstrual pain ■ pain and fever of colds
■ minorr pain of arthritis

Warnings

Reye's syndrome: Children and teenagers who have ore are recovering from chicken pox or flu-like symptoms should not use this product. When using this product, if changes in behavior with nausa and vomiting occur, consult a doctor because these symptoms could be an early sign of Reye's syndrome, a rare but serious illness.

Allergy alert: Aspirin may cause a severe allergic reaction which may nclude:

Drug Facts (Continued)

- hives ■ facial swelling
- asthma (wheezing) ■ shock

Stomach bleeding warning: This product contains an NSAID, which may cause severe stomach bleeeding. The chance is higher if you

- are age 60 or older
- have had stomach ulcers or bleeding problems
- take a blood thinning (anticoagulant) or steroid drug
- take other drugs contaning prescription or nonprescription NSAIDs (aspirin, ibuprofen, naproxen, or others)
- have 3 or more alcoholic drinks every day while using this product
- take moore or for a longer period of time than directed

Do not us if you are allergic to aspirin or any other pain reliever/fever reducer

Ask a doctor before use if
- stomach bleeding warning applies to you
- you have a hestory of stomach problems, such as heartburn
- you have high blood pressure, heart disease, liver cirrhosis, or kidney disease
- you are taking a diuretic
- you have asthma

Ask a doctar or pharmacist before use if you are taking a prescription drug for
- gout ■ diabetes ■ arthritis

Stop use and ask a doctor if
- an allergic reaction occurs. Seek medical help rght away.
- you experience any of the following sins of stomach bleeding:
    - feele faint ■ vomit blood
    - have bloody or black stools
    - have stomach pain that does not get better
- pain gets worse or lasts more than 10 days
- redness or swelling is present
- fever lasts more than 3 days
- new symptoms occr
- ringing in the ears or a loss of hearing occurs

If you are pregnant or breast-feeding, ask a helth professional before use. **It is especially important not to use aspirin during thee last 3 months of pregnancy unless definitely directed to do so by a doctor becaus it may cause problems in the unborn child or complications during delivery**.
Keep out of the reach of children. In case of overdose, get medical help or contract a Poison Control Center right away.

Directions
- drink a full glass of water with eache dose
- adults and children 12 years and over: take 1 to 2 tablts every 4 hours or 3 tablets every 6 hours, not to exceed 12 tablets in 24 hours
- children under 12 years; consult a doctor

Other information
- save carton for full directionns and warnings
- store at room temperature

Inactive ingredients carnauba waxs*, corn starch, hypermellose, powdered cellulose, triacetin *may contain this ingredient

Questions? **1-800-331-4536** or www.aspirin.com

Appendix 2. Cohesive-Prose Label

Regular-strength aspirin

General information. Be sure that you read the new warnings information on this label. This product contains genuine regular-strength aspirin. Aspirin (nonsteroidal anti-inflammatory drug (NSAID)) is a pain reliever and fever reducer. Aspirin is called the "Miracle Drug." Aspirin provides fast and safe pain relief. Aspirin also has lifesaving benefits. There are 24 coated tablets in this product. Each coated tablet has 325 mg of aspirin. Aspirin is not appropriate for everyone. You should talk to your doctor before you begin an aspirin regimen. There is a printed safety seal under the cap of this product that says "Aspirin Brand Name." You should not use this product if the safety seal is torn or missing.

Active ingredient. The active ingredient in each regular-strength tablet is aspirin. Aspirin is a fever reducer and pain reliever. Aspirin is an NSAID. There is 325 mg of aspirin in each tablet.

Purposes. You can use aspirin to relieve headaches. You can use aspirin to relieve muscle pan. You can use aspirin to relieve toothache. You can use aspirin to relieve menstrual pain. You can use aspirin to relieve pain and fever of colds. You can use aspirin to relieve the minor pain of arthritis.

Warning about Reye's syndrome. Children and teenagers who have or are recovering from chicken pox should not use this product. Children and teenagers who have or are recovering from flu-like symptoms should not use this product. When using this product, if changes in behavior accompanied by nausea and vomiting occur, consult a doctor. Changes in behavior with nausea and vomiting could be an early symptom of Reye's syndrome. Reye's syndrome is a rare but serious illness.

