Predicting and Preventing Drug Name Confusion Errors: Case-Control Analyses and Memory Experiments Based on Computerized Measures of Similarity

Bruce L. Lambert, Ph.D.
Department of Pharmacy Administration
Department of Pharmacy Practice
College of Pharmacy
University of Illinois at Chicago
lambertb@uic.edu

Presentation given to the U. S. Food and Drug Administration
Parklawn Building-Conference Room B
Rockville, MD
June 24, 1998

Acknowledgments

- This work is supported in part by grants and contracts from the U. S. Pharmacopeia, the Drug Information Association, the Institute for Safe Medication Practices, the U.S. Food and Drug Administration, and the UIC Campus Research Board.

- The U.S. Pharmacopeia graciously provided the author with electronic versions of the *USP DI–Volume I* and the *USAN Dictionary*.

- The assistance of Swu-Jane Lin, Sanjay Gandhi, Ken-Yu Chang, Dan Boring, Mike Cohen, Bill Brewer, Gary Dell, Keith Johnson, Diane Cousins, Don Rucker and Alan Baddeley and others is gratefully acknowledged.

Overview

- Basic facts about look-alike/sound-alike (LASA) errors
- Why do they occur?
- What can be done to predict or prevent them?

Basic Facts About LASA Errors

- What are they?
- When do they occur?
- How often do they occur?
- What are the consequences?

Why do LASA Errors Occur?

- Cognitive psychological factors
- Environmental/workplace factors
- Interaction between psychological and workplace factors

Basic Prevention Strategy

- Identify which psychological processes are involved in each type of error

- Use theories from cognitive psychology to guide our efforts at error minimization

LASA Error Types Broken Down by Psychological Faculty

- Memory Errors (e.g., forgetting, misremembering)
- Perceptual Errors (e.g., misperceiving)
- Action/motor control errors (e.g., typographical errors, order entry errors)

Example: Memory Errors

Known Psychological Phenomena (From Baddeley)

- Phonological similarity effect
- Word length effect
- Unattended speech effect
- Articulatory suppression effect
- Word frequency effect

Baddeley's Working Memory Model

- Consists of central executive, visuo-spatial sketchpad, and phonological loop.
- Phonological similarity effects are explained by the phonological loop.
- Phonological representations of words are subject to partial loss due to decay and interference (Gathercole and Baddeley, 1993).
- Loss and decay are most consequential when an item is phonologically similar to another item already in the phonological store.

Using the Theory to Guide Prevention

- Develop automated measures of similarity
- Use automated measures to study relationship between similarity and probability of error

Automated Measures of String Similarity

- Bigram

Atarax	Marax	Common bigrams	
(at, ta, ar, ra, ax)	(ma, ar, ra, ax)	(ar, ra, ax)	3/9 = 0.33

- Trigram

Atarax	Marax	Common trigrams	
(ata, tar, ara, rax)	(mar, ara, rax)	(ara, rax)	2/7 = 0.29

- Edit distance

Atarax Marax

How many insertions or deletions would it take to transform one word into the other? In this case, 2: (1) Change A to M, (2) delete t.

