Title: Auditory Perception of Drug Names: Neighborhood Effects

Team: Bruce L. Lambert, Ph.D. (PI), Department of Pharmacy Administration, University of Illinois at Chicago (UIC), Laura Walsh Dickey, Ph.D., Department of Communication Science and Disorders, University of Pittsburgh, William M. Fisher, Ph.D., William M. Fisher Consulting, Inc., Gaithersburg, MD, Robert D. Gibbons, Ph.D., Center for Biomedical Statistics, Department of Psychiatry, UIC, Swu-Jane Lin, Ph.D., Department of Pharmacy Administration, UIC, Paul A. Luce, Ph.D., Department of Psychology, University at Buffalo, Conor T. McLennan, Ph.D., Department of Psychology, Cleveland State University, John W. Senders, Ph.D., John W. Senders Consulting, Inc., Toronto, Ontario, Canada, Clement T. Yu, Ph.D., Department of Computer Science, UIC, David Tcheng, National Center for Supercomputing Applications, Urbana, IL.

Inclusive Dates of Project: June 1, 2003–May 31, 2008

Federal Project Manager: Robert Borotkanics, MPH

Acknowledgment: This research was supported in part the Agency for Healthcare Research and Quality.

Grant Award Number: #1 R01 HS011609

Abstract

Purpose: Measure the impact of similarity, familiarity, prescribing frequency, background noise, and other factors on clinicians' and lay people's abilities to identify spoken drug names. Develop an empirically validated, user-friendly software tool that can be used to screen proposed drug names against databases of existing drug names.

Scope: Drug names that sound alike contribute to the roughly 4 million wrong-drug errors committed annually in outpatient pharmacies, representing 0.13% of all dispensed prescriptions. Errors are made by physicians, nurses, pharmacists, and lay people in all types of settings. The FDA and industry try to prevent errors by screening new drug names against databases of existing names and by conducting safety studies on new names. We sought to validate and improve on these methods.

Methods: We conducted auditory perception experiments on pharmacists, physicians, nurses, and lay people, and we designed and evaluated a drug name search engine.

Results: Accuracy in auditory perception increased as familiarity, prescribing frequency, signal-to-noise ratio, and frequency-weighted neighborhood probability increased. We were able to produce a drug name search engine that performed well in validation tests using expert judgments of confusability as the gold standard.

Key Words: drug name confusion, nomenclature, patient safety, auditory perception, medication errors

Purpose

Health professionals and lay people confuse similar sounding drug names (e.g., *Altenol*®/atenolol, *Celebrex*®/*Cerebyx*®, *Benylin*®/*Ventolin*®, vinblastine/vincristine). One in six medication errors involves name confusion. Drug companies and regulators screen names prior to approval, but the screening process is itself error prone due to an over-reliance on subjective assessments of similarity. Our long-term objective is to minimize the incidence of name confusion errors. Our short-term goal is to develop an empirically validated, user-friendly software tool that can be used to screen proposed drug names against databases of existing drug names. Given a name as input, the software will return a list of existing names ranked in descending order of confusability. Confusability ratings will be based on validated, objective criteria derived from studies of clinicians' and lay persons' auditory perceptual errors. Auditory perception experiments will be based on Luce's Neighborhood Activation Model (NAM). The NAM predicts that errors in auditory perception depend on the intelligibility of the target word as well as the similarity and frequency of words in the target word's perceptual "neighborhood." This prediction is embodied in Luce's Frequency-Weighted Neighborhood Probability Rule (FWNPR). Using NAM as the theoretical framework, we will test three hypotheses:

1. The number of errors in auditory perceptual identification will *increase* as frequency-weighted neighborhood probability *decreases*.
2. The effects of frequency-weighted neighborhood probability on auditory perceptual identification will be the same among adult lay people, physicians, nurses, and pharmacists.

3. A model can be developed that accurately predicts a drug name's probability of confusion in an auditory perceptual identification task. In addition to the overall probability of confusion, the model will also predict which specific names are most likely to be confused with the target name.

To test these hypotheses, we propose studies with the following specific aims:

1. To generate confusion data from pharmacists, physicians, nurses, and adult lay people using a noise-masked auditory perceptual identification task;
2. To use the confusions—in conjunction with computational tools and the theoretical model—to develop and refine a model for predicting confusions;
3. To incorporate the best predictive models into a user-friendly software tool that can be used to support decision making during the drug name approval process.

Scope

The Institute of Medicine (IOM) reports that between 44,000 and 98,000 Americans die each year as the result of medical errors, making such errors the fourth leading cause of death in the US.[1] Experts dispute the number of deaths attributable to medical errors, but most agree that errors are a threat to patient safety.[2,3] Errors involving medication are the most common type.[1,4,5] One in six *reported* errors in the US involves drug names that look or sound alike (e.g., cisplatin/carboplatin, Retrovir®/Ritonavir®, Toradol®/Tapazole®).[6-8] Observational studies of dispensing errors in community pharmacies suggest that the rate of wrong-drug errors is 0.13%. Assuming roughly 4 billion outpatient prescriptions annually in the US, this translates to 5.2 million wrong-drug errors per year. The IOM recommended that "the Food and Drug Administration (FDA) should…require pharmaceutical companies to test (using FDA-approved methods) proposed drug names to identify and remedy potential sound-alike and look-alike confusion with existing drug names."[1] The federal Quality Interagency Coordination Task Force committed within 1 year to "develop additional standards for proprietary drug names to avoid name confusion" (p. 26).[9] The FDA recently has issued a pilot paper asking for input into a new process for reviewing and approving new drug names.[10] In 2007, the US Pharmacopeia published a review of hospital-based medication errors focusing exclusively on name confusions. Thus, the development of methods for minimizing drug name confusions addresses a significant public health problem.

STUDY 1a: RELATIONSHIP BETWEEN OBJECTIVE PRESCRIBING FREQUENCY AND SUBJECTIVE FAMILIARITY WITH DRUG NAMES

Efforts to predict and prevent drug name confusions have focused primarily on similarity—the idea being that, the more similar two names are, the more likely they are to be confused. However, research in psycholinguistics suggests that familiarity with a word may play an even larger role than similarity in determining which drugs are likely to be confused. For example, more familiar words are recognized more quickly and more accurately than less familiar words. A familiar word with several similar "neighbors" is less likely to be misperceived than an unfamiliar word with several neighbors.

Recognizing the importance of familiarity to our work on drug name confusion, we set about to develop measures of drug name familiarity that could be used in our future research. We used the psycholinguistic literature as our guide. In psycholinguistics, familiarity is rarely measured directly. Instead, objective word frequency counts (i.e., the number of times a word appears in some large collection of text) are used as proxy measures of familiarity. This is a valid approach because, for ordinary English words, the correlation between objective frequency and subjective familiarity is normally quite high (> 0.9). In the context of drug names, we concluded that objective data on prescribing frequency might be used as a proxy measure for people's subjective familiarity. But the validity of this assumption was untested. The correlation between prescribing frequency and subjective familiarity was not known. Therefore, we designed a study to quantify the magnitude of the correlation between objective measures of prescribing frequency and subjective measures of familiarity with a drug name. The study was guided by the following research questions:

1. What is the relationship between objective frequency and subjective familiarity?
2. How does the source of frequency data affect the relationship between frequency and familiarity?

3. How does the relationship between frequency and familiarity change across occupations?
4. Which frequency database best predicts subjective familiarity for each occupation?

Methods

Design

A cross-sectional, questionnaire-based survey was designed to assess subjective familiarity with drug names. The protocol was approved by the UIC IRB under exemption category 2.

Participants

Fifty-four pharmacists, 32 physicians, 30 nurses, and 32 lay people completed the questionnaire. All the healthcare professionals were licensed and currently practiced in a Midwestern academic medical center. Thirty-two lay people were recruited from among the employees, patients, and patient companions in the outpatient center of the medical center. Each participant was paid $50 to complete the survey.

Materials

Drug names and their prescribing frequencies were taken from five national sample databases. Two freely available datasets, NAMCS and HAMCS, and three commercial datasets were used in the study. The commercial datasets are the IMS National Prescription Audit Plus (outpatient), the Solucient Hospital Drug Utilization Database (inpatient), and the Solucient Claims Data Warehouse (outpatient). Both NAMCS and HAMCS are national probability sample surveys conducted annually by National Center for Health Statistics (NCHS). In the current study, NAMCS and HAMCS data collected between 1996 and 2000 were used. For IMS, data from December 1999 to June 2003 were included. Solucient inpatient data collected from July 1999 to Jun 2002 and Solucient outpatient data collected from January 2000 to December 2002 were used.

Five brand and five generic names were randomly selected from five frequency quintiles of each database. These 250 drug names were randomly ordered in the familiarity questionnaire. Due to the sampling technique, 11 drug names were selected twice, and one drug name was selected three times.

Data Collection

Participants rated their familiarity with each of 250 drug names on a 5-point semantic differential scale ranging from 1 (i.e., not at all familiar) to 5 (i.e., extremely familiar). The questionnaire also collected demographic data about age, gender, race, ethnicity, professional degree, practice context, specialty, and years of experience.

Analysis Plan

The subjective familiarity of each drug name was transformed into a continuous 0-1 scale. A two-step transformation was conducted for the drug frequency. In the first step, each drug frequency was converted into natural logarithm units. The second step transformed the log frequency into a continuous 0-1 scale using the minimum and maximum frequencies of the 50 names in the same database. To model the subjective familiarity and objective frequency, mixed effect regression models were built using the SAS PROC MIXED procedure and MIXOR.

In SAS PROC MIXED, the variable of the interest was the absolute difference between transformed log familiarity and transformed frequency. The database that best predicted the subjective familiarity would have the lowest absolute difference. The dependent variable was the absolute difference between transformed log frequency and transformed subjective familiarity. The independent variables were frequency source (i.e., which database), occupation, gender, ethnicity, and an interaction term between frequency source and occupation. Each independent variable was dummy coded using NAMCS, pharmacists, male, and white ethnicity as the reference group. A random subject effect regression model was fitted using PROC MIXED. The random subject effect allowed subjects to vary at their average familiarity with the drug names.

