

Similarity of Drug Names: Comparison of Objective and Subjective Measures

Bruce L. Lambert, PhD^{a, b}

Don Donderi, PhD^c

John W. Senders, PhD^{d, e}

^aDepartment of Pharmacy Administration

^bDepartment of Pharmacy Practice

University of Illinois at Chicago

^cDepartment of Psychology

McGill University

^d Department of Mechanical and Industrial Engineering

University of Toronto

&

^e Osgoode Hall School of Law

York University

Resubmitted to *Psychology & Marketing* July 20, 2001

Short Title: Drug Name Similarity

Address: Department of Pharmacy Administration,
833 S. Wood Street (M/C 871), Chicago, IL 60612-7231

Phone: 312-996-2411

Fax: 312-996-0868

Email: lambertb@uic.edu

Word Count: (excluding Abstract, References, and Appendix), 4940 total

Similarity of Drug Names: Comparison of Objective and Subjective Measures

Abstract

Previous research has shown that objective measures of orthographic (i.e., spelling) similarity can predict the probability of drug name confusion, but it is not clear how these objective measures relate to subjective judgments of similarity. This study examined the association between one objective measure of orthographic similarity, the Dice coefficient on trigrams, and one subjective measure, based on the Proscale multidimensional scaling system. Twenty-seven participants, divided into three groups, performed a similarity grouping task on one of three sets of 70 drug names drawn at random from a larger set of similar and dissimilar name pairs. Subjective groupings were converted to dissimilarity scores using the Proscale multidimensional scaling program. The association between subjective and objective measures was assessed by correlation and regression analyses. Correlations between subjective and objective measures were $-.70$, $-.48$, and $-.53$ for the three groups respectively ($p < 0.001$). Regression models with trigram similarity as the main predictor accounted for between 22 and 48 percent of the variance in subjective dissimilarity scores. We concluded that objective measures of orthographic similarity between drug names are valid but incomplete measures of subjective similarity.

Introduction

In spite of their skill, training, and vigilance, health professionals still make mistakes. Errors involving medication are among the most common of all medical mistakes (Allan & Barker, 1990; Anonymous, 1995; Cohen, 1999; Kohn, Corrigan, & Donaldson, 2000), in part because drug therapy is the most common form of treatment offered by physicians (Woodwell, 1999). In addition, the complexity of drug therapy, especially the size of the drug name lexicon, complicates the task of getting the right drug to the right patient in the right dose at the right time. A standard drug reference for the U.S. lists about 8000 brand names and 4000 generic names (U. S. Pharmacopeia, 1998a). Many of these names look and sound the same (e.g., *Accupril*[®] and *Accutane*[®], cisplatin/carboplatin, dopamine and dobutamine, *Lodine*[®] and codeine, *Premarin*[®] and *Primaxin*[®], *Xenical*[®] and *Xeloda*[®]) (Davis, 1997; U. S. Pharmacopeia, 1995b).

It is not surprising, then, that 15-25% of all reported medication errors in the U. S. involve confusions between drug names that look and/or sound too much alike (Davis, 1997; Lambert, Chang, & Lin, 2001; Lambert, Lin, Gandhi, & Chang, 1999b; U. S. Pharmacopeia, 1995c; U. S. Pharmacopeia, 2001). Recognizing the importance of drug names to patient safety, pharmaceutical manufacturers and federal regulators have recently increased their pre-marketing efforts to screen new drug names for potential confusion (Boring, 1997a; Boring, Homonnay-Weikel, Cohen, & Di Domizio, 1996; Boring, 1997b; Boring, Stein, & Di Domizio, 1998; DiDomizio & Cohen, 1995; Phillips, 2000). Most of these efforts involve collecting perceptual judgments and subjective similarity judgments from panels of practicing health professionals. Clearly, these

approaches reflect the belief that subjective judgments are a valid predictor of eventual marketplace confusion.

Another method of screening involves automated comparison of a new name with a database of existing names, using objective measures of lexical (i.e., word-to-word) similarity. This paper evaluates one of these objective measures, the trigram measure of orthographic similarity, by examining its ability to predict subjective similarity judgments of lay people. Although the subjective similarity measure is, itself, not a perfect criterion measure (the perfect measure would be name confusion errors in real-world settings), we reasoned that (a) subjective measures are related to error, (b) subjective measures are often used to make name approval decisions in the real world, and (c) evidence of the association between objective and subjective measures would provide evidence of the convergent validity of the objective measure.

Among those who study human error, the detrimental effects of similarity on human performance are well known (Reason, 1990). These effects are most notable in psycholinguistics, where research has shown that similarity between words degrades performance in recall, recognition, auditory perception, and visual perception (Anderson, Bothell, Lebiere, & Mantessa, 1998; Baddeley, 1986; Grainger & Jacobs, 1996; Lambert et al., 2001; Luce & Pisoni, 1998; Reason, 1990; Shulman, 1971; Wallace, Stewart, Shaffer, & Wilson, 1998; Wallace, Stewart, Sherman, & Mellor, 1995). In most of these investigations, similarity has been defined simplistically. Orthographic (i.e., spelling) similarity has often been defined as two words that differ by one letter in one position (e.g., *bat* and *cat*) (Coltheart, Davelaar, Jonasson, & Besner, 1977). Phonological

similarity has been defined in much the same way, with the differences pertaining to single phonemes rather than single letters (Luce & Pisoni, 1998). (The term *phoneme* refers to the smallest unit of sound in a language. There are 44 different phonemes in English, e.g., the /m/ in *me* /r/ in *ran* (Pickett, 1999).) With few exceptions, these studies have been done on monosyllabic words. Monosyllabic words were chosen to simplify processing assumptions and also because similarity measures for multisyllabic words had not been defined. So, in spite of their relevance to central questions in psycholinguistics, general measures of lexical similarity, especially for polysyllabic words, have not been adequately investigated (Luce & Pisoni, 1998).

The absence of lexical similarity measures in psycholinguistics does not reflect a general lack of such measures. In fact, computer science is replete with measures of both orthographic (i.e., look-alike) and phonetic, (i.e., sound-alike) similarity (Aoe, 1994; Stephen, 1994; Zobel & Dart, 1995; Zobel & Dart, 1996). In retrospective studies of error databases and in prospective studies of short term memory, Lambert has demonstrated that simple, objective measures of similarity can reliably predict which drug names are most likely to be confused (Lambert, 1997; Lambert et al., 2001; Lambert, Chang, & Lin, under revision; Lambert, Lin, Gandhi, & Chang, 1999a; Lambert et al., 1999b). In a comparative evaluation of 22 different orthographic and phonetic similarity measures, the trigram measure with two spaces added to each word performed best (Lambert et al., 1999b). To compute the trigram similarity score between two drug names, the names are first broken down into three-letter subsequences. For example, the trigrams for the drug *Accupril*[®] (when two spaces are added to the beginning of the name), are {-A, -Ac, Acc,

ccu, cup, upr, pri, ril}. For *Accupril*[®], the trigrams are {--A, -Ac, Acc, ccu, cut, uta, tan, ane}. The similarity score is computed as a function of the number of common trigrams (see Methods section for a complete description of the similarity metric).

The work by Lambert and others has shown that objective measures of similarity can be used to predict drug name confusion errors. However, similarity is a multidimensional construct. At a minimum, similarity between words involves orthography (i.e., spelling) and phonology (i.e., sound). Surely there is a semantic (i.e., meaning) component to similarity as well that has yet to be thoroughly explored. The question remains, then, as to how objective measures such as trigram relate to subjective judgments of similarity. After all, the trigram measure is designed to be a proxy for subjective judgments of similarity. If we can predict subjective judgments of similarity, it should be possible to guide the design of systems that could predict and prevent drug name confusion errors. Therefore, in designing the present study, we sought to answer the following research question:

RQ1: How well does the trigram similarity measure predict lay participants' subjective judgments of similarity between drug names.