Results

- Case-control studies
- Recall and recognition memory experiments

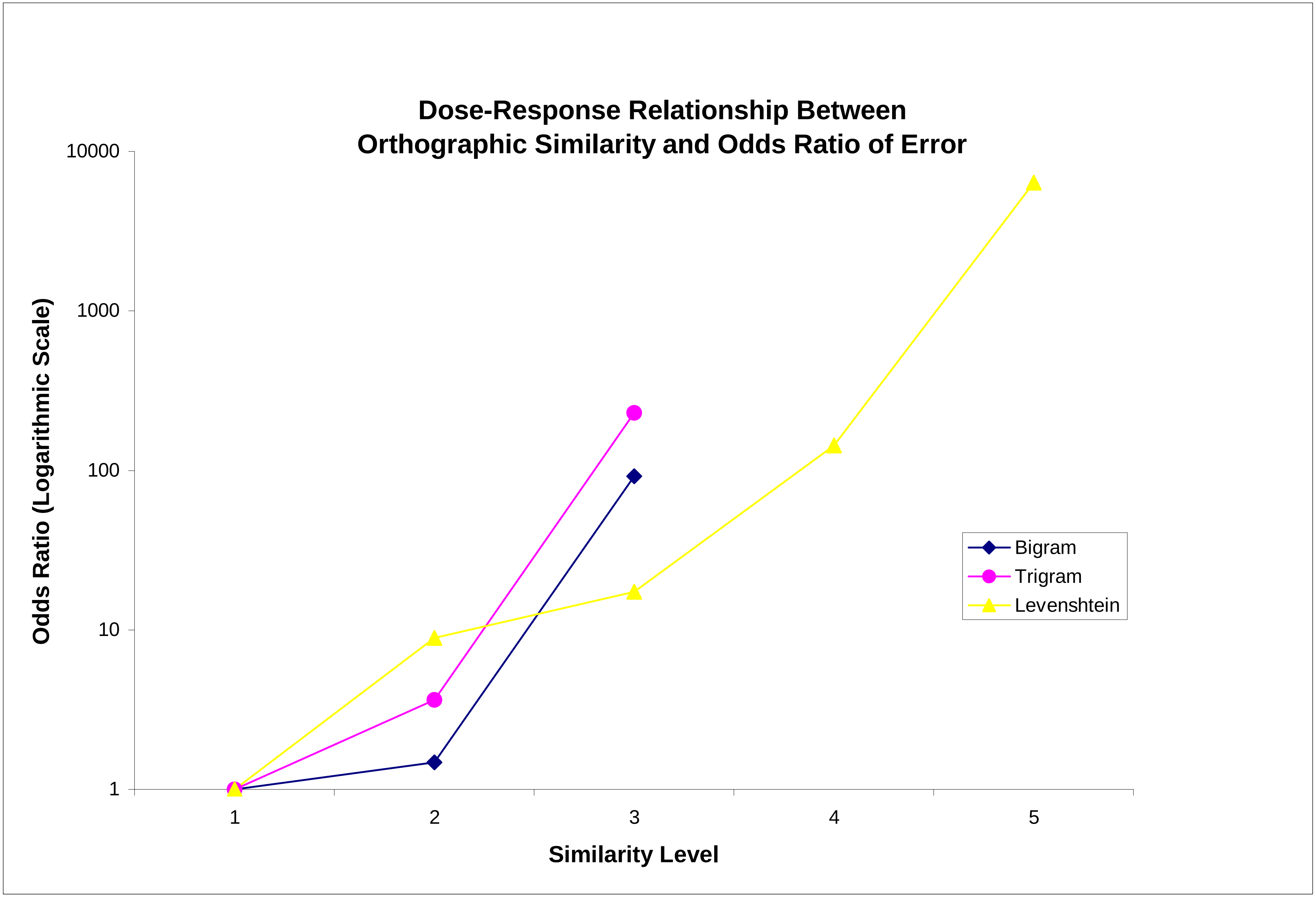
Case-Control Study #1

- Compared 969 known error pairs to 969 randomly selected control pairs
- Compared distribution of similarity scores
- Measured association between similarity and probability of being an error pair

Case-Control Studies (cont'd.)

- Evaluated prognostic test

- Analyzed dose-response relationship


Dose-Response Relationship Between
Orthographic Similarity and Odds Ratio of Error
Odds Ratio (Logarithmic Scale)
10000
1000
100
10
1
1
2
3
4
5
Similarity Level
Bigram
Trigram
Levenshtein