MIXOR is a statistical tool designed for mixed-effects logistic regression modeling of dichotomous and ordinal data. In MIXOR, the dependent variable was the dichotomized familiarity score. Here, we re-coded 1 and 2 on the original semantic differential scale as unfamiliar (coded as 0) and 3,4, 5 as familiar (coded as 1). We were aware that the cutoff point was somewhat arbitrary. The independent variables were the same as above. Two mixed effect models were tested. The first one included only random subject effect. In the second model, both subject and frequency were considered random effects.

Results

Descriptive statistics of familiarity and frequency

A total of 148 subjects participated in the study, including 98 women (66%) and 72 Caucasians (49%). The average age was 37.7 years old (SD 10.7). Among the four groups of subjects, pharmacists had the highest familiarity with the drug names. The average transformed familiarity score of the 250 drug names was 0.61 for the pharmacist group, 0.47 for the physician group, 0.38 for the nurse group, and 0.10 for the lay people ($p<0.01$ between physician and nurse, $p<0.001$ for all the other paired-comparisons based on ANOVA and Tukey's post hoc test). The mean of the transformed frequency ranged from 0.41 for IMS to 0.52 for ED, but there was no significant difference ($p>0.05$) between any of the five databases.

Relationship between objective frequency and subjective familiarity

The overall correlation between transformed familiarity and frequency was 0.38 ($p<0.001$). When stratified by database, the highest correlation was found in drug names from the NAMCS ($r=0.55$, $p<0.001$), and the lowest was from IMS ($r=0.25$, $p>0.05$). When stratified by occupation, the pharmacist group had the highest correction between familiarly and frequency ($r=0.41$, $p<0.001$), and the lay people group had the lowest ($r=0.11$, $p>0.05$).

Familiarity, frequency, and source of frequency database

Across the 148 subjects, the discrepancy score between familiarity and frequency was largest in drug names selected from IMS database, followed by ED and Solucient outpatient databases, and was smallest for Solucient inpatient and NAMCS databases. A large discrepancy score indicated a greater discrepancy between objective frequency and subjective familiarity. The interpretation of the findings is that, compared with other databases, the frequency estimates from the NAMCS database more closely matched self-rated familiarity. Similarly, the discrepancy score between subjective familiarity and prescription frequency was the largest in drug names selected from IMS; that is, frequency from the IMS database was not matched closely with self-rated familiarity.

Familiarity, frequency, and occupation

Across the 250 drug names from the five databases, the discrepancy score between familiarity and frequency was largest among lay people, followed by nurses and then pharmacists and physicians, who had similar absolute difference scores.

Random effects logistic regression model to predict subjective familiarity

The dichotomous familiarity score (familiar vs. unfamiliar) was used as dependent variable in the logistic regression model with two random effects (subjects and frequency). The purpose of the model was to predict the subjective familiarity ratings based on objective data about demographic group and prescribing frequency.

Random effects

The estimates of the model indicated both the random subject effect and random frequency effect were significant ($p<0.001$), which indicates the significant heterogeneities between subjects in the familiarity with drug names (intercept) and in the relationship between familiarity and frequency (slope). Moreover, there was a significant negative correlation ($p<0.001$) between the two random effects (intercept and slope), which means that the frequency had a larger effect on subjects with lower familiarity score or that frequency was a better predictor of familiarity in people with low familiarity than those with high familiarity with drug names.

Fixed effects

None of the demographic variables was significant in the fixed effect model. Pharmacists had significantly higher familiarity with drug names than physicians and nurses did. Lay people, as expected, had the lowest familiarity. Frequency had a significant effect on familiarity ($p<0.001$). No matter which database a drug name and frequency derived from, frequency had a positive relationship with familiarity. The higher the drug frequency, the higher the familiarity rated by the subjects. In terms of database, drug frequency derived from NAMCS and ED had the highest correlation with familiarity, which means that the drug frequency from the two databases can predict familiarity better than other databases can.

Random effects linear regression model

The absolute distance between transformed familiarity and frequency, a continuous variable, was used as dependent variable in a linear regression model with subjects as the random effect. The purpose of the model was to identify the frequency database that could best be used to predict the familiarity for members of each subject group. In constructing the model, the "best" database was defined as the one that minimized the absolute difference between transformed frequency and familiarity.

Random effects

Similar to the logistic model, the estimates from the linear regression model indicated that both the random subject effect and random frequency effect were significant ($p<0.001$), which means there were significant heterogeneities between subjects in the familiarity with drug names (intercept) and in the relationship between familiarity and frequency (slope).

Fixed effects

Summing up the coefficients of occupation and database main effects and the respective interaction terms, we found that, for pharmacists, the Solucient inpatient source was the best frequency database to predict their familiarity with the drug names, followed by the two equally good databases of NAMCS and ED; IMS was the worst. A similar pattern was observed for the nurse and physician groups, for whom the best database was NAMCS and the worst was IMS. For lay people, NAMCS was the best database and Solucient inpatient was the worst when used to predict the familiarity with the drug names. None of the demographic variables were significant.

STUDY 1b: FREQUENCY ESTIMATES FROM PRESCRIPTION DRUG DATASETS

We conducted a study to compare the frequency estimates given by several different commercially and freely available data sets. We found substantial differences in the definitions, sampling frames, nomenclature, estimates, update frequency, and cost of the databases. The results of this study were published in 2006.[11] In the interest of space, the details are not repeated here.

STUDY 2: EFFECT OF NOISE, SIMILARITY, AND FAMILIARITY ON AUDITORY PERCEPTION OF DRUG NAMES BY PHARMACISTS, PHYSICIANS, NURSES, AND LAY PEOPLE

Methods

Design

This experiment used a cross-sectional, within- and between-subjects design to examine the effect of background noise, prescribing frequency, subjective familiarity, and similarity neighborhoods on a physician's ability to correctly identify spoken drug names. Each subject heard and attempted to identify the same set of 197 spoken drug names.

Participants

All participants were paid $100 for their participation. The experiment was reviewed and approved as exempt by the IRB at the University of Illinois at Chicago and at Cleveland State University.

Pharmacists. Sixty-seven licensed, practicing pharmacists were recruited, but five were eventually excluded due to equipment malfunction or non-native language accents, leaving 62 participants for the main analysis. Thirty-six were women (58%), and 26 were men (42%). The majority identified their race as White (80.7%), five (8%) identified as African American, three (4.8%) identified as Asian, and four (6.5%) identified as other. Three (5%) identified their ethnicity as Hispanic or Latino. Participants practiced in retail pharmacy (33, 53.2%), clinics (13, 21%), hospitals (7, 11.3%), and other contexts (9, 14.5%). Participants were recruited from among the attendees at the 2005 annual meeting of the American Pharmacists Association in Orlando, FL.

Physicians. Seventy-six licensed, practicing, family physicians were recruited, but two were eventually excluded due to equipment malfunction or non-native language accents, leaving 74 participants for the main analysis. Fifty were men (68%), and 24 were women (32%). The majority identified their race as White (54, 74%), two (3%) identified as African American, 16 (22%) identified as Asian, one (1%) identified as other, and one identified as unknown. Participants practiced in clinics (61, 82%), hospitals (6, 8%), and other contexts (7,

9%). Two (3%) identified their ethnicity as Hispanic or Latino. Participants were recruited from among the attendees at the 2005 annual meeting of the American Association of Family Physicians in San Francisco, CA.

Nurses. Seventy-four licensed, practicing medical-surgical nurses were recruited, but four were excluded due to equipment malfunction or non-native language accents, leaving 70 participants for the main analysis. Seventy were women (99%). The majority identified their race as White (66, 96,%), one (1%) identified as Asian, and three (3%) identified as more than one race, other, and unknown. Three (4%) identified their ethnicity as Hispanic or Latino. Participants practiced in hospitals (69, 99%) and other contexts (1, 1%). Participants were recruited from among the attendees at the 2005 annual meeting of the Academy of Medical Surgical Nurses in Las Vegas, NV.

Lay people. Sixty lay people were recruited from the community surrounding Cleveland State University. Funds were available to transcribe only 43 of the 60 sets of responses, so only 43 participants were included in the main analysis. Thirty-two (74%) were women, and 11 (26%) were men. The majority identified their race as White (24, 57%), 11 (26%) identified as Black or African American, four (10%) identified as more than one race, three (7%) identified as Native Hawaiian or Other Pacific Islander, and one left the question blank. None identified as Hispanic or Latino.

Stimulus Materials

As stimuli for the main auditory perception experiment, we selected 99 brand and 99 generic drug names. Names and prescribing frequency information were drawn from four sources: (1) the National Ambulatory Medical Care Survey (NAMCS), (2) the National Hospital Ambulatory Medical Care Survey (HAMCS), (3) the IMS-Health National Prescription Drug Audit (NPDA), and (4) the Solucient Hospital Drug Utilization Database. NAMCS data reflect a nationally representative sample of US office-based physicians involved in direct patient care, excluding anesthesiology, pathology, and radiology.[12] HAMCS data reflect hospital emergency departments and outpatient clinics in a nationally representative sample of US hospitals.[13] Solucient's Hospital Drug Utilization Database (HDUD) contains data from nearly 150 acute-care hospitals in the United States, covering more than 1.9 million discharges. The NPDA measures the retail outflow of new and refill prescriptions from a random sample of 20,000 retail pharmacies out of 34,000 stores in IMS's reporting network. Roughly 76% of all prescriptions dispensed in the US move through the retail channel (independent, chain, and food store pharmacies).[14] Closed HMO pharmacies, dispensing physicians, hospitals, clinic pharmacies, and home health care pharmacies are excluded from the NPDA. Sampled stores are stratified by region, type, and size, and national projections are based on the stratified sample. Additional details about these prescription databases can be found elsewhere.[15]

Frequency-weighted neighborhood probability was computed for each name, according to a procedure described elsewhere.[16] Names were stratified by prescribing frequency by taking 10 brand and 10 generic from each decile of FWNP. One brand name and one generic name were removed to make the total set evenly divisible into three signal-to-noise conditions. The names were then digitally recorded in a professional recording studio. They were spoken by a woman who was a trained phonetician and an experienced voice-over actor. Pronunciations were based on the phonemic transcriptions from experienced clinicians collected for a different project. Each recorded name was then edited into a separate AIFF audio file. Because our perceptual identification experiments were intended to mimic the bandwidth limitations of telephone audio, frequencies below 300 Hz and above 3000 Hz were digitally filtered out.[17] One name was recorded incorrectly and was dropped. A list of the 197 stimulus names, along with their reference phonetic transcriptions and frequency-weighted neighborhood probabilities, can be obtained from the first author.

