EVALUATION OF DRUG LABEL DESIGNS USING EYE TRACKING

Agnieszka Bojko, Catherine Gaddy, Gavin Lew, Amy Quinn
User Centric, Inc.
Oakbrook Terrace, IL

Edmond Israelski
Abbott
Abbott Park, IL

Eye movement measures and conventional performance metrics were used to compare existing drug labels to a new label template that was created based on human factors principles and user research. Twenty pharmacy practitioners were asked to locate a particular drug among others using sets of existing labels and their redesigned counterparts. For most tasks, the new design led to faster responses, either due to a decrease in the number of fixations required to complete a task or a decrease in the mean fixation length. The number and sequence of fixations within a single label and across labels (used as indicators of search efficiency) and fixation duration (used as a measure of information processing efficiency) provided insight into the origins of the noted speed improvements, helping assess which of the multiple design changes introduced in the new template had impact on performance. Application of eye tracking to redesign evaluation is discussed.

INTRODUCTION

We were asked by a major drug manufacturer to redesign prescription drug labels used by pharmacy staff in a way that would help reduce dispensing errors. Based on the literature on medication error and drug label design, internal drug label standards of the manufacturer, observations conducted in pharmacies, as well as interviews and a participatory design session with pharmacists and pharmacy technicians, we created the new label template (Quinn, Bojko, Gaddy, & Israelski, 2005), which is shown in Figure 1.

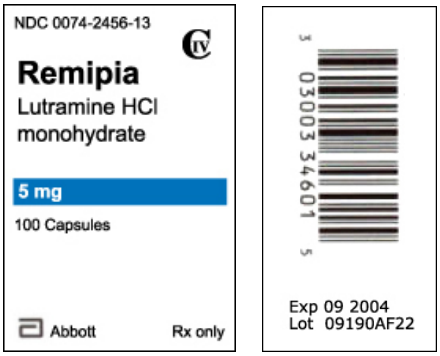

Figure 1. Front and side panels of the new label template. Imaginary brand and generic drug names and modified colors for dosage strength were used to control for pharmacy staff experience. (NOTE: The new label designs described in this study were created solely for research purposes and are not intended to depict any actual labels.)

Because introducing a new label design is costly, the new template had to be objectively evaluated against the existing label designs to determine whether or not it was likely to improve performance of pharmacy staff. We conducted an evaluation which consisted of two components: a usability test with labels on actual bottles, and an eye tracking study using images of labels presented on a computer screen (constraint imposed by the type of eye tracker we used). Both components involved the existing labels and their redesigned alternatives.

While the usability test ensured higher ecological validity for our evaluation, the response time data collected with a stopwatch were subject to reduced precision. The eye tracking test would provide finer time measurement (automatically logged by the computer) and gather eye movement data, which would help us detect and explain potential differences in performance between the existing labels and the new design. In this paper, we focus on the second component of the study which involved eye tracking. The usability test is described in Quinn et al. (2005).

When designing a drug label, there are two main goals related to users' eye movements, which translate into the ease of visual search and information processing. The first goal is to optimize eye movement patterns *across labels*. A drug label is often seen next to other labels when the drugs are arranged on a shelf, and identifying the right label among all others should be as easy as possible. The second design goal involves optimizing eye movement patterns *within a label*, so it is easy to locate the relevant information when a label is examined on its own. Achieving these goals would ultimately lead to increased speed and accuracy when selecting drugs from a shelf.

In the process of designing the new drug label template, we intended to improve the efficiency of visual search and information processing *across labels* by:

- Increasing location consistency of label elements by placing the same type of information in the same positions on different labels (Wickens, Gordon, & Liu, 1998)
- Increasing format consistency of elements by applying the same design treatment (e.g., font type and size, order of sub-elements, letter case) to the equivalent elements across labels (Wickens, Gordon, & Liu, 1998).

Improving the efficiency of search and information processing *within a label* was attempted by:

- Ensuring consistency with user expectations by placing information where users tend to anticipate it and formatting it according to established rules (Parasuraman, 1986)
- Increasing information legibility by replacing all-uppercase text with mixed case (Tinker, 1963)
- Reducing clutter by removing less important information from the front of the label and left-aligning most of the remaining elements (Wickens, Gordon, & Liu, 1998)
- Increasing the discriminability of the important elements through the use of color, different font sizes, and higher contrast between the text and the background (Fisher & Tan, 1989; Parkhurst & Niebur, 2004).