On both theoretical and empirical grounds, we had reason to believe that the trigram measure would be strongly associated with subjective judgments of similarity. Intuitively, it seems that similarity in spelling ought to be associated with perceived similarity. Empirically, the objective measures have been strongly associated with name confusion errors in both retrospective and prospective designs (Lambert, 1997; Lambert et al., 2001; Lambert et al., 1999a; Lambert et al., 1999b). Theoretically, similarity is a

key component in every major cognitive model of recall, recognition, and perceptual identification (Anderson et al., 1998; Anderson & Lebiere, 1998; Baddeley, 1986; Grainger & Dijkstra, 1996; Luce & Pisoni, 1998; Shulman, 1971). Although we could provide detailed conceptual models of recall, recognition, and visual and auditory perceptual identification, doing so would not be relevant to our purpose, which is not to test a model of one psychological faculty but rather to cross-validate two distinct but related methods for assessing similarity. Readers interested in the details of the various models should consult the references cited above.

In generating our experimental hypothesis, we reasoned as follows:

- (1) the subjective approach measures both perceptual (i.e., formal) and semantic similarity;
- (2) orthographic similarity is a major component of perceptual similarity;
- (3) perceptual similarity is one basis for confusion;
- (4) things that are more confused should show more, and those less confused less, of both cognitive and orthographic similarity.

Donderi's previous work on subjective similarity supports the first contention (Donderi, 1988; Donderi, 1997; Donderi & Aspler, 1994; Donderi & Aspler, 1998; Donderi & Aspler, in press; Donderi, Jordan, Aspler, & O'Neill, 2001). We take the second and third statements as axiomatic. Thus, our experimental hypothesis, derived from (4), was that the trigram measure of orthographic similarity would account for a significant proportion of the variance in participants' subjective judgments of similarity.

Method

Measures of Similarity or Dissimilarity

The Dice coefficient of on letter trigrams. Words were first broken into three-letter subsequences called trigrams. The dice similarity coefficient (D) for a pair of words is calculated as $D = 2C/(A+B)$, where A = the number of unique trigrams in word 1, B = the number of unique trigrams in word 2, and C = the number of identical trigrams in the two words (Lambert, 1997; Lambert et al., 1999b; Stephen, 1994). In counting trigrams, two spaces were added to the beginning of each word in order to emphasize similarity in the beginning of words. This measure can range from zero (no identical trigrams in the two words) to one (identical words; all trigrams identical). Using this measure, the trigram similarity between *Accupril*[®] {--A, -Ac, Acc, ccu, cup, upr, pri, ril} and *Accutane*[®] {--A, -Ac, Acc, ccu, cut, uta, tan, ane}, which share 4 trigrams {--A, -Ac, Acc, ccu }, is $(2*4)/(8 + 8) = 0.5$.

The Proscale measure of dissimilarity. Proscale is a multidimensional scaling method that is based on an isomorphism between information measurement and euclidean distance. By means of this isomorphism, similarity classifications are transformed into distances in a euclidean dissimilarity space (Donderi, 1988). Multidimensional scaling methods including Proscale have been used to study the relative quality of images printed on different types and qualities of paper (Donderi & Aspler, 1994; Donderi & Aspler, in press; Donderi et al., 2001; Parush & Lyne, 1983). In all of these studies the goal has been to correlate subjective measures of preference and similarity with objective characteristics of the printed paper samples, and in each of the

studies a correlation was found between some physical measure of print or paper quality, and the subjective similarities among the samples as well as the preference rating for each sample.

The participant in a Proscale study is shown a set of samples (as many as 70), and is asked to place the samples into groups so that “each group has something in common that distinguishes it from the other groups.” The only forbidden groupings are: all samples in the same group or each sample in its own group. Other than that, the participants are free to group the samples any way they choose, finding similarities among them that justify their grouping. After the participant has classified the samples into groups, he or she assigns a unique preference value to each group, so that all the samples in that group are ranked the same in preference, and each group has a different preference ranking. Finally, the participant indicates which preference value (the “cutoff value”) is the lowest that he or she would prefer to buy or use if given a choice. Although we make passing mention of the preference data here, a complete analysis of those data will be reported elsewhere.

The Proscale program (Donderi, 1997; Donderi & Aspler, 1998) uses the information-distance metric to transform each participant's classifications into distances in a euclidean space. The distance between any two samples is interpreted as a sample pair dissimilarity, and the distance between a sample and the origin of the space is interpreted as the distinctiveness of that sample. The Proscale program also generates a principal components solution to the dimensionality of the average dissimilarity space

generated by the participants, and generates a quartimax factor rotation to simple structure for the number of retained spatial dimensions selected by the user.

Drug Names

We used a total of fifty-four pairs of high similarity drug names and fifty-one pairs of low-similarity names. The actual name pairs used (210 names in all) are presented in Appendix A, along with the trigram similarity scores calculated between the words in each high-similarity and low-similarity pair. The high-similarity name pairs had trigram scores (S) > 0.65 , while all the low-similarity pairs had $S = 0$.

The participants in the study did not, however, see “high similarity pairs” or “low similarity pairs” of drug names. Instead, we divided the set of 210 names into three sets of 70 names each, and tested each set with a different group of subjects. Each participant simply saw 70 individual drug names, each typed on a 3 x 5 card. Each set contained equal numbers of high-similarity and low-similarity name pairs. The participants were told nothing about the words in each set. They were not told which words had been paired as high or low similarity, and they were not told about any objective similarity measures relating to the words.

Set 1 consisted of the words in the first 18 pairs of high-similarity names and the words in the first 17 pairs of low-similarity names Set 2 consisted of the words in high-similarity pairs 19 through 36, and low-similarity pairs 18 through 35. Set 3 consisted of the words in high-similarity pairs 37 through 54, and words in low-similarity pairs 36 through 51.

Trigram Similarity Scores were Calculated for all Name Pairs

We calculated trigram similarity scores among all possible name pairs within each of the three sets of 70 words. Therefore we could compare the subjective dissimilarity measure generated by the Proscale similarity classification task with the objective Dice coefficient for each possible name pair, including all the pairs within each set that were not matched as either high or low-similarity as shown in Appendix A. In each set there were $(70*69)/2 = 2415$ different name pairs with measured trigram similarity scores (S) and Proscale dissimilarity (P) values.

Participants

The 27 participants, 10 men and 17 women, ranged in age from 17 to 55 (average 26). Their education ranged from high school student to university graduate, with an average 13 years of schooling. The participants were divided at random into three groups of 10 (Set 1), 8 (Set 2) and 9 (Set 3).

Test Conditions

Participants were tested individually. Each participant sat in front of a large table in an empty university seminar room. The 70 drug names in each set were spread out on the table in a different (shuffled) random order for each participant. As described earlier, each participant classified the names into groups based on his or her own criteria of similarity among the names, then assigned a unique preference rating to each group, and then assigned a cutoff preference value for the entire set. The testing sessions took between twenty and 45 minutes per participant.

Analysis

Results from each of the three Sets were analyzed using the Proscale multidimensional scaling program. This program generated a preference value for each name, a distinctiveness measure for each name, and a dissimilarity measure for each name pair. It also generated a principal components solution for the dimensionality of the euclidean dissimilarity space that included the distinctiveness vectors for each name and the dissimilarity vectors between each name pair (a total of 2485 euclidean distances per set).