Stimulus degradation. Drug names were played back against a background of standard 20-speaker babble (obtained from Auditec of St. Louis). The noise was played at a mean amplitude of 65 dB and was not bandwidth limited. The stimuli were played at either 63 dB, 68 dB, or 73 dB, resulting in three signal-to-noise conditions of -2 dB, +3 dB, and +8 dB, respectively.

Experimental Procedure

To recruit participants, invitation letters were mailed to registered attendees in advance, and individuals were approached at the convention center and asked (a) if they were currently a licensed, practicing physician

in the US and (b) if they were interested in participating in a study of drug name confusion. Those who agreed were directed to a meeting room where the experiment was being conducted. Participants began by reading a consent form and completing a demographic questionnaire. They then took a pure tone hearing threshold test to screen for anything worse than normal, age-related hearing loss (i.e., pure tone thresholds of 50 dB or lower were accepted). Participants were then seated at a Macintosh PowerBook computer and fitted with a pair of headphones with an attached headset microphone (Beyerdynamic BT190). The participant then read the instructions. Playback of the 20-speaker babble was initiated, and this noise continued throughout the duration of the experiment. The PsyScope experiment program was used to run the main experiment. The task began with 21 practice trials and then continued with 198 trials comprising the main experiment (one of the 198 names was subsequently dropped from the analysis due to an error in recording). On each trial, the participant was asked to repeat back the name they thought they heard. Spoken responses were captured through the laptop's built in microphone and digitally recorded on the computer. Reaction times (i.e., time between end of word playback and onset of response) were triggered through the headset microphone and were also captured by PsyScope but are not analyzed in this paper. After completing the main experiment, participants moved to a different computer, put on a headset mic, and then pronounced the 198 experimental names as they were visually presented on a computer screen. After reading and pronouncing each name, participants rated their subjective familiarity with the name on a 7-point semantic differential scale (extremely familiar to extremely unfamiliar). This final part of the task allowed us to measure pronunciation variation and subjective familiarity.

Scoring Spoken Responses

All spoken responses were transcribed into the ARPAbet[18] phonetic alphabet by a experienced phonetician (e.g., Zyvox was transcribed as /Z '1AY V '2AA K S/). Responses were scored as correct or incorrect by comparing the transcribed responses to the reference transcriptions for each of the 197 stimulus names. However, natural variation in pronunciation made verbatim matching to the reference transcription too strict a criterion. We developed additional procedures to capture legitimate pronunciation variants. The first was a computer program that applied generally accepted rules for pronunciation variation to the reference pronunciations. For example, in fluent speech, unstressed vowel sounds are reduced to the schwa sound. Responses were scored as correct if they could be produced by applying the variation rules to the reference pronunciations. Even after applying these rules, there still appeared to be legitimate variants that were being scored as incorrect. Thus, our phonetician examined all incorrectly scored responses, identified those which she deemed legitimate variants, and provided linguistic justification for each case (e.g., what phonetic transformation would produce this legitimate variant). In the end, a response was scored as correct if it matched the reference pronunciation exactly, if it could be automatically generated as a rule-generated variant, or if it was recognized by an expert in phonetics as a legitimate variant.

Analysis Plan

The goal of our analysis was to determine the main effects of FWNP, noise, frequency, and familiarity on accuracy in auditory perception. Statistical analyses were done using SAS version 9.1 and SuperMix, a system for doing mixed-effects logistic regression modeling of dichotomous and ordinal data. The mixed-effects logistic regression model accommodates nesting of experimental conditions within subjects for a binary outcome and a mixture of discrete and continuous covariates that can vary at the level of either the subject or the experimental condition.

The dichotomous dependent variable was correct identification. Words correctly identified were scored as 1, and words incorrectly identified were scored as 0. The independent variables were (1) the FWNP score, a continuous variable on the interval 0 to 1, reflecting the predicted probability of identification; (2) signal-to-noise ratio; an ordinal variable with three levels (-2 dB, +3 dB, +8 dB); (3) familiarity, an ordinal variable reflecting a participant's subjective familiarity with the name (ranging from 1 to 7); (4) prescribing frequency, a continuous variable representing the log (base 10) of the maximum frequency from our multiple sources of prescribing frequency data; (5) phoneme frequency, a continuous variable representing the frequency of a given phoneme in a given position in our database of drug names; and (6) biphone frequency, a continuous variable representing the frequency of a two-phoneme sequence in a given position in our database of drug names.[19]

The control variables were (1) participant demographics, including age, gender, race, practice context, and years of experience; (2) pure tone threshold, eight continuous variables reflecting the sensitivity of a participant's hearing in each ear at 500 Hz, 1 kHz, 2 kHz, and 3 kHz; (3) length, and ordinal variable reflecting the number of phonemes in the word; and (4) trial, an ordinal variable representing the sequential position of a given response within the set of 198 responses. A small number of missing values for familiarity scores was replaced with the means for those names from other subjects.

The first step was to identify the correct scale for each independent and control variable. We did this by separately plotting the log odds of error as a function of each independent or control variable. If these plots were linear, terms were entered as simple linear terms. If the plot revealed an obvious nonlinearity, we selected a scale to fit the nonlinear form of the function.[20,21] In this case, we primarily considered quadratic terms. Having identified the appropriate scale for each independent and control variable, we used Kleinbaum's method of backward elimination to decide which variables to include in the final model.[22] According to this method, the analyst begins with a full model and then proceeds to eliminate as many terms as possible, using likelihood ratio tests to decide which terms contribute significantly to the model's fit.

The final step in our modeling strategy was to assess goodness of fit. We assessed fit in several ways. First, we examined the fit between observed and predicted accuracy rates for each of the 197 stimulus names. Observed rates were taken directly from the data. For each subject, we used the parameter estimates from the fitted model and covariate values from the data to generate a predicted proportion of correct identification for each name. Plots of predicted versus observed frequencies are provided. All statistical tests used an alpha of 0.05.

Results

Physicians

Each of the 74 participants responded to 197 stimuli, producing 14,578 total responses. Mean accuracy on the perception task was 34.45% (s.d.=5.52, range=19.8%–47.2%). This may seem unrealistically low compared with the rates observed in real-world practice settings. Because one aim of our investigation was to generate a large number of errors for subsequent analysis, we intentionally inflated the overall error rate by increasing the amount of degradation and by including many low-frequency names with which many participants were not familiar. As a result, we now have a large database of actual errors that can be examined in an effort to learn more about the way similarity is represented in the minds of family physicians. Table 1 presents descriptive statistics for the continuous independent variables.

Table 1. Descriptive statistics for independent variables for correct and incorrect responses

Variable	Mean (s.d.)		Min	Max	Z	P
	Incorrect (n=9555)	Correct (n=5023)				
R500 Hz	19.64 (6.26)	19.11 (6.34)	0	35	-3.16	0.002
L500 Hz	18.18 (13.68)	17.78 (12.54)	5	120	-1.06	0.29
R1000 Hz	14.01 (6.45)	13.55 (6.28)	5	35	-2.64	0.01
L1000 Hz	12.44 (7.13)	12.22 (7.12)	0	35	-1.09	0.27
R2000 Hz	10.31 (8.97)	9.60 (8.20)	-5	45	-3.14	0.002
L2000 Hz	11.15 (9.53)	10.55 (8.92)	-10	40	-3.67	0.0002
R3000 Hz	11.63 (11.01)	11.21 (10.77)	-10	45	-1.42	0.16
L3000 Hz	13.17 (12.81)	12.80 (12.25)	-10	50	-1.10	0.27
Age	44.40 (11.07)	44.11 (10.59)	28	69	-1.53	0.13
Experience	15.00 (10.48)	14.67 (10.19)	1	40	-1.09	0.27
Familiarity	2.61 (2.37)	4.43 (2.72)	1	7	39.01	0.0000
FWNP	0.52 (0.34)	0.61 (0.36)	0	1	13.45	0.0000
Rx Frequency	3.13 (1.64)	3.95 (1.77)	-0.23	7.82	26.63	0.0000
No. of Phonemes	8.04 (2.26)	8.29 (2.84)	4	18	5.94	0.0000
Trial	121.64 (57.02)	123.94 (57.44)	24	221	2.31	0.02

Note. Mixed-effects logistic regression models were built with Supermix with only an intercept and one independent variable in the model. Z scores come from the Wald test on the parameter estimate for the variable in question. See text for details.

Table 2 gives the proportion of correct and incorrect responses as a function of the nominal predictors as well as results of tests of the bivariate association between predictors and the outcome.

Table 2. Percent correct and incorrect responses by nominal covariates

Variable	%		Chi-Square	P	Z	P
	Correct	Incorrect				
S/N Ratio			1238.94	<0.0001	34.02	<0.0000
-2 dB	16.32	83.68				
+3 dB	37.12	62.88				
+8 dB	49.94	50.06				
Gender			1.01	0.31	-0.62	0.54
Male	34.73	65.27				
Female	33.88	66.12				
Name Type			126.95	<0.0001		
Brand	30.04	69.96				
Generic	38.91	61.09				
Race			23.90	<0.0001	-3.22	0.001
White	35.63	64.37				
Non-White	31.29	68.71				
Ethnicity			0.09	0.77	-0.17	0.86
Hispanic	33.76	66.24				
Non-Hispanic	34.48	65.52				
Context			4.91	0.09		
Hospital	31.64	68.36			-1.31	0.19
Clinic	34.79	65.21			1.14	0.25
Other	33.94	66.06			-0.26	0.79

Note. Chi-square values came from standard 2 x N tables. Z scores come from Wald tests on parameter estimates generated by Supermix mixed-effects regression models.