Effect of Orthographic Similarity on Immediate Free Recall

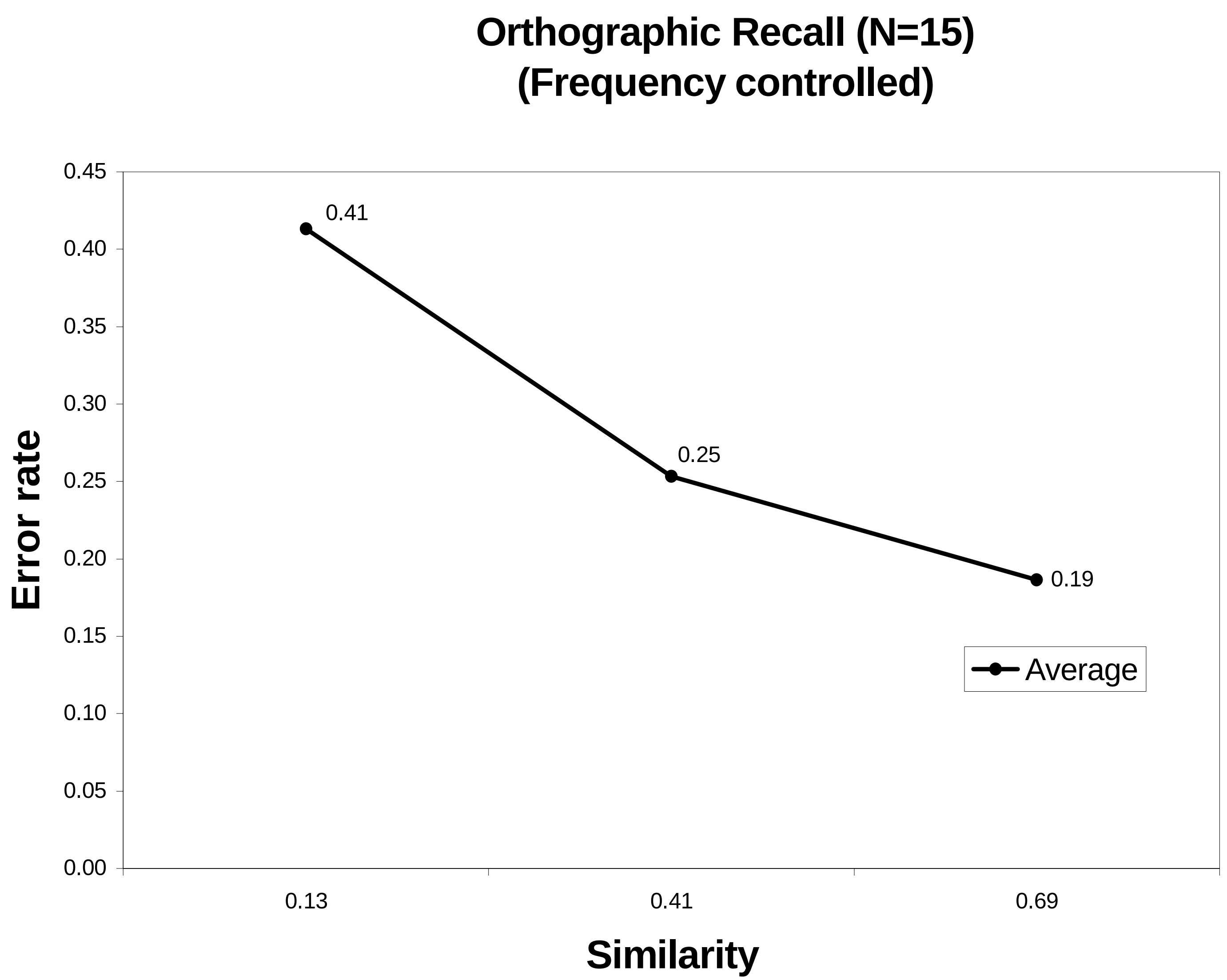
- N=15 licensed pharmacists
- Superlab experiment program, 1 word / 2 sec
- Words taken from combined 1992-1994 NAMCS data
- Bigram + 1 space at start and end of word as similarity measure
- Frequency matched, 15 trials
- Any misspelling was an coded as an error

Orthographic Recall: Stimulus Materials

Stimulus Words for Orthographic Recall Experiment

NAMCS Freq.	Bigram Sim.	Names		
161010.67	0.1035	correctol	formalin	bellergal
152828.33	0.3895	midamor	cetamide	blephamide
151138.33	0.7320	nolamine	calamine	alamine
790665.0	0.1072	florinef	fedahist	beclovent
776249.0	0.4862	lubriderm	estraderm	eryderm
649728.7	0.7245	thorazine	norazine	clorazine
29615.666	0.1603	dihistine	filgrastim	decaspray
29222.334	0.3602	cataflam	cotazym	azactam
26289.0	0.6777	prostigmin	prolastin	progestin
246775.0	0.1111	flaxedil	cyclogyl	cetapred
297534.34	0.4351	iberet	fibermed	fibercon
251222.0	0.6850	calcidrine	dalcaine	alcaine
618969.0	0.1603	eldercaps	aspercreme	aldactone
522223.34	0.3619	imuran	iophen	anuphen
548433.7	0.6210	pertussin	histussin	detussin

Orthographic Recall: Results



Orthographic Recall: Results

- Single factor repeated measures ANOVA
- Single factor is orthographic similarity, 3 levels: hi (0.69), med (0.41), lo (0.13)
- $F(2, 14) = 17.15$, $MSE = 0.201$, $\underline{p} < 0.0001$
- All but hi-med post hoc comparisons are significant
- Contrary to our prediction, recall is actually ENHANCED by similarity
- Due to rhyming heuristic and/or confounding by number of unique syllables to be remembered. Baddeley calls this the ‘availability effect’.