Because the existing labels were not standardized and differed from one another (sometimes quite substantially), the above design changes affected some labels more than others. They also affected each of the label elements to a different extent. The objective of the study was to determine whether our overall design efforts were successful and if so, which of the introduced changes would lead to improvements in performance for pharmacists and pharmacy technicians.

Two types of eye movement measures were collected: quantitative – the number and duration of fixations needed to accomplish the tasks, and qualitative – the spatial sequence of these fixations (i.e., scanpath). Number and sequence of fixations in search tasks are considered to be indicators of layout effectiveness (Kotval & Goldberg, 1998). A large number of fixations and long scanpaths with frequent backtracking (i.e., rapid changes in direction) suggest inefficient search caused by poor layout and unintuitive information organization. Mean fixation duration, on the other hand, measures information clarity and/or density in the stimulus (Halverson & Hornof, 2004). Long fixations indicate difficulties extracting and interpreting information due to ambiguous or unclear information, or high information density. In addition to the eye movement measures, response time and accuracy data were automatically logged.

When using the above metrics to compare two sets of labels, we assumed that, in general, the set consisting of more effective designs would require less time to find information, lead to fewer errors, involve fewer and shorter fixations, and/or produce more direct scanpaths.

Specifically, we predicted an overall decrease in the time to select the correct label for the new designs as compared to the existing designs. Due to the relatively low (1.7%) error rates estimated based on actual pharmacy error data (Flynn, Barker, & Carnhan, 2003), we did not expect to observe many errors, and thus did not anticipate any notable effects of design on accuracy.

With regard to the eye movement measures, we predicted that the new designs would show a reduction in the number of fixations in tasks requiring to find the label elements for which we specifically increased location consistency across labels and/or discriminability. We also expected a decrease in fixation duration for the new designs in (1) tasks involving labels whose existing versions have much higher information density and clutter, and a low level of differentiation between the various elements of the label, and (2) tasks involving label elements that underwent an improvement in format consistency and legibility.

METHOD

Participants

A total of twenty participants (50% female), 10 pharmacists and 10 pharmacy technicians, with normal or corrected-to-normal vision, ranging in ages from 18 to 67 were recruited from various pharmacies in the Chicagoland area. All received monetary compensation for their time. Four participants were excluded from the analyses due to tracking difficulties which caused over 10% of their data to be invalid.

Apparatus

The stimuli were presented on a 17" monitor interfaced with a PC with a 1.79 GHz AMD Athlon XP 2100+ processor. The screen resolution was set to 1024 x 768 pixels. Each participant used a mouse to indicate responses. Eye movements were recorded with a Tobii 1750 binocular remote eye tracker with 50 Hz temporal resolution and a 0.5° spatial resolution.

Stimuli and Procedure

Participants were shown 30 sets of labels on a computer screen. Each set consisted either of three existing labels or three labels created based on the new template. The labels in each set represented drugs that were likely to be next to each other on a pharmacy shelf. Each label was 220 pixels high and 148 pixels wide, which is equivalent to the actual size of the largest label currently available.

Each set of labels was preceded by task instructions that were displayed on the screen until the participant clicked on the Start link. There were five types of tasks, each associated with a different label element:

1. BRAND NAME: *Pick the drug named [brand name].*
2. GENERIC NAME: *Pick the drug named [generic name].*
3. DOSAGE FORMULATION: *Pick the drug named [brand name] ER.*
4. DOSAGE STRENGTH: *Pick the drug with the strength [X mg or X mg/X mg].*
5. EXPIRATION DATE: *Pick the drug that expired on [date].*

Once the stimuli appeared, participants would click on the label containing the characteristic specified in the task. Figure 2 presents a sample sequence of events during a trial. Participants were asked to respond as quickly as possible but without sacrificing accuracy. The order of the tasks and sets of stimuli was random for each participant.