The upper right panel of Figure 1 shows the relationship between familiarity and accuracy for physicians. The influence of signal-to-noise ratio is depicted Figure 2.

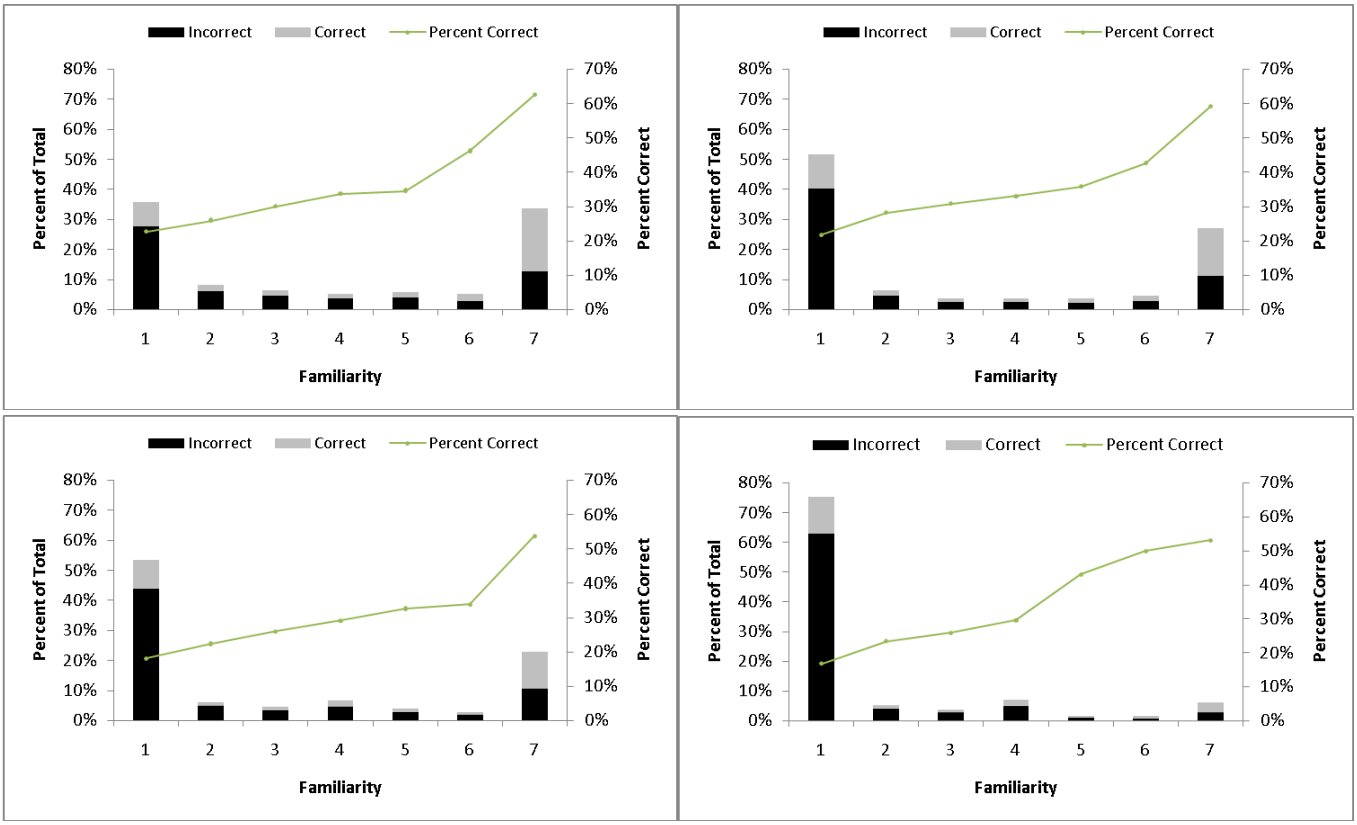


Figure 1. Auditory perception accuracy as a function of familiarity. Clockwise from top left: Pharmacists, physicians, lay people, nurses. For each level of familiarity, the dark bar represents the percent of incorrect responses as a percent of the total number of responses for that participant group. For lay people, data are shown only from subjects in the condition with background noise.

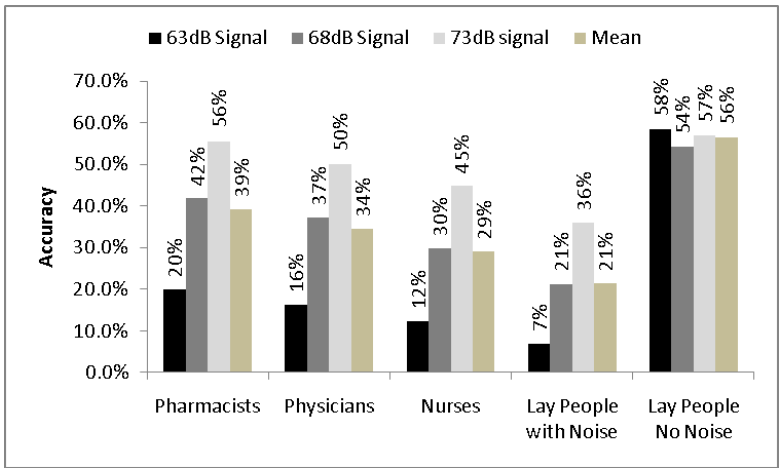


Figure 2. Effect of signal strength on accuracy for pharmacists, physicians, nurses, and lay people. Noise was played at constant 65-dB amplitude for all subject groups, except where noted for lay people.

Table 3 gives the parameter estimates and associated probabilities for the final random effect logistic regression model (i.e., the model arrived at after following Kleinbaum's backward elimination procedure). Background noise decreased accuracy in physicians' auditory perception of drug names (signal-to-noise ratio, b=0.20, OR=1.23, 95%CI=1.20-1.25, p<0.0001). Drug names with many similar neighbors or with commonly prescribed neighbors were perceived less accurately than were names with few neighbors or names whose neighbors were rarely prescribed (FWNP[2], b=1.44, OR=4.21, 95%CI=2.72-6.51, p<0.0001). Familiar drug

names were perceived more accurately than were unfamiliar names (familiarity^3, b=0.02, OR=1.02, 95%CI=1.01-1.03, p<0.0001). Drug names with high national prescribing frequency were perceived more accurately than were names with low national prescribing frequency (frequency, b=0.12, OR=1.13, 95%CI=1.09-1.16, p<0.0001). Brand name drugs were perceived less accurately than generic names were (brand, b=-0.18, OR=0.84, 95%CI=0.77-0.91, p<0.0001). Some of the main effects were contained in significant higher-order interactions (details not shown).

Table 3. Parameter estimates for random effects logistic regression model of accuracy in auditory perception

Parameter	Estimate	Standard Error	P	Odds Ratio	95% Lower	95% Upper
Intercept	-2.26	0.22	0.00	0.10	0.07	0.16
FWNP	-1.56	0.24	0.00	0.21	0.13	0.34
FWNP^2	1.44	0.22	0.00	4.21	2.72	6.51
Rx Frequency	0.12	0.02	0.00	1.13	1.09	1.16
S/N Ratio	0.20	0.01	0.00	1.23	1.20	1.25
Familiarity	0.89	0.21	0.00	2.44	1.62	3.66
Familiarity^2	-0.23	0.06	0.00	0.79	0.70	0.89
Familiarity^3	0.02	0.01	0.00	1.02	1.01	1.03
Phoneme Freq.	-0.28	0.05	0.00	0.75	0.68	0.83
Biphone Freq.	0.50	0.06	0.00	1.66	1.48	1.85
Familiarity x S/N Ratio	0.00	0.00	0.01	1.00	0.99	1.00
Familiarity x Phoneme Freq.	0.04	0.01	0.00	1.04	1.01	1.06
Familiarity x Biphone Freq.	-0.05	0.01	0.00	0.95	0.93	0.98
Experience	-0.01	0.00	0.00	0.99	0.98	1.00
R500 Hz	-0.02	0.01	0.00	0.98	0.97	0.99
L1000 Hz	0.01	0.01	0.01	1.01	1.00	1.02
R2000 Hz	-0.02	0.00	0.00	0.98	0.97	0.99
Nonwhite	-0.27	0.08	0.00	0.76	0.66	0.89
Hospital	-0.29	0.12	0.01	0.75	0.60	0.94
Trial	0.00	0.00	0.00	1.001	1.00	1.002
Brand	-0.18	0.04	0.00	0.84	0.77	0.91

Note. FWNP=frequency weighted neighborhood probability. Rx Frequency is the log base 10 of the maximum frequency observed across four different prescription databases. S/N ratio is the signal-to-noise ratio. See text for details.

To assess goodness of fit of the model, we compared the observed and predicted percent correct for each of the 197 drug names. The root mean squared error of prediction was 17.52%, and the mean absolute error was 14.65%. Figure 3 shows the scatter plot of predicted versus observed proportion of correct responses for each drug name. There was substantial variability across names (range, 0% to 98.38%). With an R^2=0.48, the model accounted for 48% of the variance in accuracy at the level of the individual name. Several names were never identified correctly (e.g., sutilains, *Kira*, sparfloxacin, and tromethamine), while others were rarely missed (e.g., hydrochlorothiazide, *Zithromax*).