Effect of Orthographic Similarity on Recognition Memory

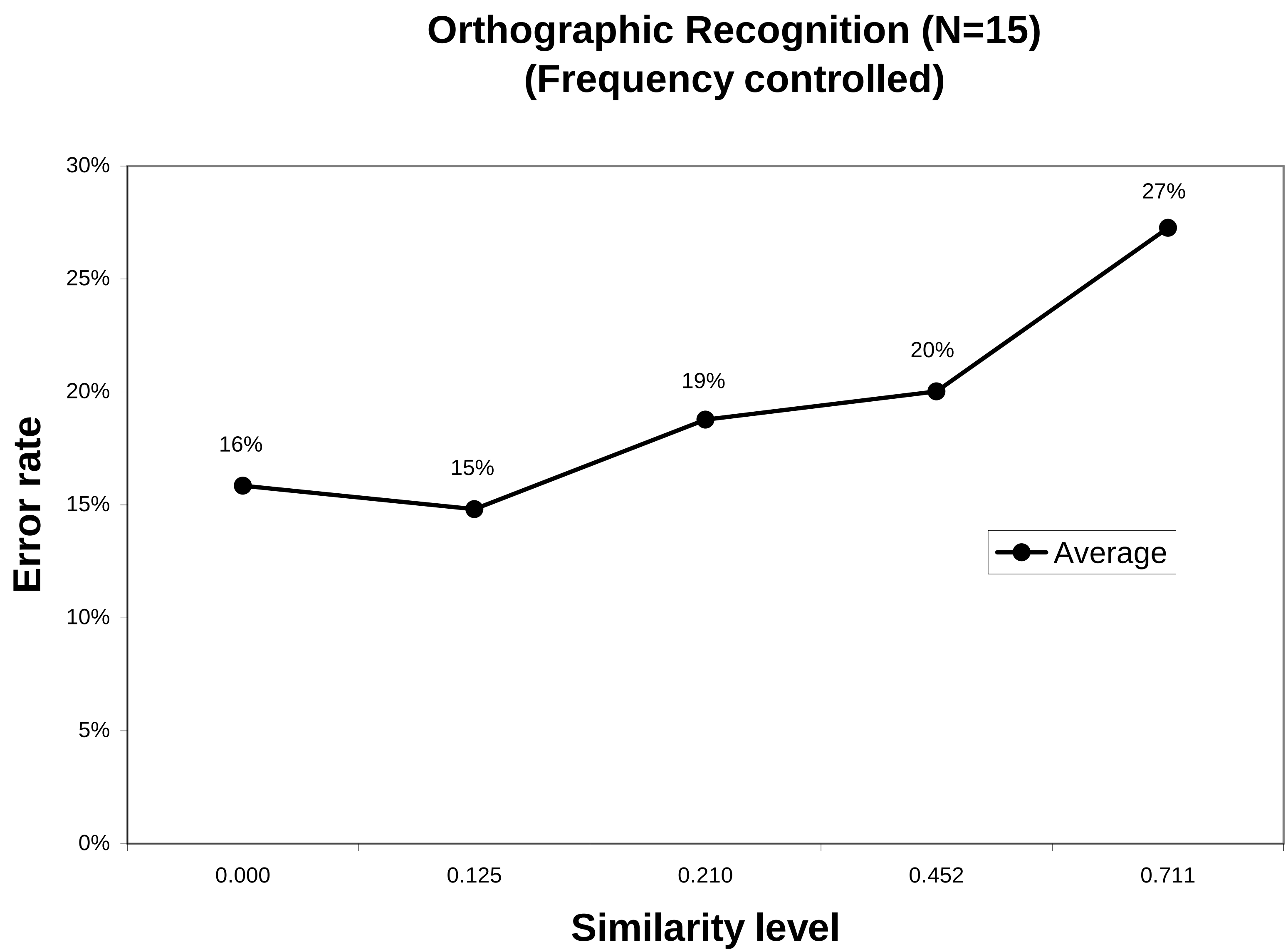
- N=15 licensed pharmacists
- Superlab experiment program, 1 word/sec
- Words taken from combined 1992-1994 NAMCS data
- Bigram + 1 space at start and end of word as similarity measure
- Frequency matched
- 5-word study list followed by 10-word test list

Orthographic Recognition (cont'd.)

- Single factor, repeated measures ANOVA
- Single factor is orthographic similarity, 5 levels
- Order or presentation randomized
- 16 trials