Within each set of 70 drug names, the Proscale dissimilarity measures generated for each name pair were compared directly with the trigram similarity score calculated for that pair. The comparison within each set took the form of a scatter plot of Proscale similarity versus the trigram similarity score for each name pair, and a linear regression equation relating the Proscale dissimilarity rating to the trigram similarity scores within each Set.

Since there were seventeen pairs of similar drug names in each set, and since, as a matter of fact, seventeen principal components accounted for 100% (set 1), 97% (set 2) and 92% (set 3) of the variance of euclidean distances within each set, seventeen dimensions were retained and rotated to simple structure to describe the dissimilarity space for each set.

Results

Name Pair Dissimilarity

Figures 1, 2 and 3 show the relationship between trigram similarity as measured by the Dice coefficient (on the x-axis) and Proscale dissimilarity (on the y-axis) for each of the 2415 word pairs in test sets 1, 2 and 3. The correlation coefficients describing the scatter are $r = -.70$ ($p < .001$) for Set 1, $r = -.48$ ($p < 0.001$) for Set 2 and $r = -.53$ ($p < 0.001$) for Set 3. The regression equations relating Proscale dissimilarity (P) to the trigram similarity scores (S) are $P = -2.12S + 2.18$ for Set 1, $P = -1.25S + 2.19$ for Set 2, and $P = -1.80S + 2.15$ for Set 3. The values for the the constant of each formula fall within the confidence intervals for all three sets. Therefore the mean constant, which is the Proscale dissimilarity value when the trigram similarity score equals zero, is 2.17 for all of the 7245 word pairs tested across the three sets. The slope coefficient of -2.13 for Set 1 and -1.80 for Set 3 are not significantly different. These coefficients are significantly different from the coefficient of -1.25 calculated for Set 2. The relationship between Proscale dissimilarity and the trigram similarity score was weaker for Set 2 than for Sets 1 and 3. We have no explanation for the difference in the strength of the Proscale – Dice coefficient relationship among Sets 1, 2 and 3, other than that the differences were cause by uncontrolled variation among the observers who carried out the similarity grouping task on each set.

Insert Figures 1-3 about here.

The trigram similarity score accounted for about 48 percent of the variance of Proscale dissimilarity measures in Set 1, about 22 percent in Set 2, and about 27 percent in Set 3. This means that about between 50 and 80 percent of the variance of the Proscale dissimilarity measure of the word pairs in the three sets is not predicted by trigram similarity as measured by the Dice coefficient.

One characteristic of the Dice coefficient is that it is zero for a word pair which has no trigrams in common, regardless of any other properties of the pair. The components of the Dice coefficient are C , the number of common trigrams between the two words, A , the number of trigrams in one word of the pair, and B , the number of trigrams in the other word of the pair. We studied the relationship between each of these components, and the Proscale dissimilarity measure. We found essentially no relationship between A and B , the total number of trigrams in each word, and Proscale dissimilarity. In addition, C , which is the number of trigrams the two words have in common, was less closely related to Proscale similarity than was the complete Dice coefficient $2C/(A + B)$. But for all three drug name sets, when $C = 0$ and therefore $D = 0$, Proscale dissimilarity = 2.17, an upper limit for the many word pairs in the three sets that had no trigrams in common. A variety of functional relationships involving C , A and B was explored to determine whether it was possible to improve on the one-variable linear regression equations described above. No combination reduced variance more over the entire range of word pairs than the simple two-part functions described above: $P = 2.17$ when $S = 0$, and P a decreasing linear function of S when $S > 0$.

Dimensional Analysis of the Words in Each Set

The Proscale dissimilarity space defines both the distance of each drug name from the origin of the space, which is its distinctiveness, as well as the distance between the members of each drug name pair, which is their dissimilarity. This complete euclidean space can be factor analyzed to determine how many orthogonal principal component dimensions are required to account for all the distances in the space. Then the set of orthogonal dimensions can be rotated to an optimized factor analytic criterion of simple structure, which means that the position of each word in the space will be defined by high absolute values on a small number of dimensions, and low absolute values on the rest (Gorsuch, 1983). We did this for the dissimilarity spaces defined by Proscale measures in Sets 1, 2 and 3. Almost all of the variance of the euclidean dissimilarity space for each set was accounted for by seventeen retained and rotated orthogonal principal component dimensions (Set 1, 100 percent, Set 2, 97 percent, Set 3, 92 percent). These dimensions represent seventeen independent characteristics of the 70 words in each set. Figures 4, 5, 6, and 7 display the similarity dimension profiles for two similar and two dissimilar drug names respectively. These figures show that words high in objective similarity tend to have nearly identical profiles in the reduced-dimension space defined by the Proscale measure. Conversely, words low in objective similarity tend to have divergent profiles in subjective similarity space. This is just another way of illustrating, in addition to the scatterplots of Dice scores versus Proscale scores, the extent of the agreement between subjective and objective similarity measures.

Insert Figures 4-7 about here.

Discussion and Conclusions

Cross-Validation of Subjective and Objective Measures

Trigram similarity scores, as measured by the Dice coefficient, have been used by Lambert to predict confusions between drug name pairs (Lambert, 1997; Lambert et al., 1999a; Lambert et al., 1999b). This study provides additional cross-validation of the Dice/trigram measure by showing that it also correlates significantly with measures of subjective dissimilarity among all possible pairs of drug names. In each set of 70 drug names, half of the names were pairs of highly similar drug names, and half were highly dissimilar. The trigram similarity score accounted for between 22 and 48 percent of the variance in dissimilarity across all possible pairs of drug names; not just those that were paired as confusable, but among all possible pairs of the confusable names, all possible pairs of the non-confused names, and all possible pairs drawn one from each set. This leaves between 52 and 78 percent of the variance of drug name dissimilarity across each set to be accounted for. Although the observed relationship between the Dice coefficient score and the Proscale similarity dimensions is not as strong as one might wish, it suggests that, with additional refinements, the relative confusability of two drug names may be predicted in advance, rather than being left to determination in the marketplace.

Practical Applications

The current study, which shows only a modest association between objective and subjective measures, does not itself justify the claim that objective measures should be used to predict future errors. However, when taken in the context of previous work, the claim is stronger. For example, in retrospective, case-control analyses of reported errors, objective measures such as those tested here were able to separate cases (known errors) and controls (randomly selected pairs of names) with 94% accuracy (Lambert, 1997; Lambert et al., 1999b). Also, in our analyses of recognition memory errors, the fit between the data and a model based on orthographic similarity was quite good (predictive accuracy between 65-70%) (Lambert et al., 2001). Models of pharmacists' visual perception errors show equally good classification performance (Lambert, Chang, Lin, & Gupta, 2000). Thus, we have reason to believe that objective similarity measures can form the basis of an automated decision-support tool that would assist experts during pre-approval screening of new drug names.

Prediction of errors based on objective similarity should *not* be the *sole* basis for name-approval decisions. Rather, the results of a computerized search of existing name databases should be *part* of a more elaborate name-vetting process such as that currently used by the U. S. Food and Drug Administration (FDA) or the Institute for Safe Medication Practices, both of whom still rely, at least in part, on health professionals' subjective similarity judgments (Institute for Safe Medication Practices, 2001; Phillips, 2000). Software created by the first author has already been used by major pharmaceutical companies to screen proposed drug names prior to FDA approval. A total

of more than 150 distinct searches have been performed between 1997 and 2000 (where each search involves the comparative evaluation of up to 10 names). The same software system has also been used by the United States Adopted Names Council (USANC) to screen new non-proprietary (generic) names. The USANC used the system for one year, for a total of more than 150 searches. In both cases, the client submitted a list of candidate names. Each name was submitted as a query against a database of brand and generic names (U. S. Department of Commerce--Patent and Trademark Office, 1994; U. S. Pharmacopeia, 1995a; U. S. Pharmacopeia, 1998b). Orthographic similarity or distance scores were computed for each name in the database, and the top n most similar/least distant names were displayed to the user ($n = 50$ names by default). The results were used by decision makers within the respective agencies to eliminate candidate names that were too similar to existing names.