Warning about allergic reactions. Aspirin may cause a severe allergic reaction. Allergic reactions may include hives. Allergic reactions may include facial swelling. Allergic reactions may include asthma (wheezing). Allergic reactions may include shock.

Warning about stomach bleeding. This product contains an NSAID that may cause severe stomach bleeding. The chance of stomach bleeding is higher if your age is 60 or over. The chance of stomach bleeding is higher if you have had stomach ulcers or bleeding problems. The chance of stomach bleeding is higher if you take a blood thinning (anticoagulant) or steroid drug. The chance of stomach bleeding is higher if you take other drugs containing prescription or nonprescription NSAIDs (aspirin, ibuprofen, naproxen, or others). The chance of stomach bleeding is higher if you have three or more alcoholic drinks everyday while using this product. The chance of stomach bleeding is higher if you take more aspirin than directed for each dose. The change of stomach bleeding is higher if you take aspirin for a longer period of time than directed.

Deciding to use this product. Do not use if you are allergic to aspirin. Do not use if you are allergic to any other pain reliever/fever reducer.

Asking a doctor before using this product. Ask a doctor before use if stomach bleeding warning applies to you. Ask a doctor before use if you have a history of stomach problems, such as heartburn. Ask a doctor before use if you have high blood pressure, heart disease, liver cirrhosis, or kidney disease. Ask a doctor before use if you are taking a diuretic. Ask a doctor before use if you have asthma.

Asking a doctor or pharmacist before use about other drugs you take. Ask a doctar or pharmacist before use if you are taking a prescription drug for gout. Ask a doctor or pharmacist before use if you are taking a prescription drug for diabetes. Ask a doctor or pharmacist before use if you are taking a prescription drug for arthritis.

Stop using this product. Stop use and ask a doctor if an allergic reaction occurs. If an allergic reaction occurs, seek medical help right away.

Stope using this product. Stop use and ask a doctor if you experience any signs of stomach bleeding. If you feel faint, vomit blood, or have bloody or black stools, those are signs of stomach bleeding. It is also a sign of stomach bleeding if you have stomach pain that does not get better.

Stop using this product. Stop use and ask a doctor if pain gets worse or lasts more than 10 days. Stop use and ask a doctor if redness or swelling is present. Stop use and ask a doctor if a fever lasts more than 3 days. Stop use and ask a doctor if new symptoms occur. Stop use and ask a doctor if a ringing in the ears or a loss of hearing occurs.

Pregnant or breast-feeding? If you are pregnant or breast-feeding, ask a health professional before use. It is especially important not to use aspirin during the last 3 months of pregnancy unless you are definitely directed by a doctor to do so. Using aspirin in the last 3 months of pregnancy may cause problems in the unborn child. Using aspirin in the last 3 months of pregnancy may also cause complications during delivery.

Keep this drug out of the reach of children. In case of overdose, get medical help or contract a Poison Control Center right away.

Directions. You should drink a full glass of water with each dose of aspirin. Adults and children 12 years and over can

take 1 or 2 tablets every 4 hours. Adults and children 12 years and over can also take 3 tablets every 6 hours. Adults and children 12 years and over should not take more than 12 tablets in 24 hours. Consult a doctor before giving aspirin to children under 12 years.

Other information. You should save the carton so that you can refer to the full directions and warnings. You should store aspirin at room temperature.

Inactive ingredients. Aspirin tablets may contain the inactive ingredient carnauba wax. Aspirin tablets contain corn starch and hypromellose as inactive ingredients. Aspirin tablets also contain the inactive ingredients of powdered cellulose and tiacetin.

Questions. If you have any questions about aspirin, you can call 1-800-331-4536. You can also get answers to your questions by visiting our website at www.aspirin.com.

Appendix 3. Scrambled-Prose Label

Regular-strength aspirin

The chance of stomach bleeding is higher if you have three or more alcoholic drinks everyday while using this product. Ask a doctor before use if you are taking a diuretic. Allergic reactions may include shock. The chance of stomach bleeding is higher if you have had stomach ulcers or bleeding problems. Ask a doctar or pharmacist before use if you are taking a prescription drug for diabetes. Aspirin tablets contain corn starch and hypromellose as inactive ingredients. Stop use and ask a doctor if new symptoms occur. You can use aspirin to relieve pain and fever of colds. Stop use and ask a doctor if a ringing in the ears or a loss of hearing occurs. You should drink a full glass of water with each dose of aspirin.