Stimulus Materials for Orthographic Recognition Test

Frequency	Similarity	Name1	Name2
861364.0	0.7778	PROLIXIN	PROCOLIN
863195.0	0.4706	TENUATE	TRILISATE
849898.0	0.2105	ZITHROMAX	CAPITROL
858872.0	0.125	ROGENIC	BENYLIN
860639.0	0.0	ACLOVATE	NICOBID
8625.0	0.75	ARAMINE	ANAMINE
8269.0	0.4706	TRANTOIN	TRIACIN
8191.0	0.2105	NEPHRAMINE	SINOPHEN
8264.5	0.125	DIALOSE	PINOVAL
8269.0	0.0	PARAFLEX	OTRIVIN
71847.5	0.7368	HYDROCORT	HYDROCET
72387.5	0.4706	ANTIMINTH	TIMENTIN
72002.0	0.2105	MINIZIDE	ALLERGINE
71961.0	0.125	HEXALOL	TEMARIL
71857.5	0.0	BELEXAL	MARAZIDE
557534.5	0.7143	URISEP	URISED
568357.5	0.4444	EXELDERM	ELDEPRYL
557069.5	0.2105	DERMATOP	CAPTOPRIL
558434.5	0.125	MYLICON	EMPIRIN
557307.5	0.0	SENOKOT	EFUDEX
159777.0	0.7059	RIFAMPIN	RIFADIN
158092.0	0.4444	CHOLEDYL	CHOLYBAR
158877.0	0.2105	TRINSICON	ATABRINE
159785.0	0.125	GENORA	DESFERAL
159724.0	0.0	CLAFORAN	MERITAL
218197.5	0.6667	PRAMASONE	ORASONE
226041.5	0.4706	DRIXORAL	FLUORAL
222957.0	0.2105	ENDURON	DANTROLENE
224622.0	0.125	GLUCOLA	TALACEN
224380.0	0.0	ROWASA	FERRALET
3801978.0	0.6667	ISORDIL	ISOMIL
3903035.0	0.4	INDOCIN	DOXEPIN
3593548.5	0.2105	ANTIVERT	ASCRIPTIN
3764190.0	0.125	ZOLADEX	RELAFEN
3804967.5	0.0	LOTENSIN	NIZORAL
130319.5	0.6667	PANADOL	NADOLOL
135120.0	0.4444	HALOTHANE	LOXITANE
128733.0	0.2105	THERAPLEX	HEXADROL
130320.5	0.125	ESTINYL	VEPESID
130718.0	0.0	BETALIN	RYNATUSS



Orthographic Recognition (N=15)
(Frequency controlled)
Error rate
30%
25%
20%
15%
10%
5%
0%
16%
15%
19%
20%
27%
Average
0.000
0.125
0.210
0.452
0.711
Similarity level

Orthographic Recognition: Results

- $F(4, 14) = 9.14$, $MSE = 0.04$, $\underline{p} < 0.0001$
- Highest level is significantly different from all other levels ($p < .05$)
- Supports hypothesized effect of similarity on recognition errors
- Recognition is a more ecologically valid task (i.e., it more closely resembles what pharmacists actually do)

Phonological Similarity: Automated Measure

- USP Dictionary provides pronunciation guides for some but not all listed names
- Lincomycin = (lin koe mye′ sin)
- Tobramycin = (toe bra mye′ sin)
- Used guides as basis for quasi-phonological measure