The Multi-Dimensionality of Similarity

Seventeen similarity space dimensions accounted for almost all of the variance in the distances of the Proscale dissimilarity spaces for all three sets of drug names, and we found tentative phonemic or graphemic descriptions for most of the dimensions in the three sets. (A *grapheme* is a basic unit of written language, a letter or set of letters, which corresponds to a distinct *phoneme* in spoken language.) But only between 22 and 48 percent of the variance of the Proscale dissimilarity measures was accounted for by the Dice/trigram similarity scores. Why is this? The answer lies in the multidimensionality of similarity. For drug names, these dimensions are likely to include orthography (i.e., spelling), phonology (i.e., sound pattern), semantics, (i.e., literal and connotative

meaning), frequency of use, recency of use, ease of pronunciation, word shape, print color, type and size of font, and more. For drug *products*, as opposed to just drug *names*, the dimensions of similarity must include strength (in milligrams, milliliters, drops, etc.), schedule (e.g., twice a day), dosage form (e.g., capsule, tablet, oral solution, cream, ointment, suppository), color, shape, packaging, indication (i.e., what it is used for), mechanism of action (i.e., how it works), and so on. Thus, having used in this experiment *only* a measure of orthographic similarity, we did not expect to account for all of the variance in the subjective similarity measure. We did expect the association to be significantly larger than zero (and it was). More importantly, we wanted to quantify the *magnitude* of the contribution of spelling similarity to overall subjective similarity. We now have evidence that spelling similarity accounts for, at most, 50% of the observed variability in subjective similarity. Subsequent research will allow us to replicate this finding and to quantify the contributions of various other dimensions of similarity, some of which we discuss below.

Other phonetic, graphemic, and semantic measures (Lambert et al., 1999b), either in combination with or independently of the Dice/trigram measure, may better describe the dissimilarities found using the Proscale multidimensional scaling technique. For example, it is possible to capture sound-alike similarities using orthographic (i.e. spelling) representations. This is done by grouping similar-sounding letters (e.g., *p* and *b*) into equivalence classes and then treating them as if they were the same during similarity computations. Several methods have been described recently, and Lambert evaluated most of them in the context of drug name confusion (Lambert et al., 1999b; Zobel &

Dart, 1996). It is possible to transform the orthographic (i.e., spelling) representations of words in to phonological (i.e., pronunciation-based) representations, and subsequently compute similarity scores between the phonological representations (Fisher, 1999a; Fisher, 1999b; Fisher & Fiscus, 1993; Fisher, Fiscus, Martin, Pallett, & Pryzbocki, 1995).

Semantic features will probably also play some role. One can imagine, for example, defining similarity functions for drug strengths, schedules, routes of administration, and dosage forms (e.g., tablets are more similar to capsules than they are to oral solutions), but devising, combining, and evaluating these measures is a task for another study. Until then, we conclude that the Dice/trigram measure of orthographic similarity between drug names is a valid, albeit incomplete, measure of lay persons' subjective judgments of the similarity between drug names. This conclusion enhances the convergent validity of the Dice/trigram measure and reinforces our belief that the likelihood of potentially dangerous, real-world drug name confusions can be minimized by using objective measures of similarity during the pre-approval screening process for new drug names.

References

- Allan, E. L., & Barker, K. N. (1990). Fundamentals of medication error research. American Journal of Hospital Pharmacy, 47, 555-571.
- Anderson, J. R., Bothell, D., Lebiere, C., & Mantessa, M. (1998). An integrated theory of list memory. Journal of Memory and Language, 38, 341-380.
- Anderson, J. R., & Lebiere, C. (Eds.). (1998). The atomic components of thought. Mahwah, NJ: Erlbaum.
- Anonymous. (1995). Understanding and preventing drug misadventures. American Journal of Health-System Pharmacists, 15(February 15), 369-416.
- Aoe, J. (1994). Computer algorithms: String pattern matching strategies. Washington, DC: IEEE Computer Society Press.
- Baddeley, A. D. (1986). Working memory. Oxford: Oxford University Press.
- Boring, D. (1997a). The development and adoption of nonproprietary, established, and proprietary names for pharmaceuticals. Drug Information Journal, 31, 621-634.
- Boring, D., Homonnay-Weikel, A. M., Cohen, M., & Di Domizio, G. (1996). Avoiding trademark trouble at FDA. Pharmaceutical Executive, 16, 80-8.
- Boring, D. L. (1997b). The CDER labeling and nomenclature committee: Structure, function, and process. Drug Information Journal, 31, 7-11.
- Boring, D. L., Stein, I. A., & Di Domizio, G. (1998, July). United States: Trademark trainwrecks at FDA. Trademark World, 29-33.
- Cohen, M. (Ed.). (1999). Medication errors. Washington, DC: American Pharmaceutical Association.

Coltheart, M., Davelaar, E., Jonasson, J. T., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), Attention and Performance VI (pp. 535-555). New York, NY: Academic Press.

Davis, N. M. (1997). Drug names that look and sound alike. Hospital Pharmacy, 32(11), 1558-1570.

DiDomizio, G., & Cohen, M. R. (1995). Drug names can be a factor in medication errors. Product Liability Law and Strategy, 13(7), 1-4.

Donderi, D. C. (1988). Information measurement of distinctiveness and similarity. Perception and Psychophysics, 44(6), 576-584.

Donderi, D. C. (1997). A multidimensional scaling program for distinctiveness and similarity based on stimulus classification and information measurement. Behavior research methods, instruments & computers, 97(4), 549-555.

Donderi, D. C., & Aspler, J. A. (1994). Correlates of preference and similarity in 45 paper samples from 9 newsprint machines. Nordic Pulp and Paper Research Journal, 2, 88-93.

Donderi, D. C., & Aspler, J. A. (1998). Proscale: A research method and computer program for sample quality evaluation and multidimensional scaling (Miscellaneous Report MR 382): Pulp and Paper Research Institute of Canada.

Donderi, D. C., & Aspler, J. S. (in press). A fast new way to measure subjective print and paper quality. Pulp and Paper Canada.

Donderi, D. C., Jordan, B., Aspler, J., & O'Neill, M. (2001). The subjective print quality of commercial heatset offset prints on lightweight coated paper . Montreal, Quebec: Pulp and Paper Research Institute of Canada.

Fisher, W. M. (1999a). aldistsm-1.2 (1.2). National Institute of Standards. Available: <http://www.itl.nist.gov/iaui/894.01/software.htm>.

Fisher, W. M. (1999b, March 15-19). A statistical text-to-phone function using ngrams and rules. Paper presented at the IEEE International Conference on Acoustics, Speech and Signal Processing, Phoenix, AZ.

Fisher, W. M., & Fiscus, J. G. (1993). Better alignment procedures for speech recognition evaluation. Proceedings of the International Conference on Acoustics, Speech and Signal Processing, 59-62.

Fisher, W. M., Fiscus, J. G., Martin, A., Pallett, D. S., & Pryzbocki, M. A. (1995). Further studies in phonological scoring. Proceedings of the Spoken Language Systems Technology Workshop, 181-186.