Substitution errors. Physicians produced two types of error in the perception task. The most common type was just a mispronunciation of the stimulus name (90.3%). However, approximately one in 10 errors was what we term a substitution error, when the response corresponded to another real drug name (9.7%). Substitution errors are more interesting, because they correspond to potentially harmful wrong-drug errors. One important question concerns the effect of frequency on substitution errors. Psycholinguistic theory predicts that rare names will be misheard as common names but not vice versa. We tested this hypothesis. Overwhelmingly, substitution errors tended to go in the direction of the more frequently prescribed name. There were 1408 substitution errors overall. In 883 (62.71%) of these errors, the substituted name had a higher prescribing frequency than the stimulus name did. In 433 cases (30.75%), the prescribing frequency of the substituted name was not known, because it was not in our prescribing databases. Considering only the cases when the prescribing frequency of the substituted name was known, errors went in the direction of the more

frequently prescribed drug 90.56% of the time. The mean difference in log prescribing frequency between the substituted name and the target name was 2.27 (s.d.=1.76), meaning the substituted name was, on average, prescribed 186 times more frequently prescribed than the target. For errors that went in the direction of the more frequently prescribed name, the mean difference in log frequency was 2.54. For errors that went in the direction of the lower-frequency name, the mean difference in log prescribing frequency was only -1.40. This suggests that confusions go in the direction of the less common name when the prescribing frequencies are relatively similar. As the difference in prescribing frequency increases, so does the tendency for the errors to go in the direction of the higher-frequency name. Once the difference in log frequency increases above about 3, nearly 100% of the errors are in the direction of the higher-frequency name (see Figure 4).

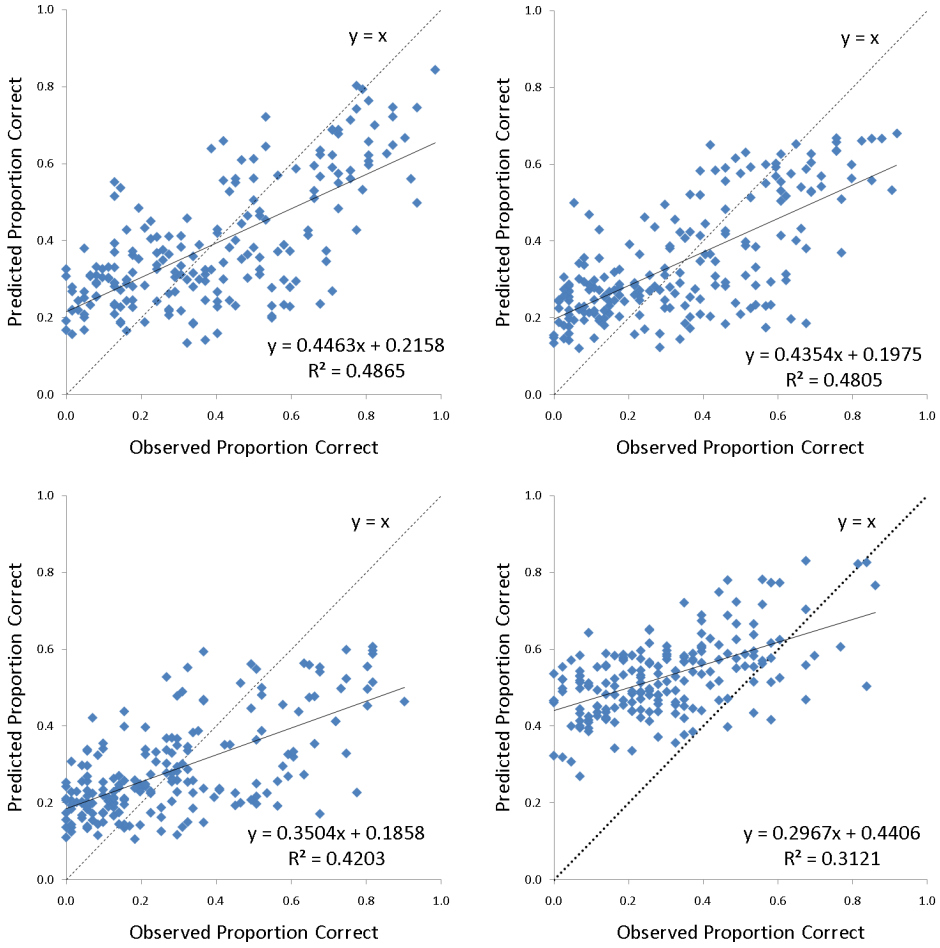


Figure 3. Predicted and observed proportion correct. Clockwise from top left: pharmacists, physicians, lay people, nurses.

Pharmacists, Nurses, and Lay People

Space constraints prevent us from showing the detailed results from pharmacists, nurses, and lay people. The main results are summarized in Figures 1-4 and in Table 4.

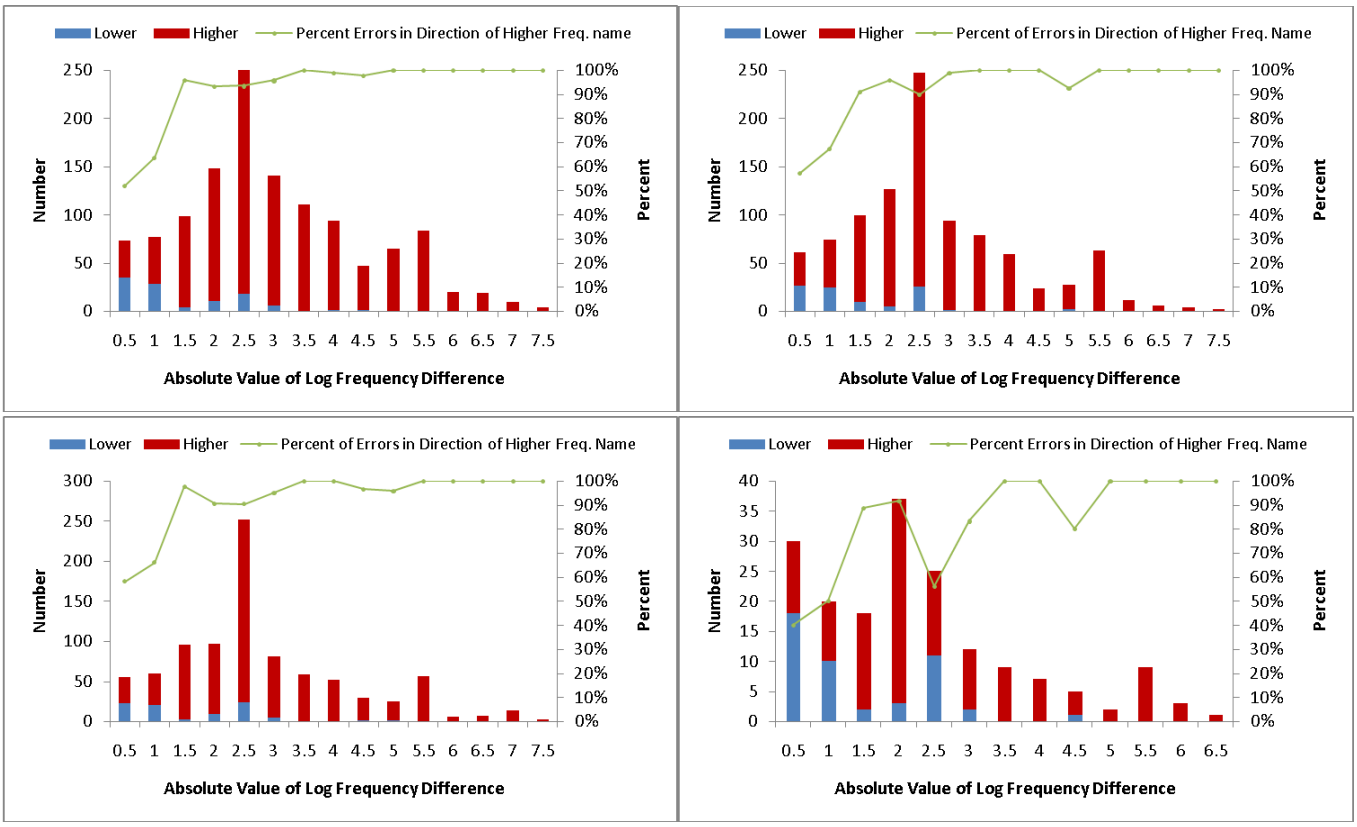


Figure 4. Direction of substitution errors as a function of log frequency difference. Clockwise from top left: Pharmacists, physicians, lay people, nurses. The bottom part of each vertical column represent the number of substitution errors that went in the direction of the lower-frequency name. The top portion of each vertical column represents the number of substitution errors that went in the direction of the higher-frequency name. The line is the percent of errors in the direction of the higher-frequency name.

Table 4. Summary of auditory perception results for pharmacists, physicians, nurses, and lay people

Variable	Pharmacists	Physicians	Nurses	Lay People
Participants	62	74	70	43
Num. Responses	12214	14578	13790	8471
Accuracy				
Mean (s.d.)	39.14 (5.83)	34.46 (5.52)	29.02 (5.15)	29.54 (16.47)
Range	16.75%–49.74%	19.8%–47.2%	13.71%–38.58%	12.18–65.99
Goodness of Fit				
Root mean squared error	18.5%	17.5%	17.8%	28.1%
Mean absolute error	15.5%	14.7%	14.6%	24.8%
R^2 predicted vs. observed	0.49	0.48	0.42	0.31
Substitution Errors				
Total number (%)	1783 (14.6%)	1408 (9.66%)	1268 (9.20)	285
% of errors that were substitution errors	23.99%	14.74%	12.95%	4.78%
No. with known freq.	1266	975	887	178
% in direction of higher-frequency name	1163 (91.9%)	883 (90.6%)	803 (90.5%)	131 (73.6%)
Mean (s.d.) log freq. difference between target and substituted name	2.57 (1.80)	2.27 (1.76)	2.30 (1.83)	1.42 (2.06)

Note. Lay people included 33 subjects with noise and 10 without any noise, so lay results are not strictly comparable to the others.

STUDY 3: EFFECTIVENESS OF A COMPUTER PROGRAM FOR RANKED RETRIEVAL OF DRUG NAMES

Our goal in this part of the project was to build, evaluate, and refine a computerized search engine that would take a queried drug name as input and retrieve a list of other drug names, ranked in descending order of similarity to the queried name. We have produced such a search engine.[23] It is available via the web (http://dolphin.cs.uic.edu:8080/DrugSearch/). Figure 5 is a screenshot of the main query screen for the search engine. Figure 6 is a screenshot of an example search result for the query Zetia.