Phonological Similarity Measure

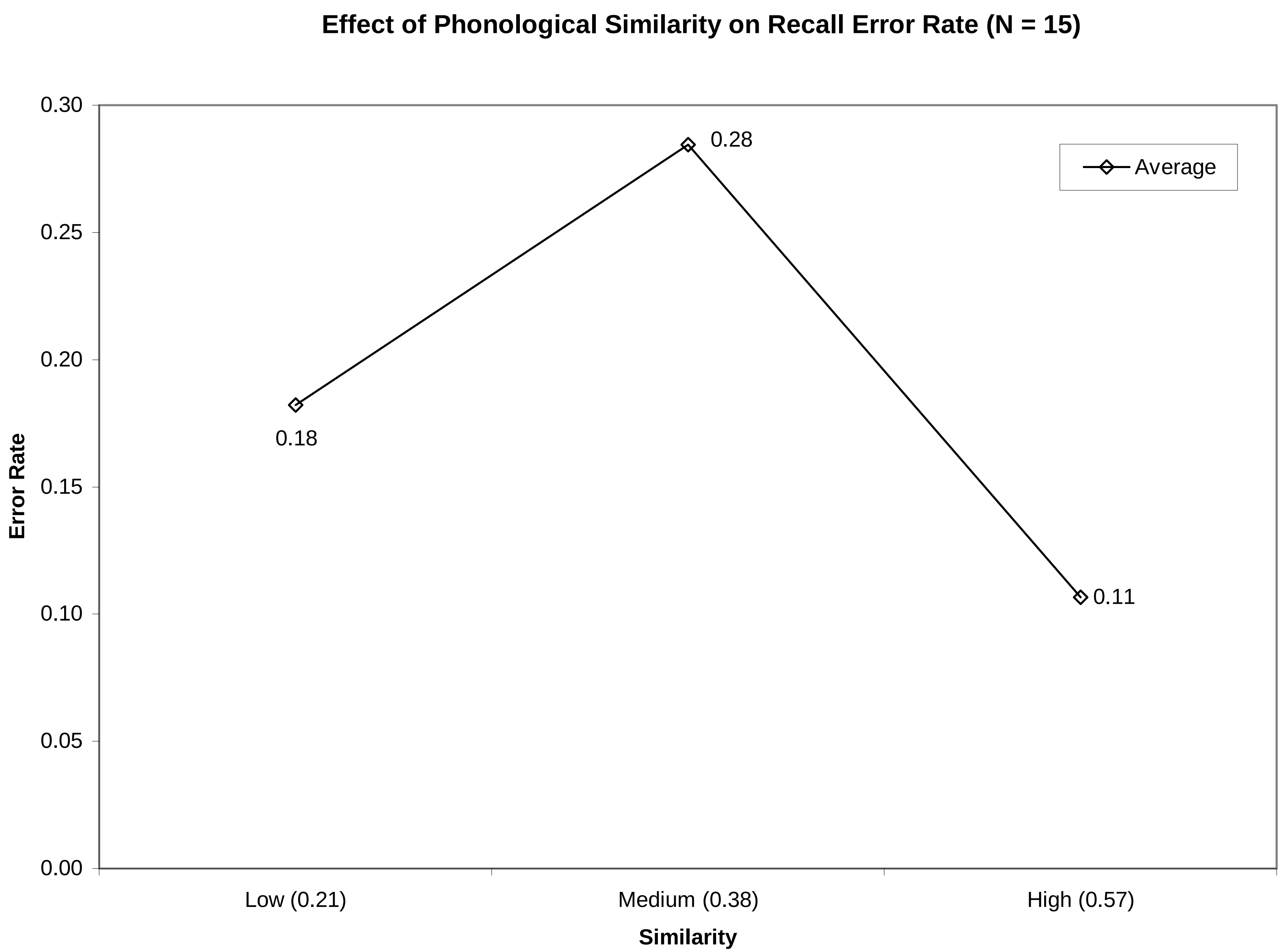
- Several phonological features identified as important in previous research (Drewnowski & Murdock, 1980, The role of auditory features in memory span for words)
- Number of syllables, location of stressed syllable, initial syllable (phoneme), terminal syllable, stressed vowel
- I used initial syllable, terminal syllable, accented syllable, accent position, number of syllables, and number of common syllables

Phonological Similarity and Recall

- N=15 licensed pharmacists
- Free recall task
- Any misspelling coded as an error
- Single factor, repeated measures design
- Single factor is phonological similarity, 3 levels
- 15 trials

Phonological Recall Stimulus Sets

Frequency	Phono. Sim.		Names	
279781.66	0.6250	lincomycin	tobramycin	vancomycin
254758.67	0.3888	cimetidine	minoxidil	simethicone
274705.34	0.2083	clotrimazole	cytarabine	temazepam
56608.0	0.5833	carbidopa	levodopa	methyldopa
61170.0	0.4027	astemizole	indapamide	miconazole
60907.332	0.2083	adenosine	chlorzoxazone	nevirapine
77035.664	0.5833	thioguanine	thiotepa	thiothixene
80775.0	0.3888	methenamine	methimazole	metolazone
77858.0	0.2096	carbamazepine	clozapine	isradipine
103313.0	0.5416	famotidine	nizatidine	ranitidine
133977.67	0.3750	alprazolam	triazolam	trimethoprim
118570.336	0.2096	acetone	amiodarone	norfloxacin
245547.33	0.5099	fenoprofen	ketoprofen	metoprolol
246407.0	0.3611	chlorthalidone	piroxicam	risperidone
209988.33	0.2142	amoxapine	cefazolin	prednisolone


Effect of Phonological Similarity on Recall Error Rate (N = 15)
Error Rate
0.30
0.25
0.20
0.15
0.10
0.05
0.00
0.28
0.18
0.11
Average
Low (0.21)
Medium (0.38)
High (0.57)
Similarity

Phonological Recall: Results

- $F(2, 14) = 10.73$, $MSE = 0.12$, $p < 0.0003$
- All but hi-lo comparisons were significant using Tukey's test, minimum signigficant difference = 0.0947.
- Results partially support our predictions