Gorsuch, R. L. (1983). Factor analysis. Hillsdale, NJ: Erlbaum.

Grainger, J., & Dijkstra, T. (1996). Visual word recognition: Models and experiments. In T. Dijkstra & K. de Smedt (Eds.), Computational psycholinguistics (pp. 139-165). Bristol, PA: Taylor & Francis.

Grainger, J., & Jacobs, A. M. (1996). Orthographic processing in visual word recognition: A multiple read-out model. Psychological Review, 103(3), 518-565.

Institute for Safe Medication Practices. (2001). Med-E.R.R.S., [Web page]. Institute for Safe Medication Practices. Available: www.ismp.org.

Kohn, L. T., Corrigan, J., & Donaldson, M. S. (Eds.). (2000). To err is human: building a safer health system. Washington, DC: Institute of Medicine.

Lambert, B. L. (1997). Predicting look- and sound-alike medication errors. American Journal of Health-System Pharmacy, *54*, 1161-1171.

Lambert, B. L., Chang, K. Y., & Lin, S. J. (2001). Effect of orthographic and phonological similarity on false recognition of drug names. Social Science & Medicine, *52*, 1843-1857.

Lambert, B. L., Chang, K.-Y., & Lin, S.-J. (under revision). Effect of orthographic and phonological similarity on immediate free recall of drug names. Journal of Applied Cognitive Psychology.

Lambert, B. L., Chang, K.-Y., Lin, S.-J., & Gupta, P. (2000). Effect of prescribing frequency, neighborhood frequency, and neighborhood density on pharmacists' visual perception of drug names. PharmSci, *2*, S-1276.

Lambert, B. L., Lin, S.-J., Gandhi, S. K., & Chang, K.-Y. (1999a). Predicting and preventing drug name confusion errors: A summary of findings. In A. L. Scheffler & L. A. Zipperer (Eds.), Proceedings of Enhancing Patient Safety and Reducing Errors in Health Care (pp. 221-225). Rancho Mirage, CA: National Patient Safety Foundation.

Lambert, B. L., Lin, S.-J., Gandhi, S. K., & Chang, K.-Y. (1999b). Similarity as a risk factor in drug name confusion errors: The look-alike (orthographic) and sound-alike (phonological) model. Medical Care, *37*(12), 1214-1225.

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. Ear & Hearing, *19*(1), 1-36.

Parush, A., & Lyne, M. B. (1983). The print quality of offset and letterpress newsprint. Journal of Pulp and Paper Science, 9(TR 117).

Phillips, J. (2000,). The name game: New realities at FDA. Pharmaceutical Executive, July, 66-69.

Pickett, J. M. (1999). The acoustics of speech communication. Needham Heights, MA: Allyn and Bacon.

Reason, J. (1990). Human error. New York: Cambridge University Press.

Shulman, H. G. (1971). Similarity effects in short-term memory. Psychological Bulletin, 75, 399-415.

Stephen, G. A. (1994). String searching algorithms. River Edge, NJ: World Scientific.

U. S. Department of Commerce--Patent and Trademark Office. (1994). Trademark CD-Rom User's Guide. Washington, DC: Author.

U. S. Pharmacopeia. (1995a). USP DI, Vol. II; Advice for the patient. Rockville, MD: U. S. Pharmacopeia.

U. S. Pharmacopeia. (1995c). USP Quality Review (46). Rockville, MD: U. S. Pharmacopeia.

U. S. Pharmacopeia. (1995b). USP Quality Review (49). Rockville, MD: U. S. Pharmacopeia.

U. S. Pharmacopeia. (1998a). USP DI, Vol. I: Drug information for the health care professional. Rockville, MD: U. S. Pharmacopeia.

U. S. Pharmacopeia. (1998b). USP dictionary of USAN and international drug names. Rockville, MD: U. S. Pharmacopeia.

U. S. Pharmacopeia. (2001, March). USP Quality Review (76). U. S. Pharmacopeia. Available: <http://www.usp.org/reporting/review/qr76.pdf>.

Wallace, W. P., Stewart, M. T., Shaffer, T. R., & Wilson, J. A. (1998). Are false recognitions influenced by prerecognition processing? Journal of Experimental Psychology: Learning, Memory, and Cognition, *24*(2), 299-315.

Wallace, W. P., Stewart, M. T., Sherman, H. L., & Mellor, M. D. (1995). False positives in recognition memory produced by cohort activation. Cognition, *55*, 85-113.

Woodwell, D. A. (1999). National Ambulatory Medical Care Survey: 1997 summary. US National Center for Health Statistics. Advance Data from Vital and Health Statistics, *305*, 1-26.

Zobel, J., & Dart, P. (1995). Finding approximate matches in large lexicons. Software-Practice and Experience, *25*(3), 331-345.

Zobel, J., & Dart, P. (1996, August 18-22). Phonetic string matching: Lessons from information retrieval. Paper presented at the 19th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Zurich, Switzerland.

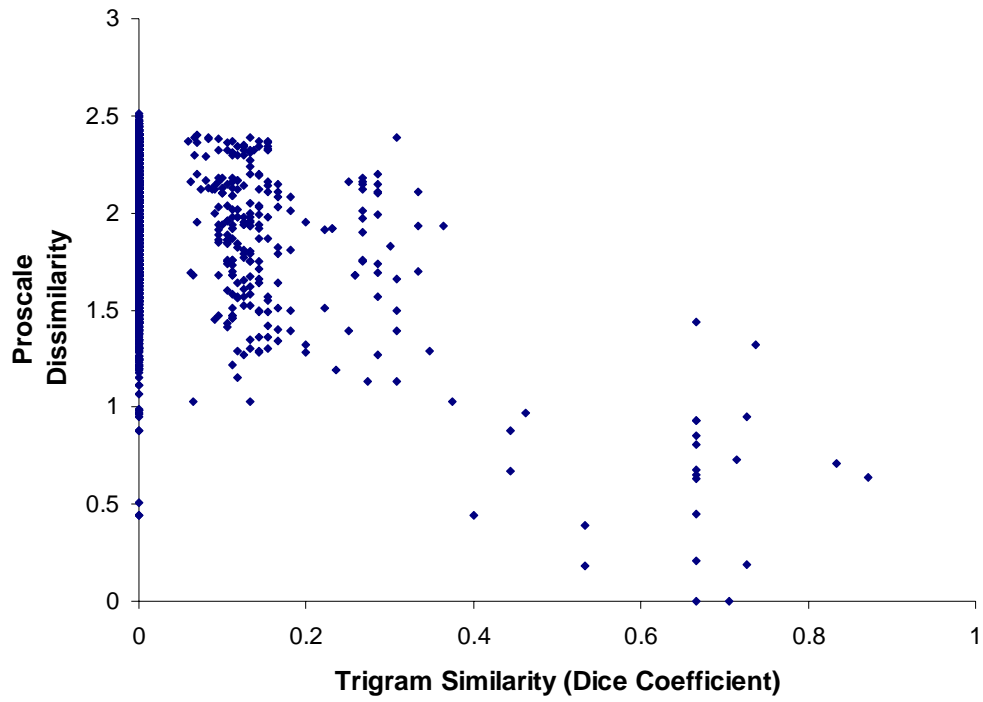


Figure 1. Scatterplot of Proscale distance and trigram similarity for group 1.