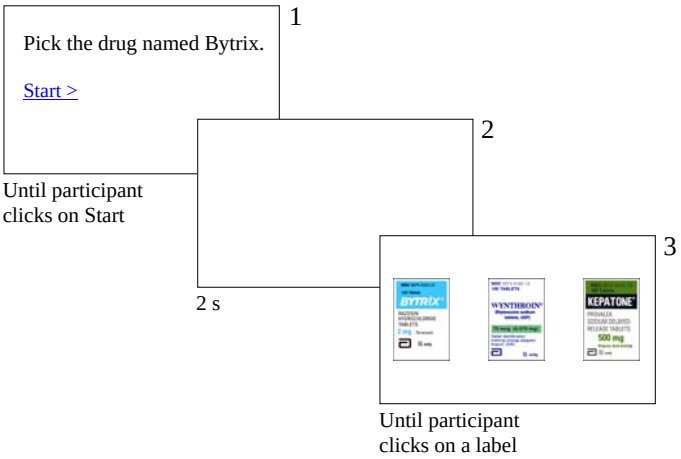

Figure 2. Sequence of events in a trial.

RESULTS AND DISCUSSION

We computed a 2 (design: existing and new) x 5 (task: brand name, generic name, dosage formulation, dosage strength, and expiration date) within-subjects analysis of variance (ANOVA) for response times, fixation count, and mean fixation length. All error trials were excluded from the analysis because only seven errors occurred and it was often impossible to determine whether they were caused by the participant forgetting the task or by design factors. We also excluded the last fixation in each trial from the fixation duration analysis. The last fixation tended to be approximately twice as long as the other fixations due to the fact that it coincided with the response (i.e., mouse click), which required aiming with the cursor. Including this fixation in the analysis would have had an uneven effect on the average fixation duration, artificially increasing it for tasks with fewer fixations.

While there were no interactions between the factors ($p > .05$), both main effects – for design and for task – were found to be significant for all three measures. Overall, the new designs required less time to complete tasks, $F(1, 15) = 20.7$, $p < .001$, fewer fixations, $F(1, 15) = 18.5$, $p < .001$, and shorter fixation durations, $F(1, 15) = 4.6$, $p < .05$, than the existing designs. The reduction in the number of fixations needed to complete the tasks suggests that the new label design decreased the overall overt search demands placed on the pharmacy staff. The reduction in fixation length, on the other hand, indicates lowered information processing demands with the new designs as compared to the existing designs. The main effects for task (response time: $F(4, 15) = 7.6$, $p < .0001$; fixation count: $F(4, 15) = 6.6$, $p < .001$; fixation duration: $F(4, 15) = 5.2$, $p < .005$) suggest that the tasks differed in their difficulty level.

Paired t-tests were conducted to compare participants' performance using the existing labels with their performance using the new labels for each of the five types of tasks. This was done to assess how specific design changes impacted performance. Table 1 summarizes the results of these comparisons.

Brand Name Task

No significant differences were found in response times and eye movement data between the existing labels and the new labels when participants searched for the drug's brand name. In both designs, participants tended to locate the element in question (i.e., brand name) on the label they first fixated almost immediately, without allocating much attention to other areas of the label (Figure 3).

The same was true of the other labels that were fixated during the task. The observed behavior suggests that in all labels, both existing and new, the brand name was easy to find, likely due to the fact that it was a highly salient element compared to the other elements of the label and was placed in the general location where the participants expected it based on their experience.

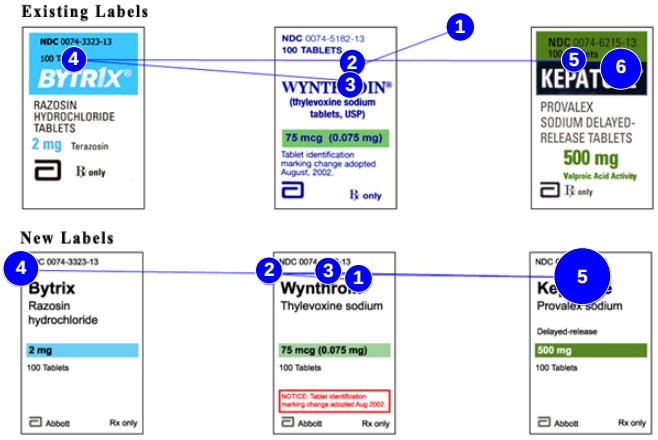

Figure 3. Sample scanpaths used to find a label with the brand name specified in the task (here: *Kepatone*). The circles represent fixations. Dot size is proportional to fixation duration. The numbers indicate the order of fixations.