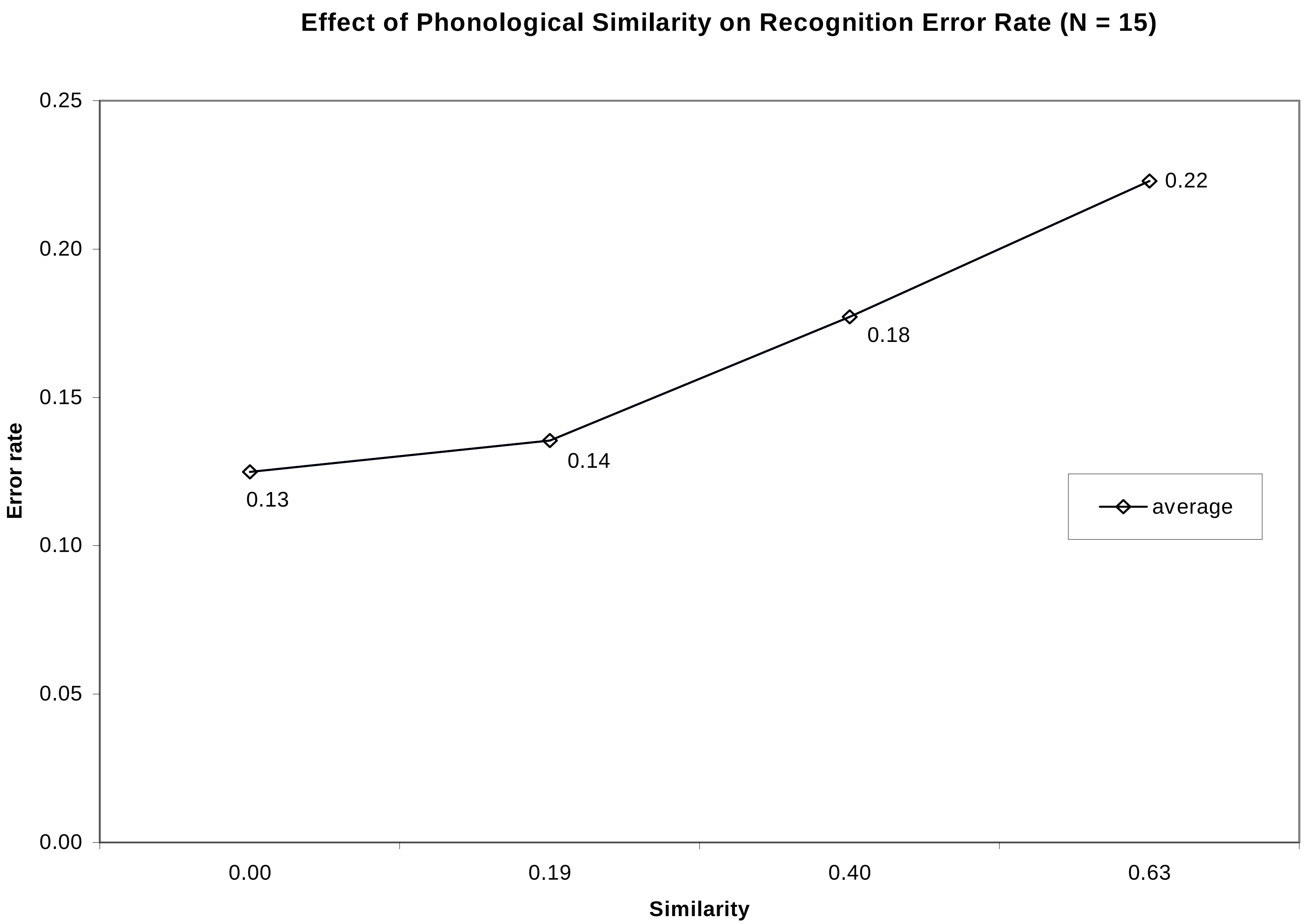
Phonological Similarity and Recognition

- N = 15 licensed pharmacists
- 8 word study list, 16 word test list
- 1 word / sec presentation rate
- Randomized sequence
- Single factor repeated measures ANOVA
- Single factor is phonological similarity (4 levels)
- 8 trials

Phonological Similarity and Recognition

Stimulus Materials

NAMCS Freq.	Phono. Sim.	Names	
35516.5	0.75	chloroform	chloroquine
57995.0	0.42	glycerin	tolmetin
25862.5	0.25	cisapride	urea
46868.0	0.00	benzoin	filgrastim
684121.0	0.71	acyclovir	ganciclovir
689955.5	0.38	felodipine	nifedipine
811208.5	0.17	isosorbide	oxytocin
651481.0	0.00	aminophylline	baclofen
722570.0	0.67	betamethasone	dexamethasone
354297.0	0.43	mephobarbital	metronidazole
334157.5	0.19	griseofulvin	riboflavin
530160.0	0.00	cyclosporine	nitrofurantoin
94649.0	0.67	tolazamide	tolbutamide
105210.0	0.42	mannitol	sorbitol
99321.5	0.17	melphalan	propofol
93620.0	0.00	phenol	probenecid
669189.0	0.58	atropine	loxapine
605468.5	0.42	calamine	phentermine
635358.5	0.17	captopril	ipecac
610143.0	0.00	dapsone	meprobamate
93151.0	0.58	digitalis	digitoxin
81987.0	0.38	didanosine	dienestrol
84583.0	0.17	glucagon	ichthammol
81848.0	0.00	flutamide	pentoxifylline
1793161.0	0.50	amikacin	bacitracin
2510795.5	0.38	ceftazidime	cephalexin
2038120.0	0.21	methotrexate	ofloxacin
1513675.5	0.00	indomethacin	nystatin
154957.5	0.58	cefaclor	cephradine
187119.5	0.42	carbachol	carmustine
180663.0	0.17	lactulose	succimer
170942.0	0.00	estrone	misoprostol