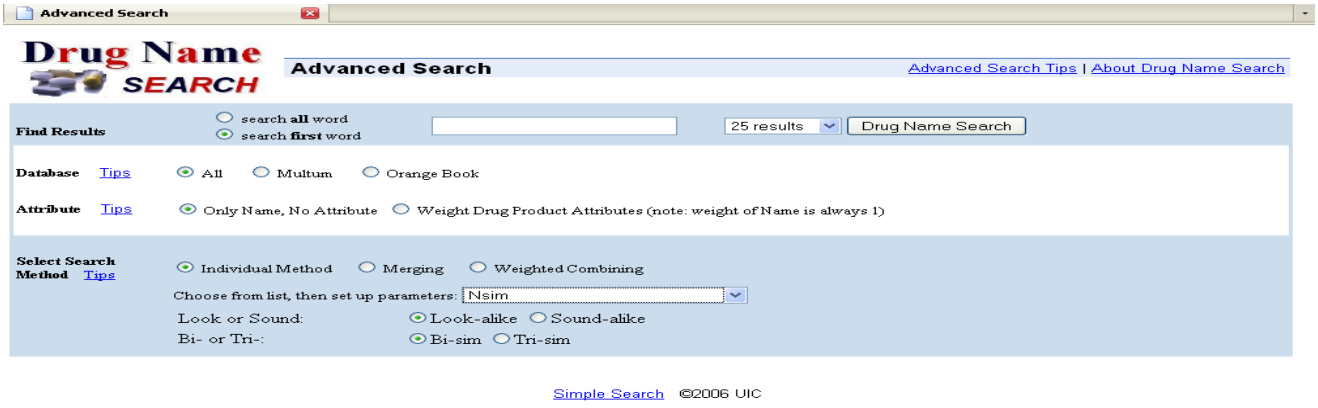
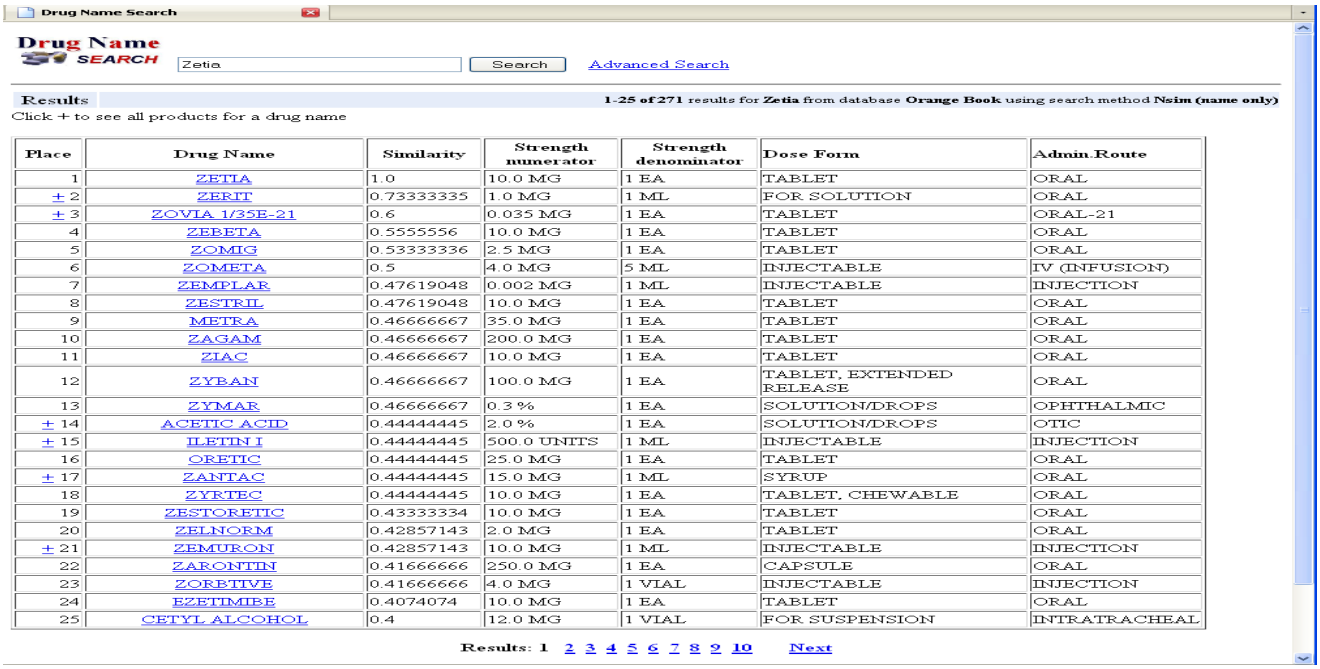


Figure 5. Main query screen from drug name search engine.



Place	Drug Name	Similarity	Strength numerator	Strength denominator	Dose Form	Admin Route
1	ZETIA	1.0	10.0 MG	1 EA	TABLET	ORAL
+ 2	ZERIT	0.73333335	1.0 MG	1 ML	FOR SOLUTION	ORAL
+ 3	ZOVIA 1/35E-21	0.6	0.035 MG	1 EA	TABLET	ORAL-21
4	ZEBETA	0.5555556	10.0 MG	1 EA	TABLET	ORAL
5	ZOMIG	0.53333336	2.5 MG	1 EA	TABLET	ORAL
6	ZOMETA	0.5	4.0 MG	5 ML	INJECTABLE	IV (INFUSION)
7	ZEMPLAR	0.47619048	0.002 MG	1 ML	INJECTABLE	INJECTION
8	ZESTRIL	0.47619048	10.0 MG	1 EA	TABLET	ORAL
9	METRA	0.46666667	35.0 MG	1 EA	TABLET	ORAL
10	ZAGAM	0.46666667	200.0 MG	1 EA	TABLET	ORAL
11	ZIAC	0.46666667	10.0 MG	1 EA	TABLET	ORAL
12	ZYBAN	0.46666667	100.0 MG	1 EA	TABLET, EXTENDED RELEASE	ORAL
13	ZYMAR	0.46666667	0.3 %	1 EA	SOLUTION/DROPS	OPHTHALMIC
+ 14	ACETIC ACID	0.44444445	2.0 %	1 EA	SOLUTION/DROPS	OTIC
+ 15	ILETIN I	0.44444445	500.0 UNITS	1 ML	INJECTABLE	INJECTION
16	ORETIC	0.44444445	25.0 MG	1 EA	TABLET	ORAL
+ 17	ZANTAC	0.44444445	15.0 MG	1 ML	SYRUP	ORAL
18	ZYRTEC	0.44444445	10.0 MG	1 EA	TABLET, CHEWABLE	ORAL
19	ZESTORETIC	0.43333334	10.0 MG	1 EA	TABLET	ORAL
20	ZELNORM	0.42857143	2.0 MG	1 EA	TABLET	ORAL
+ 21	ZEMURON	0.42857143	10.0 MG	1 ML	INJECTABLE	INJECTION
22	ZARONTIN	0.41666666	250.0 MG	1 EA	CAPSULE	ORAL
23	ZORETIVE	0.41666666	4.0 MG	1 VIAL	INJECTABLE	INJECTION
24	EZETIMIBE	0.4074074	10.0 MG	1 EA	TABLET	ORAL
25	CETYL ALCOHOL	0.4	12.0 MG	1 VIAL	FOR SUSPENSION	INTRATRACHEAL

Figure 6. Output screen from drug name search engine.

To evaluate the search engine, we used the method of pooled relevance judgments. This method is described in detail in our previous progress reports and publications.[24,25] We also developed a novel mathematical approach to estimating mean average precision. Space limitations prevent us from presenting the full details of this new estimation method, but the details are available upon request and will be published. Table 5 lists the first-order ranking methods currently available in the search engine. Table 6 gives the ranking of these methods based on pooled relevance judgments and our new estimation method. Java code for all of the individual similarity measures will be made publicly available via the web. Evaluation and refinement are ongoing, with complete results to be reported via a published manuscript.

Table 5. Current methods in the search engine

Method	Remarks
Bi-gram Similarity	Orthographic and phonetic
Tri-gram Similarity	Orthographic and phonetic
N-sim Similarity	Orthographic and phonetic, Bi-gram and Tri-gram
Edit Distance	Orthographic and phonetic
Longest Common Subsequence	Orthographic and phonetic
Ukkonen Key	Orthographic and phonetic
Learning Edit Distance	Phonetic
Bipartite Matching Distance	Orthographic and phonetic
Jaro Distance	Orthographic and phonetic
Jaro Winkler Distance	Orthographic and phonetic
Cohort 4 and 5	Phonetic

Table 6. Estimated mean average precision for various methods

Method	Estimated Mean Avg. Precision
Weighted Sum of Multiple Methods	0.02548
Longest Common Subsequence (phonetic)	0.02443
Longest Common Subsequence (spelling)	0.02436
N-sim Similarity (phonetic)	0.02436
Jaro Distance (phonetic)	0.02424
N-sim Similarity (spelling)	0.02416
Jaro Distance (spelling)	0.02412
Jaro-Winkler Distance (phonetic)	0.02402
Jaro-Winkler Distance (spelling)	0.02378
Editex Distance	0.02346
Bipartite Matching Distance(phonetic)	0.02193
Bipartite Matching Distance(spelling)	0.02187
Learning Edit Distance	0.02063

Other Software. In addition to the search engine, we produced several other potentially useful pieces of software (in C++), all of which we will release on the web at www.uic.edu/~lambertb. The first of these is a program for text-to-phoneme mapping. It takes the orthographic (i.e., spelling) representation of a drug names and converts it to the ARPAbet phonemic representation. This software is useful when phonemic transcriptions of words are not available from another source and need to be generated automatically. Its error rate at the level of individual phonemes was measured at 7.9%.The second is a program for retrieving drug names based on phonemic similarity. This program is integrated into our web-based search engine but is also available as a standalone module. The third is a program for calculating frequency-weighted neighborhood probability based on a target name and a lexicon of names, all with associated frequencies. The fourth is a program to automatically detect acceptable variants of a canonical pronunciation. The system takes as input a reference pronunciation, a test pronunciation (in ARPAbet), and a set of inverse phonological rules. The rules are applied to the test pronunciation, and, if any of the resulting pronunciations matches the reference pronunciation, the test pronunciation is flagged as a legitimate variant. We used this program to score our auditory perception data.