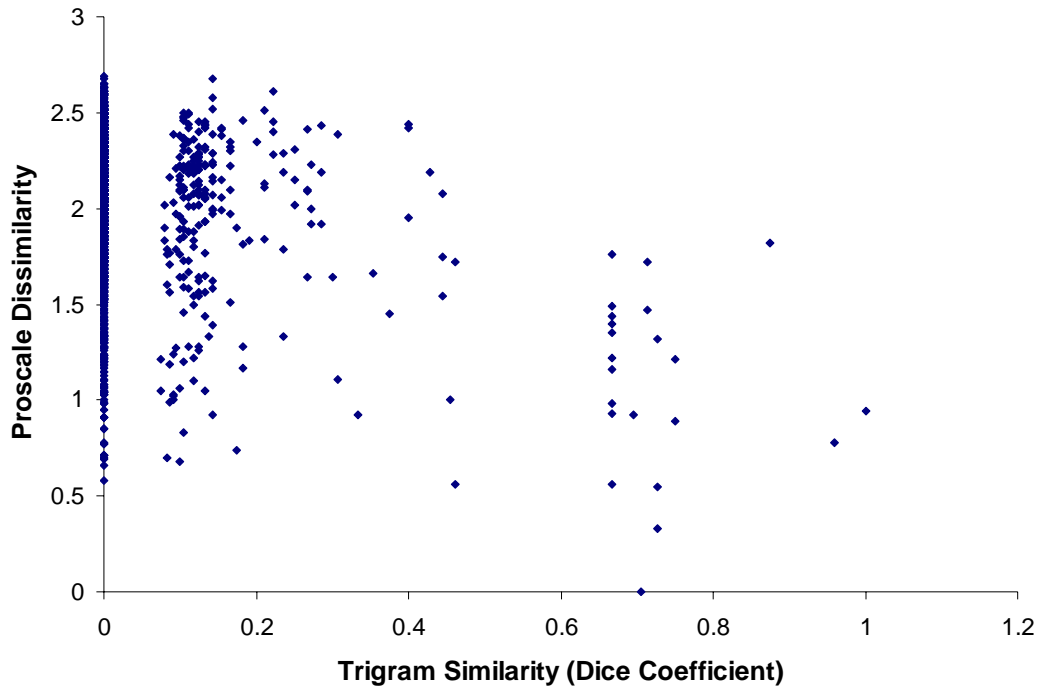


Figure 2. Scatterplot of Proscale distance and trigram similarity for group 2.

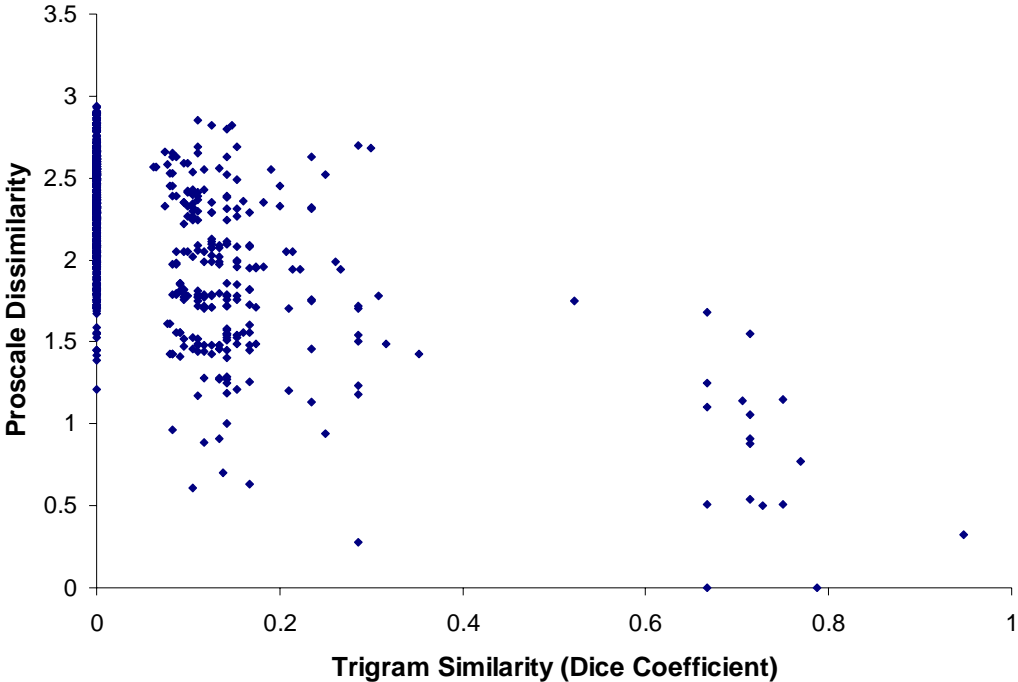
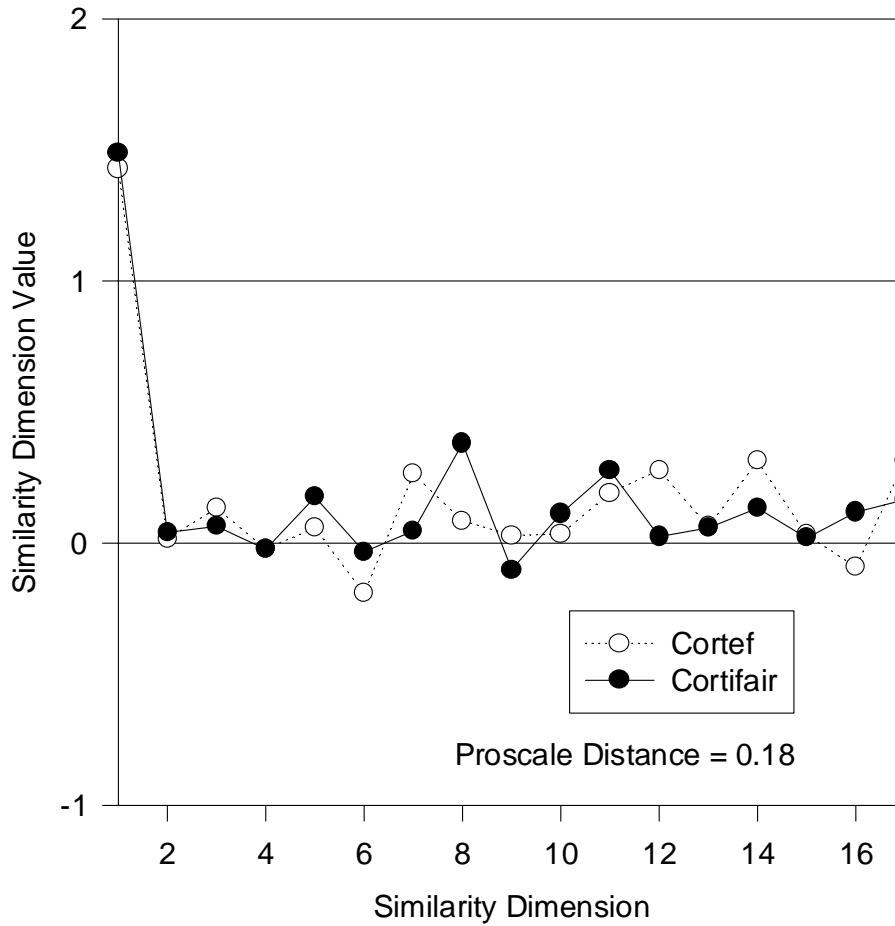


Figure 3. Scatterplot of Proscale distance and trigram similarity for group 3.