TASKS	Response Time			Number of Fixations			Fixation Duration		
	Existing Labels	New Labels	*t*(15)	Existing Labels	New Labels	*t*(15)	Existing Labels	New Labels	*t*(15)
Brand Name	1.76 s ± .59	1.71 s ± .49	0.3	5.63 ± 1.63	4.63 ± .69	1.9	260 ms ± 81	275 ms ± 88	-0.5
Generic Name	2.41 s ± .59	1.93 s ± .73	3.9*	5.63 ± 1.63	5.13 ± 1.50	1.2	392 ms ± 87	335 ms ± 80	2.7*
Dosage Formulation	2.18 s ± .53	1.83 s ± .64	2.3*	6.00 ± 1.46	4.75 ± 1.18	3.4*	300 ms ± 85	292 ms ± 103	0.3
Dosage Strength	1.97 s ± .42	1.66 s ± .33	3.0*	4.81 ± 1.60	4.69 ± .62	0.3	357 ms ± 106	285 ms ± 71	2.9*
Expiration Date	2.72 s ± .72	2.22 s ± .62	2.3*	7.13 ± 1.59	5.19 ± 1.97	4.0*	340 ms ± 101	336 ms ± 98	0.1

Table 1. Means and standard deviations ($M \pm SD$) for response time, number of fixations, and fixation duration for different types of tasks performed with existing labels, as compared to the tasks performed with new labels. *Results significant at $p < .05$.

Even though we increased the consistency of the brand name's position and formatting across labels, these modifications were not substantial enough to produce a visible difference in performance. Compared to the other tasks, the brand name task was among the easiest to complete and it is possible that the current performance is already close to optimal (given the design constraints) and cannot be improved any further.

Generic Name Task

A statistically significant difference was observed for the time required to find a label with the correct generic name, with the new designs leading to faster responses than the existing designs. Of the eye movement measures, only fixation duration was significantly lower for the new designs than it was for the existing designs (Figure 4). This decrease was likely caused by the improved consistency in the name formatting (i.e., font type and size). Another possible contribution to the reduction in fixation length involved the increased legibility of the generic name in some labels achieved by replacing all-uppercase text, which makes words more difficult to recognize (Tinker, 1963), with mixed case.

Contrary to our predictions, the number of fixations did not significantly differ between the designs, even though we increased the location consistency of the generic name across labels. This design modification did not impact search efficiency – possibly because the participants, based on their experience, expected the generic name to be positioned right below the brand name which was very easy to locate in both existing and new labels.

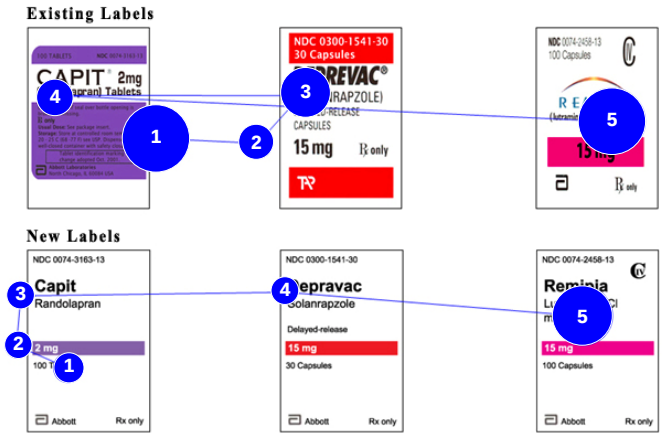

Figure 4. Sample scanpaths used to find a label with the generic name specified in the task (here: *lutramine HCl monohydrate*). Notice the longer fixations on the existing designs as compared to the new designs.

Dosage Formulation Task

In the dosage formulation tasks, the new labels significantly differed from the existing labels in terms of response time and number of fixations required to locate the search target. The new designs produced faster responses and required fewer fixations than the existing designs (Figure 5). These savings were likely caused by the increased luminance contrast between the letters "ER" and their background, which facilitated the quicker allocation of overt visual attention in the new designs (Parkhurst & Niebur, 2004).