GENERAL DISCUSSION OF SIGNIFICANT PROJECT FINDINGS

The project's main findings fall into three general categories: (a) findings about frequency and familiarity; (b) findings about auditory perception of drug names; and (c) findings about drug name retrieval systems. In this section, we briefly summarize the main findings and comment on their significance and their implications for future research and public policy.

Frequency and Familiarity

Almost all scientific and public policy discussions about drugs begin by examining how frequently a give drug or class of drugs is used. Several different databases are available that provide information about prescribing or dispensing frequency. We examined the most commonly used of these data sources: IMS, Solucient, NAMCS, and HAMCS. We conducted two studies of these databases.

The first sought simply to describe the similarities and differences between the databases and to try to make recommendations about which was suitable for certain purposes.[15] We found that correlations between frequency estimates from public and proprietary drug datasets ranged from small to large. High correlation was found between proprietary datasets. Moderate correlation was found between public and proprietary datasets. Correlations between drug frequency datasets appeared to depend on sampling frame, data collection method, and nomenclature differences. Total prescribing frequency estimated from the five datasets differed significantly. Each of the five prescription drug datasets provides a unique opportunity to investigate various issues related to drug frequency; however, researchers need to be aware of the different definition of drug frequency between the datasets.

The second examined the relationship between objective frequency, as measured by one of the databases mentioned above, and subjective familiarity, as measured by asking participants to report their familiarity with a given drug. Familiarity is important because it is believed to be the key psychological variable that affects visual perception, auditory perception, and short-term memory for drug names. In ordinary language, there are very high correlations between frequency and familiarity, so most psychological experiments use objective (printed) frequency as a proxy for subjective familiarity. We intended to do the same in our experiments on drug names, but, in a specialized technical vocabulary like drug names, the correlation between objective and subjective measures was unknown. We carried out a study to determine the correlation between objective and subjective measures of frequency and familiarity. On the whole, the correlations were statistically significant by much smaller measures than for those observed in ordinary language. We concluded that objective frequencies could serve as an adequate proxy for subjective similarity, especially for pharmacists and other health professionals; based on these findings, we decided to include a direct measure of familiarity in our perceptual experiments. Policy decisions relating to familiarity, when practical, probably should be based on direct measures of familiarity rather than on frequency as a proxy.

Auditory Perception of Drug Names

Findings from the auditory perception experiments identified several key factors that influence the probability of accurately identifying spoken drug names: (1) noise, (2) familiarity and frequency, and (3) similarity.

Noise. Noise had a dramatic effect on accuracy. Accuracy rates increased by a factor of 3 to 5 as signal-to-noise ratios increased from -2 dB to +8 dB. When the noise was completely removed, lay people performed as well as pharmacists did in the quietest condition. The implication is that areas where drugs are dispensed or orders are given or taken should be as quiet as possible. Also, as noise increased, the misperceptions got worse. In other words, substitution errors in the noisiest conditions were farther away from the target name than were errors in the quieter conditions, when distance was measured by phoneme features. The implication is that names that may not seem similar can be confused if there is a sufficient amount of noise.

Familiarity and frequency. Familiarity had substantial effects on accuracy. Across subject groups, the most familiar names were heard roughly three times more accurately than the least familiar names were. The distribution of familiarity scores was essentially bimodal: participants either knew the names or they did not. One implication, which needs to be confirmed by further research, is that increasing people's familiarity with a wider array of drug names, via continuing education for example, might decrease the likelihood of drug name confusions. The other implication has to do with the relationship between similarity and the direction of errors. Previous research and public policy has operated on the assumption that confusion errors are symmetrical, that if they are similar enough, Name A will be misheard as Name B as often as Name B will be misheard as Name A. This assumption is false. Our findings show that substitution errors overwhelmingly go in the direction

of the more frequently prescribed name. This means that policymakers, especially the FDA, must consider frequency (really, familiarity) when estimating the probability of confusion between two drug names and the probability of the resulting harm. For example, for a given proposed name, similarity to high-frequency drugs should be seen as presenting more risk of confusion than similarity to low-frequency drugs.

Similarity. Drugs with more similar neighbors, and especially with frequently prescribed similar neighbors, are more likely to be confused than are drugs with fewer and less frequently prescribed neighbors. Objective measures of similarity neighborhood were statistically significant predictors of accuracy in all our experiments, although the magnitude of their effect was smaller than that observed for frequency, familiarity, and noise. Decisionmakers who evaluate the acceptability of new drug names should consider using these validated measures of similarity as one additional piece of objective evidence when making their decisions.

Drug Name Retrieval Systems

High-quality drug name retrieval systems can be designed using techniques from computer science and computational linguistics. Our experiments validated some of these measures against human performance in an auditory perception task. Our research suggests that no single ranking method will perform as well as a method that intelligently combines multiple ranking methods. Systems such as the one we designed are used by FDA (the so-called POCA system) to analyze proposed new drug names. These systems need continual updating and refinement. By the end of 2008, we will release a test collection of drug names and relevance judgments that can be used to evaluate current and future retrieval systems. When evaluating new drugs for confusability, we believe the output of computerized drug name searches should be used as input to a panel of human experts.

List of Publications, Presentations, and Products

Test collection for evaluating drug name retrieval systems. Department of Pharmacy Administration, University of Illinois at Chicago; 2008. www.uic.edu/~lambertb/r_test_collection.shtml.

Chermak TE, Lambert BL. Label characteristics of primary packages of solid oral dosage forms. *J Am Pharm Assoc.* in press.

Lambert BL, Yu CT, Jia L, Lewis DD, Liu F, Fisher WM. A search engine for drug names. http://dolphin.cs.uic.edu:8080/DrugSearch/. Accessed August, 2008.

Lambert, B.L., Dickey, L.W., Fisher, W.M., Gibbons, R.D., Lin, S.-J., Luce, P.A., McClennan, C.T., Senders, J.W., Yu, C.T. (2008, October). *Nurses' auditory perception of drug names*. Academy of Medical-Surgical Nurses, October 3, 2008. Nashville, TN.

Lambert, B.L., Dickey, L.W., Fisher, W.M., Gibbons, R.D., Lin, S.-J., Luce, P.A., McClennan, C.T., Senders, J.W., Yu, C.T. (2008, September). *Family physicians' auditory perception of drug names: Effect of noise, familiarity and similarity*. American Association of Family Physicians Scientific Assembly, September 17, 2008. San Diego, CA.

Lambert, B.L. (2008). *Recent Developments in the Prevention and Detection of Drug Name Confusion*. In Hicks R.W., Becker S.C., Cousins D.D. (Eds.). *MEDMARX data report. A report on the relationship of drug names and medication errors in response to the Institute of Medicine's call for action (pp. 10-15)*. Rockville, MD: Center for the Advancement of Patient Safety, US Pharmacopeia.

Lambert, B.L. "Name safety evaluation: What does research tell us?" Institute for Safe Medication Practices' National Summit on Managing Drug Name Confusions, Philadelphia, PA, October 9, 2007.

Lambert, B.L., Lin, S.-J., Toh, S., et al. Auditory perception of drug names. *Annual Congress of the National Patient Safety Foundation*. Washington, DC; 2007.

Lambert, B.L., Lin, S.-J., Toh, S., Luce, P.A., McLennan, C., La Vigne, R., Fisher, W.M., Dickey, L.W., & Senders, J.W. (2006, September). *Auditory perception of drug names*. Poster presented at Translational Research at UIC Conference, Chicago, IL, September 22, 2006.

Lambert, B.L., Lin, S.-J., Toh, S., Luce, P.A., McLennan, C., La Vigne, R., Fisher, W.M., Dickey, L.W., & Senders, J.W. (2006, June). *Auditory perception of drug names*. Poster presented at AHRQ Patient Safety and Health Information Technology Conference, Washington, DC, June 5-6, 2006.

Lin S-J, Lambert BL, Tan H, Toh S. Frequency estimates from prescription drug datasets *Pharmacoepidemiology and Drug Safety.* 2006;15(7):512-520.

Lambert, B.L., Lin, S.-J., Toh, S., Luce, P.A., McLennan, C., La Vigne, R., Fisher, W.M., Dickey, L.W., & Senders, J.W. (2006). Frequency and neighborhood effects on auditory perception of drug names in noise. Proceedings of NoiseCon 2005, Institute of Noise Control Engineering, Minneapolis, MN

Database of published drug name confusions. 2005. www.uic.edu/~lambertb.

Lambert BL, Lin S-J, Toh S, et al. Frequency and neighborhood effects on auditory perception of drug names in noise. *Acoustical Society of America Journal.* 2005;118(3):1955.

Lambert, B.L., Lin, S.-J., & Tan, H.K. (2005). Designing safe drug names. *Drug Safety*, 28, 495-512.

Toh, S. (2005). *Analysis of published case reports of drug name confusion*. M.S. Thesis, University of Illinois at Chicago.

Lin, S.-J., Tan, J., Toh, D., & Lambert, B.L. (2005, April) *Frequency estimates from prescription drug datasets*. Poster presented at the Annual Meeting of the American Pharmacists Association, Orlando, FL, April 1, 2005.

Lambert, B.L. “Drug name confusion, medication reconciliation and medication safety,” Illinois Hospital Association Patient Safety Learning Collaborative—Medication Reconciliation, First Learning Session, Lisle, IL, February 16, 2005.

Tan, J., Lin, S.-J., Lambert, B.L., Toh, D. (2004, July). *An evaluation of the quality of the USP/ISMP Medication Error Reporting Program*. Poster presented at the Midwest Pharmacy Administration Conference, West Lafayette, IN, July 29, 2004.