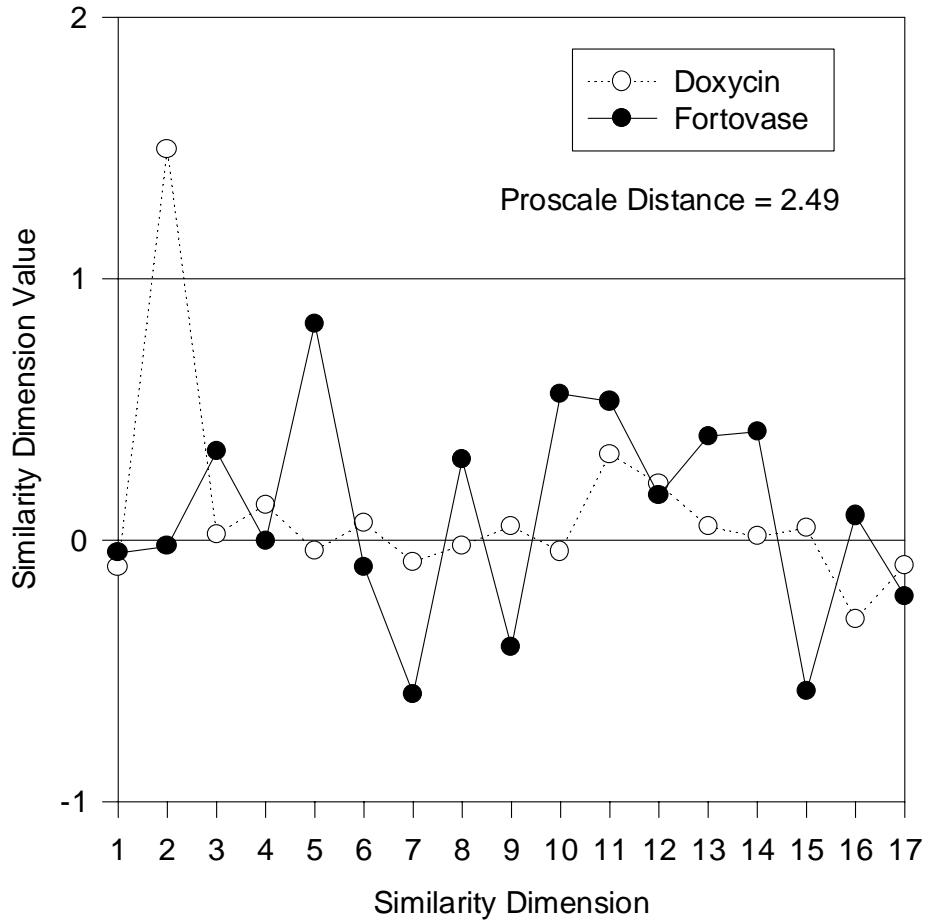
Similarity Dimension Profiles of a Similar Name Pair in Group 1



GP1HIPRO.SPW

Figure 4. Similarity dimension profiles of a similar name pair in group 1.

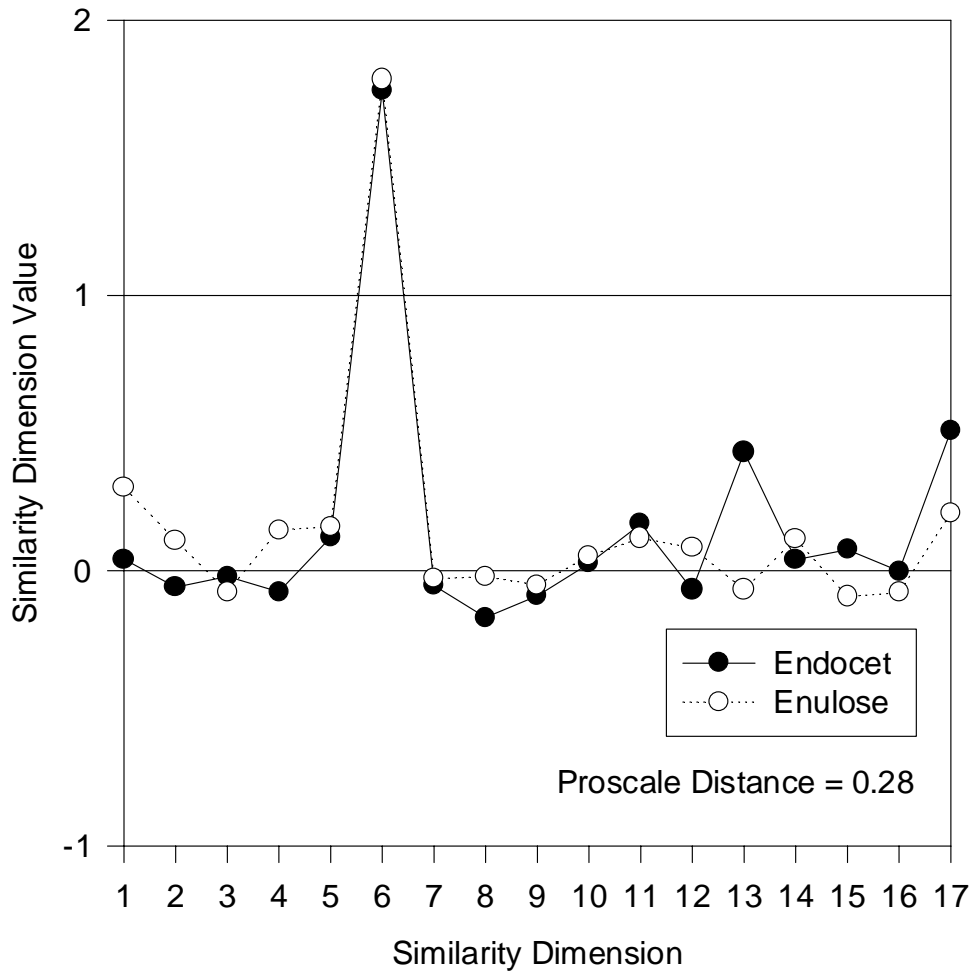
Similarity Dimension Profiles of the Least Similar Name Pair in Group 1



GP1LOPRO.SPW

Figure 5. Similarity dimension profiles of the least similar name pair in group 1.

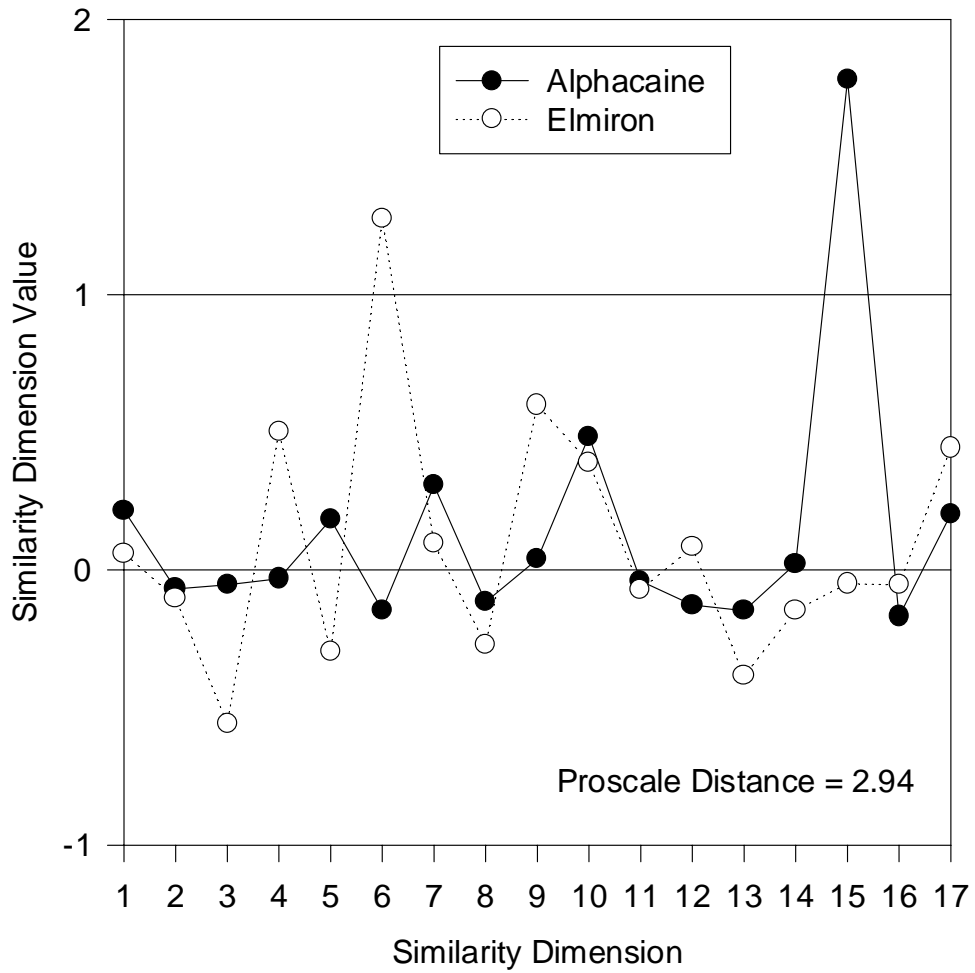
Similarity Dimension Profiles of a Similar Name Pair in Group 3