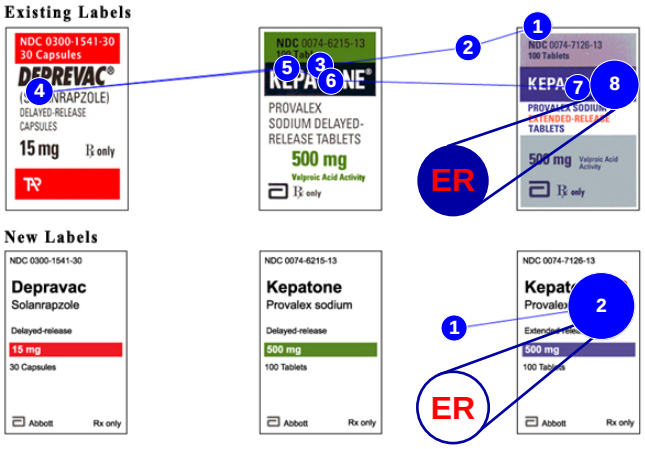

Figure 5. Sample scanpaths used to find a label with the formulation specified in the task (here: *Kepatone ER*). Notice that in both tasks the location of the participant's first fixation is similar. However, when looking at the existing labels, he did not notice the red "ER" on the blue background and scanned the other two labels before returning to the target label. When performing the task with the new labels, the red "ER" on white attracted the participant's attention immediately.

Dosage Strength Task

Significant differences in response time and average fixation duration were observed between the existing labels and the new designs in the dosage strength tasks. The new designs produced faster responses and shorter fixations than the existing designs (Figure 6). One of the design changes that likely resulted in the decrease in fixation length is the enhanced legibility of the dosage strength information achieved by the enlarged font. Removing less important details from the label, and thus decreasing its overall information density, may have also contributed to making strength information easier to process, especially because the existing designs that were used for this task had a particularly high information density.

A new graphic treatment element introduced in the redesigned labels was the colored background against which we placed the dosage strength information. We predicted that this would make the dosage strength easier to isolate from the rest of the information (Yantis & Egeth, 1999), which would decrease the number of fixations needed to locate it. This would only impact search performance for the first fixated label because the location of the dosage strength information was consistent across labels in both design conditions. However, no difference in fixation count between the existing and new designs was noted, likely because the designs were randomized and the participants never knew which (existing or new) to expect. This and the fact that the dosage strength information appeared on a colored background in the new labels but it fell *outside* of the colored areas in the existing labels impeded the top-down attention that modulates bottom-up signals, likely leading to the lack of improvement in search efficiency. Improvements might be visible with a blocked presentation of the two designs, where the participants could rely more heavily on the top-down processes.

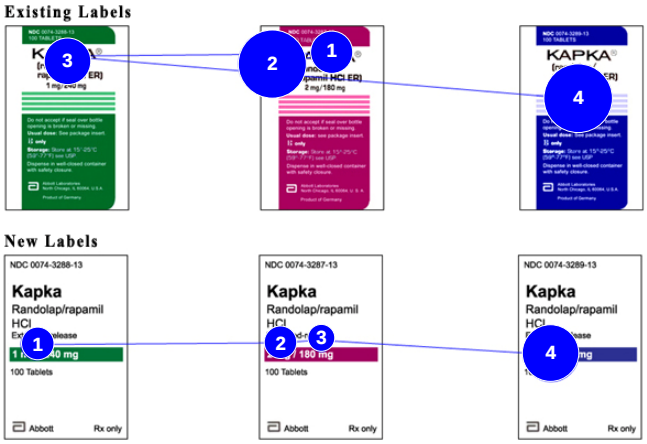

Figure 6. Sample scanpaths used to find a label with the strength specified in the task (here: *2 mg/240 mg*). Notice the longer fixations on the existing designs as compared to the new designs.

Expiration Date Task

The expiration date tasks showed significant differences in response time and number of fixations between the new and existing label designs. The redesigned side panel increased the speed of participants' responses and reduced the number of fixations required to locate the target information (Figure 7). Of the changes introduced in the redesign, it appears that the increased location consistency of the expiration date had the greatest impact on performance in this task, causing the reduction in fixation count. In the new designs, all information on the side panel was presented in the same orientation, and once the participant found the element in question in the first label he/she fixated, there was no need for reorientation on the subsequently fixated labels. The increased consistency of the date format and contrast had no significant impact on fixation duration in tasks with the new designs, possibly because these changes involved only one label of the six tested in this task.