Toh, D., Tan, J., Lambert, B.L., Lin, S.-J. (2004, July). *Assessing the quality of published case reports of look-alike and sound-alike medication errors*. Poster presented at the Midwest Pharmacy Administration Conference, West Lafayette, IN, July 29, 2004.

Lambert, B.L. “Comments on validation strategies and testing methods used to evaluate new drug names,” Public testimony before FDA’s Drug Safety and Risk Management Advisory Committee, Rockville, MD, December 4, 2003.

Lambert, B.L. “Predicting and preventing drug name confusions: Strategies for front line practitioners,” Mercy Resource Management, Inc. Annual Meeting,, Lisle, IL, October 29, 2003.

Lambert, B.L. “Computer assisted decision analysis in drug naming,” Look-Alike/Sound-Alike Health Product Names (LA/SA) Consultative Workshop, Health Canada, Ottawa, CA, October 20-21, 2003.

References

1. Kohn LT, Corrigan J, Donaldson MS, eds. To err is human: building a safer health system. Washington, DC: Institute of Medicine; 2000.
2. McDonald CJ, Weiner M, Hui SL. Deaths due to medical errors are exaggerated in institute of medicine report. Journal of the American Medical Association. 2000;284(1):93-95.
3. Leape LL. Institute of medicine medical error figures are not exaggerated. Journal of the American Medical Association. 2000;284(1):95-97.
4. Cohen MR, ed. Medication errors. Washington, DC: American Pharmaceutical Association; 1999.
5. Aspden P, Wolcott J, Bootman JL, Cronenwett LR. Preventing medication errors. Washington, DC: The National Academies Press; 2007.
6. Davis NM. Drug names that look and sound alike. Hosp Pharm. 1997;32(11):1558-1570.
7. US Pharmacopeia. Similar names. USP DI Update Volume I and II. 1997:4-7.
8. Fraterrigo CC, Frangella GM. The "Z" word. Am J Hosp Pharm. 1994;51(January 15):235.
9. Quality Interagency Coordination Task Force. Doing what counts for patient safety: Federal actions to reduce medical errors and their impact. Washington, DC: Quality Interagency Coordination Task Force; February 2000.

10. US Food and Drug Administration. PDUFA pilot project proprietary name review. http://www.fda.gov/cder/drug/MedErrors/meeting_names.pdf, 2008.
11. Lin S-J, Lambert BL, Tan H, Toh S. Frequency estimates from prescription drug datasets Pharmacoepidemiology and Drug Safety. 2006;15(7):512-520.
12. National Center for Health Statistics. Ambulatory Health Care Data: NAMCS Description. Web page] http://www.cdc.gov/nchs/about/major/ahcd/namcsdes.htm. Accessed May 31, 2001, 2001.
13. National Center for Health Statistics. Ambulatory Health Care Data: NHAMCS Description. Web page] http://www.cdc.gov/nchs/about/major/ahcd/nhamcsds.htm. Accessed May 31, 2001, 2001.
14. IMS I. National Prescription Audit Plus. Chicago, IL: IMS. Inc.; October 15, 2001 2001.
15. Lin S-J, Lambert BL, Tan H, Toh SW. Frequency estimates from prescription drug datasets. Journal of Pharmacoepidemiology and Drug Safety. 2006;15:512-520.
16. Lambert BL, Lin S-J, Toh S, et al. Frequency and neighborhood effects on auditory perception of drug names in noise. Acoustical Society of America Journal. 2005;118(3):1955.
17. Rodman J. The effect of bandwidth on speech intelligibility. http://www.polycom.com/common/pw_cmp_updateDocKeywords/0,1687,1809,00.pdf. Accessed June 17, 2005.
18. Jurafsky D, Martin JH. Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition. Upper Saddle River, NJ: Prentice Hall; 2000.
19. Storker HL. Methods for minimizing the confounding effects of word length in the analysis of phonotactic probability and neighborhood density. J Speech Lang Hear Res. 2004;47:1454-1468.
20. Hosmer DW, Lemeshow S. Applied logistic regression. New York: John Wiley & Sons; 1989.
21. Selvin S. Statistical analysis of epidemiological data. New York: Oxford University Press; 1996.
22. Kleinbaum DG. Logistic regression: a self-learning text. New York: Springer-Verlag; 1994.
23. Lambert BL, Yu CT, Jia L, Lewis DD, Liu F, Fisher WM. A search engine for drug names. http://dolphin.cs.uic.edu:8080/DrugSearch/ Accessed August, 2008.
24. Lambert BL, Yu C, Thirumalai M. A system for multi-attribute drug product comparison. J Med Syst. 2004;28(1):29-54.
25. Voorhees E. Overview of TREC 2002. PDF Document] http://trec.nist.gov/pubs/trec11/papers/OVERVIEW.11.pdf. Accessed March 23, 2004.

Program Director/Principal Investigator (Last, First, Middle): Lambert, Bruce, L.

Inclusion Enrollment Report

This report format should NOT be used for data collection from study participants.

Study Title: Relationship between subjective familiarity and objective prescribing frequency

Total Enrollment: 148 **Protocol Number:** 1

Grant Number: 1R01HS011609

PART A. TOTAL ENROLLMENT REPORT: Number of Subjects Enrolled to Date (Cumulative) by Ethnicity and Race

Ethnic Category	Sex/Gender			
	Females	Males	Unknown or Not Reported	Total
Hispanic or Latino	8	4	0	12 **
Not Hispanic or Latino	87	41	0	128
Unknown (individuals not reporting ethnicity)	3	3	2	8
Ethnic Category: Total of All Subjects*	98	48	2	148 *
Racial Categories				
American Indian/Alaska Native	0	0	0	0
Asian	27	14	0	41
Native Hawaiian or Other Pacific Islander	1	0	0	1
Black or African American	11	3	0	14
White	48	24	0	72
More Than One Race	0	0	0	0
Unknown or Not Reported	11	7	2	20
Racial Categories: Total of All Subjects*	98	48	2	148 *

PART B. HISPANIC ENROLLMENT REPORT: Number of Hispanics or Latinos Enrolled to Date (Cumulative)

Racial Categories	Females	Males	Unknown or Not Reported	Total
American Indian or Alaska Native				
Asian				
Native Hawaiian or Other Pacific Islander				
Black or African American				
White				
More Than One Race				
Unknown or Not Reported	8	4	0	12
Racial Categories: Total of Hispanics or Latinos**	8	4	0	12 **

* These totals must agree.
** These totals must agree.

Program Director/Principal Investigator (Last, First, Middle): Lambert, Bruce, L.

Inclusion Enrollment Report

This report format should NOT be used for data collection from study participants.

Study Title: EFFECT OF NOISE, SIMILARITY AND FAMILIARITY ON AUDITORY PERCEPTION

Total Enrollment: 250 **Protocol Number:** 2

Grant Number: 1R01HS011609

PART A. TOTAL ENROLLMENT REPORT: Number of Subjects Enrolled to Date (Cumulative) by Ethnicity and Race

Ethnic Category	Sex/Gender			
	Females	Males	Unknown or Not Reported	Total
Hispanic or Latino	7	1	0	8 **
Not Hispanic or Latino	150	86	0	236
Unknown (individuals not reporting ethnicity)	5	1	0	6
Ethnic Category: Total of All Subjects*	162	88	0	250 *
Racial Categories				
American Indian/Alaska Native	0	0	0	0
Asian	12	8	0	20
Native Hawaiian or Other Pacific Islander	1	3	0	4
Black or African American	14	4	0	18
White	124	71	0	195
More Than One Race	5	1	0	6
Unknown or Not Reported	6	1	0	7
Racial Categories: Total of All Subjects*	162	88	0	250 *

PART B. HISPANIC ENROLLMENT REPORT: Number of Hispanics or Latinos Enrolled to Date (Cumulative)

Racial Categories	Females	Males	Unknown or Not Reported	Total
American Indian or Alaska Native	0	0	0	0
Asian	0	0	0	0
Native Hawaiian or Other Pacific Islander	0	0	0	0
Black or African American	0	0	0	0
White	2	0	0	2
More Than One Race	0	0	0	0
Unknown or Not Reported	5	1	0	6
Racial Categories: Total of Hispanics or Latinos**	7	1	0	8 **

* These totals must agree.
** These totals must agree.

Program Director/Principal Investigator (Last, First, Middle): Lambert, Bruce, L.

Inclusion Enrollment Report

This report format should NOT be used for data collection from study participants.

Study Title: Evaluation of Drug Name Retrieval System

Total Enrollment: 20 **Protocol Number:** 3

Grant Number: 1R01HS011609

PART A. TOTAL ENROLLMENT REPORT: Number of Subjects Enrolled to Date (Cumulative) by Ethnicity and Race

Ethnic Category	Sex/Gender			
	Females	Males	Unknown or Not Reported	Total
Hispanic or Latino	1	0	0	1 **
Not Hispanic or Latino	16	3	0	19
Unknown (individuals not reporting ethnicity)	0	0	0	0
Ethnic Category: Total of All Subjects*	17	3	0	20 *
Racial Categories				
American Indian/Alaska Native				
Asian	5	0	0	5
Native Hawaiian or Other Pacific Islander	0	0	0	0
Black or African American	1	0	0	1
White	10	3	0	13
More Than One Race	1	0	0	1
Unknown or Not Reported	0	0	0	0
Racial Categories: Total of All Subjects*	17	3	0	20 *

PART B. HISPANIC ENROLLMENT REPORT: Number of Hispanics or Latinos Enrolled to Date (Cumulative)

Racial Categories	Females	Males	Unknown or Not Reported	Total
American Indian or Alaska Native	0	0	0	0
Asian	0	0	0	0
Native Hawaiian or Other Pacific Islander	0	0	0	0
Black or African American	0	0	0	0
White	0	0	0	0
More Than One Race	1	0	0	1
Unknown or Not Reported	1	0	0	1
Racial Categories: Total of Hispanics or Latinos**	1	0	0	1 **

* These totals must agree.
** These totals must agree.