SP3HIPRO.SPW

Figure 6. Similarity dimension profiles of a similar name pair in group 3.

Similarity Dimension Profiles of the Least Similar Name Pair in Group 3



SP3LOPRO.SPW

Figure 7. Similarity dimension profiles of the least similar name pair in group 3.

Appendix A

High and Low Similarity Name Pairs

Below are 54 high-similarity name pairs and 51 low similarity name pairs.

Similarity was computed with the trigram measure, adding two spaces to the beginning of each word. The actual similarity coefficient is the Dice coefficient. All of the high similarity pairs have similarity greater than 0.65; low similarity pairs have similarity equal to 0.

$$\text{Similarity} = 2C / (A + B)$$

A = number of trigrams in word 1

B = number of trigrams in word 2

C = number of common trigrams

Pairs were generated at random from a list of non-hyphenated, purely alphabetic, one-word names taken from the 3rd quarter 1998 update of the United States Pharmacopeia's *Drug Information for Health Care Professionals, Vol. 1* (U. S. Pharmacopeia, 1998a).

Table 1. High Similarity Pairs

Number	Trigram Sim.	Name #1	Name #2
1.	0.71	Normix	Normiflo
2.	0.67	Cortifoam	Cortifair
3.	0.67	Epival	Epivir
4.	0.83	Salac	Salacid
5.	0.73	Colax	Colace
6.	0.73	Secalciferol	Calciferol
7.	0.74	Tribavirin	Ribavirin
8.	0.67	Isocet	Isocal
9.	0.87	Diiodohydroxyquinoline	Diiodohydroxyquin
10.	0.67	Uritab	Uritin

Number	Trigram Sim.	Name #1	Name #2
11.	0.67	Uritin	Uritab
12.	0.67	Genamin	Genac
13.	0.67	Niacin	Niacor
14.	0.67	Cortenema	Cortef
15.	0.67	Hydromine	Hydrosine
16.	0.71	Ampicin	Ampicillin
17.	0.67	Androcur	Android
18.	0.67	Cuprimine	Cuprid
19.	0.67	Dopamet	Dopar
20.	0.67	Hydromycin	Hydromet
21.	0.96	Tetracycline	Tetracyclines
22.	0.67	Cortaid	Cortacet
23.	0.73	Pericyazine	Periciazine
24.	0.67	Levofloxacin	Ofloxacin
25.	0.70	Aprobarbital	Amobarbital
26.	0.71	Triad	Triadapin
27.	0.67	Methylergometrine	Ergometrine
28.	0.67	Keflex	Keflin
29.	0.67	Lidemol	Lidex
30.	0.73	Velban	Velbe
31.	0.87	Duretic	Dureticyl
32.	0.75	Hydromox	Hydromet
33.	0.67	Geridium	Eridium
34.	0.67	Aspirin	Aspirtab
35.	0.71	Probalan	Probar
36.	0.73	Erythrocin	Erythromycin
37.	0.71	Phendimet	Phendiet
38.	0.75	Robaxin	Robaxisal
39.	0.75	Pseudogest	Pseudo
40.	0.67	Calciparine	Calcidrine
41.	0.71	Diastat	Diastix
42.	0.95	Glucantime	Glucantim
43.	0.67	Chlorphenamine	Dichlorphenamide
44.	0.67	Hycomine	Hycomed
45.	0.67	Niacin	Niacor
46.	0.79	Sulphamethoxazole	Sulfamethoxazole
47.	0.71	Cortaid	Cortate
48.	0.71	Permax	Permapen
49.	0.67	Tetracyn	Tetracyclines
50.	0.71	Nephropure	Nephron
51.	0.73	Doxine	Doxil

Number	Trigram Sim.	Name #1	Name #2
52.	0.71	Trilafon	Trilax
53.	0.77	Suprane	Suprax
54.	0.71	Ultram	Ultracal

Table 2. Low similarity pairs

Number	Trigram Sim	Name #1	Name #2
1.	0.00	Doxycin	Methocel
2.	0.00	Megagen	Alrex
3.	0.00	Lopurin	Carisoprodol
4.	0.00	Repan	Tylosterone
5.	0.00	Vitaneed	Peptamen
6.	0.00	Ambien	Obalan
7.	0.00	Legalon	Chronulac
8.	0.00	Viprynum	Quelicin
9.	0.00	Anexate	Doktors
10.	0.00	Trilostane	Hemocitrate
11.	0.00	Tazocin	Ibuprin
12.	0.00	Fortovase	Cardiolite
13.	0.00	Raudixin	Vivol
14.	0.00	Tolbutamide	Phenobarbital
15.	0.00	Norflex	Misoprostol
16.	0.00	Lithizine	Metrodin
17.	0.00	Raxar	Larodopa
18.	0.00	Orazinc	Calan
19.	0.00	Genaphed	Deficol
20.	0.00	Revex	Klorvess
21.	0.00	Liquipake	Urimed
22.	0.00	Resyl	Naturalyte
23.	0.00	Aldara	Doryx
24.	0.00	Inspire	Natacyn
25.	0.00	Ovrette	Proscar
26.	0.00	Amoxicillin	Profenal
27.	0.00	Rofact	Sedapap
28.	0.00	Kasof	Heptalac
29.	0.00	Panadol	Erwinase
30.	0.00	Drixtab	Postacne
31.	0.00	Iodotope	Primidone

Number	Trigram Sim	Name #1	Name #2
32.	0.00	Betapace	Normodyne
33.	0.00	Restoril	Amfepramone
34.	0.00	Peridex	Sulfatrim
35.	0.00	Alphacaine	Cholecystokinin
36.	0.00	Cefizox	Zephrex
37.	0.00	Diphenhist	AeroBid
38.	0.00	Proctocort	Lignocaine
39.	0.00	Inulin	Kindercal
40.	0.00	Skelid	Elmiron
41.	0.00	Senokot	Endocet
42.	0.00	Tribenzagan	Peptamen
43.	0.00	Decadrol	Florone
44.	0.00	Itraconazole	Magnolax
45.	0.00	Cytotec	Formulex
46.	0.00	Teslac	Dimelor
47.	0.00	Parvolex	Konsyl
48.	0.00	Sporanox	Enulose
49.	0.00	Cologel	Myleran
50.	0.00	Narcan	Bentylol
51.	0.00	Terconazole	Duralex