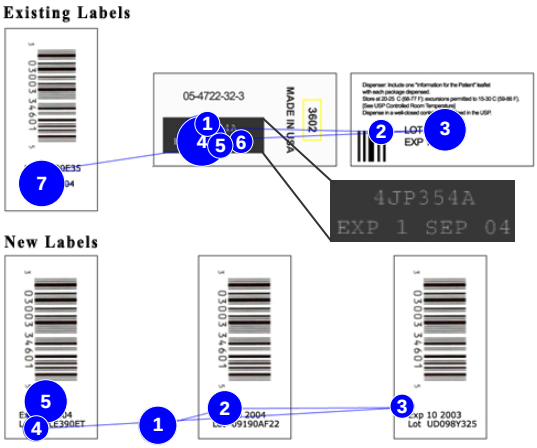

Figure 7. Sample scanpaths used to find a label with the expiration date specified in the task (here: *April of this year*). Notice the higher number of fixations on the existing designs as compared to the new designs.

CONCLUSION

The present study compared existing drug labels to a redesigned template and found overall performance improvements as a result of the introduced changes. Eye tracking data illustrated what happened "behind the scenes" to affect the noted speed increase – be it lowered search demands or lowered information processing demands posed by the new design. These insights helped uncover which types of changes had a more pronounced impact on performance than others, a true benefit in a situation when manipulating one variable at a time was simply not feasible. While this depth of analysis is usually not necessary to determine which design is more effective, the lessons learned can be of great value in redesign projects with a high level of business and/or technical constraints, where it is important to be selective in terms of the proposed changes.

While it would have been ideal to conduct the eye tracking study in a natural environment rather than in front of a computer screen, we believe that the eye movement data collected approximates the real world sufficiently to be of value (especially that the response times in the eye tracking study did not differ from the response times obtained in the test with actual bottles). We realize that the consistency across the new labels could have had a stronger impact on performance with a blocked rather than randomized design presentation, which would have been a more accurate representation of a real-world experience. Another consideration for a similar study in the future might be a better separation between search behavior and selection behavior by mapping responses to key presses rather than using the mouse.

Even though no error rate analysis could be performed, the observed savings indicate that the labels created with the new template made certain tasks *easier* for the pharmacy staff. Because the accuracy tends to increase as the task difficulty decreases, we believe the new labels should be noticeably less error-prone in the long run.

REFERENCES

Fisher, D. L. & Tan, K. C. (1989). Visual displays: The highlighting paradox. *Human Factors*, 31(1), 17-30.

Flynn, E. A., Barker, K. N., & Carnahan, B. J. (2003). National observational study of prescription dispensing accuracy and safety in 50 pharmacies. *Journal of the American Pharmaceutical Association*, 43(2), 191-200.

Halverson, T. & Hornof, A. J. (2004). Local density guides visual search: sparse groups are first and faster. *Proceedings of the 48th Annual Meeting of the Human Factors and Ergonomics Society*.

Kotval, X. P. & Goldberg, J. H. (1998). Eye movements and interface components grouping: an evaluation method. *Proceedings of the 42nd Annual Meeting of the Human Factors and Ergonomics Society*.

Parasuraman, R. (1986). Vigilance, monitoring and search. In K. Boff, L. Kaufman, & J. Thomas (Eds.), *Handbook of perception and human performance, Vol. 2: Cognitive processes and performance*. New York: Wiley.

Parkhurst, D. J. & Niebur, E. (2004). Texture contrast attracts overt visual attention in natural scenes. *European Journal of Neuroscience*, 19(3), 783-789.

Quinn, A., Bojko, A., Gaddy, C., & Israelski, E. (2005, manuscript submitted). Improving prescription drug labels: "Can you read me now?" *Ergonomics in Design.*

Tinker, M. A. (1963). *Legibility of Print.* Ames, Iowa: The Iowa State University Press.

Wickens, C. D., Gordon, S., and Liu, Y. (1998). *An introduction to human factors engineering*. New York: Addison Wesley Longman, Inc.

Yantis, S. & Egeth, H. E. (1999). On the distinction between visual salience and stimulus-driven attentional capture. *Journal of Experimental Psychology: Human Perception and Performance*, 25, 661-676.