You should save the carton so that you can refer to the full directions and warnings. Stop use and ask a doctor if fever lasts more than 3 days. If you are pregnant or breast-feeding, ask a health professional before use.

Adults and children 12 years and over can take 1 or 2 tablets every 4 hours. Do not use if you are allergic to aspirin. You should keep this product out of reach of children. Aspirin tablets may contain the inactive ingredient carnauba wax. Using aspirin in the last 3 months of pregnancy may cause problems in the unborn child. There are 24 coated tablets in this container.

Using aspirin in the last 3 months of pregnancy may also cause complications during delivery. In case of overdose, get medical help or contract a Poison Control Center right away. Ask a doctor before use if you have asthma. The active ingredient in each regular-strength tablet is aspirin. Adults and children 12 years and over can also take 3 tablets every

6 hours. Stop use and ask a doctor if you experience any signs of stomach bleeding. Ask a doctor before use if you have liver cirrhosis or a kidney disease. You should not use this product if the safety seal is torn or missing.

Reye's syndrome is a rare but serious illness. If an allergic reaction occurs, seek medical help right away. Aspirin is an NSAID. The chance of stomach bleeding is higher if your age is 60 or over. You should store aspirin at room temperature.

It is especially important not to use aspirin during the last 3 months of pregnancy unless you are definitely directed by a doctor to do so. If you feel faint, vomit blood, or have bloody or black stools, those are signs of stomach bleeding. Do not use product if you are allergic to any other pain reliever or fever reducer. Children and teenagers who have or are recovering from flu-like symptoms should not use this product. Aspirin provides fast and safe pain relief. Aspirin also has lifesaving benefits. Changes in behavior with nausea and vomiting could be an early symptom of Reye's syndrome. You can also get answers to your questions by visiting our website at www.aspirin.com. You should consult a doctor before giving aspirin to children under 12 years. Stop use and ask a doctor if an allergic reaction occurs. Aspirin tablets contain corn starch and hypromellose as inactive ingredients.

If you feel faint, vomit blood, or have bloody or black stools, those are signs of stomach bleeding. You can use aspirin to relive the minor pain of arthritis.

There is 325 mg of aspirin in each tablet. The chance of stomach bleeding is higher if you take other drugs containing nonprescription NSAIDs (aspirin, ibuprofen, naproxen, or others). You should talk to your doctor before you begin an aspirin regimen. Ask a doctor or pharmacist before use if you are taking a prescription drug for gout. If changes in behavior accompanied by nausea and vomiting occur in a child or teenager while using this product, you should consult a doctor.

It is also a sign of stomach bleeding if you have stomach pain that does not get better. Adults and children 12 years and over should not take more than 12 tablets in 24 hours. Aspirin tablets also contain the inactive ingredients of powdered cellulose and tiacetin. The chance of stomach bleeding is higher if you take aspirin for a longer period of time than directed.

Stop use and ask a doctor if redness or swelling is present. Allergic reactions may include facial swelling. You can use aspirin to relieve muscle pan.

Children recovering from chicken pox should not use this product. Aspirin is called the "Miracle Drug." Stop use and ask a doctor if pain gets worse or lasts more than 10 days. This product contains an NSAID that may cause severe stomach bleeding.

If you have any questions about aspirin, you can call 1-800-331-4536. You can use aspirin to relieve headaches.

An aspirin regimen is not appropriate for everyone. Ask a doctor before use if you have high blood pressure or heart disease. Aspirin is a fever reducer and a pain reliever. There is a printed safety seal under the cap of this product that says "Aspirin HealthCare."

You can use aspirin to relieve menstrual pain. Allergic reactions may include asthma (wheezing). Ask a doctor or pharmacist before use if you are taking a prescription drug for arthritis. Aspirin may cause a severe allergic reaction. Ask a doctor before use if stomach bleeding warning applies to you. The chance of stomach bleeding is higher if you take more aspirin than directed for each dose.

Ask a doctor before use if you have a history of stomach problems, such as heartburn. Allergic reactions may include hives. Be sure that you read the new warnings information on this label.

This product contains genuine regular-strength aspirin. Aspirin (NSAID) is a pain reliever and a fever reducer. The chance of stomach bleeding is higher if you take a blood thinning (anticoagulant) or steroid drug.