Effect of Phonological Similarity on Recognition Error Rate (N = 15)
0.25
0.20
0.15
0.10
0.05
0.00
Error rate
0.22
0.18
0.14
0.13
average
0.00
0.19
0.40
0.63
Similarity

Phonological Similarity and Recognition: Results

- $F(3, 14) = 6.91$, $MSE = 0.03$, $\underline{p} < 0.0007$
- High similarity differed from low and low-medium ($\underline{p} < .05$) using Tukey's test, minimum significant difference = 0.065.
- Results support our prediction

Case Control Study #2

- N = 1127 cases (from published reports), and N = 1127 controls
- Controls randomly selected from unique names among cases
- Twenty distinct similarity measures evaluated by 10-fold cross validation
- Best three univariate measures used to form multivariate logistic regression model
- Multivariate model also evaluated using 10-fold cross-validation

Performance of various similarity and distance measures in case-control analyses of look-alike and sound-alike medication errors (N = 1127 cases, N = 1127 controls) based on 10-fold cross-validation.

Measure	Accuracy	Sensitivity	Specificity
Trigram-2b	0.9272	0.9223	0.9321
Bigram-1b	0.9156	0.8848	0.9464
Normalized Edit Distance	0.9152	0.9009	0.9295
Bigram-1b1a	0.9080	0.9054	0.9107
Trigram-2b2a	0.8982	0.8893	0.9071
Trigram-1b1a	0.8875	0.8590	0.9170
Trigram-1b	0.8830	0.8036	0.9625
Edit Distance	0.8813	0.8223	0.9402
Bigram-1a	0.8799	0.8518	0.9080
Editex	0.8790	0.8330	0.9250
Edit-Skeleton	0.8737	0.8116	0.9357
Bigram	0.8647	0.8518	0.8777
Edit-Phonix	0.8616	0.9330	0.7902
Edit-Soundex	0.8594	0.7750	0.9438
Edit-Omission	0.8500	0.8500	0.8500
Trigram-2a	0.8478	0.8491	0.8464
Trigram-1a	0.8464	0.7750	0.9179
LCS	0.8384	0.8473	0.8295
Trigram	0.8371	0.7295	0.9446
Tapered Edit Distance	0.8112	0.7964	0.8259

Note. On each fold of cross-validation, a locally optimal cutoff point was chosen by evaluating 100 different cutoff points evenly spaced across the range of the given measure. Ninety percent of the N = 2254 pairs of cases and controls were used to select the cutoff point for each fold. Each cutoff point was then tested on the remaining 10% of the data. This process was repeated 10 times. The ten test sets were non-overlapping. Trigram-2b = trigram with two spaces added before the word, Trigram-1a = trigram with one space added after the word, etc.

Sensitivity, Specificity, and Accuracy of Logistic Regression Models Using Trigram, Normalized Edit Distance, and Editex Measures (Based on 10-Fold Cross-Validation, N = 1127 cases, N = 1127 controls)

Trial	Sensitivity	Specificity	Overall Accuracy
1	0.958	0.971	0.964
2	0.917	0.942	0.929
3	0.944	0.974	0.960
4	0.926	0.948	0.938
5	0.931	0.959	0.947
6	0.958	0.971	0.964
7	0.918	0.974	0.947
8	0.925	0.932	0.929
9	0.940	0.935	0.938
10	0.945	0.982	0.964
Total	0.937	0.959	0.948

Note. On each fold of cross-validation, a logistic regression model was formed using 90% of the N = 2254 pairs of cases and controls. Each model was then tested on the remaining 10% of the data. This process was repeated 10 times. The ten test sets were non-overlapping

General Discussion

- What’s going on with recall?
- Need better, more explicit cognitive models of the underlying processes
- Overall error rate is quite low in phonological tests
- Recognition is clearly adversely affected by similarity
- Recognition is ecologically valid (more or less)
- Need to address illegibility in handwritten materials
- Need to examine neighborhood frequency and neighborhood density effects on perceptual identification and progressive demasking tasks
- Need ‘truly’ ecologically valid tasks (mock dispensing situations)
- Will be collecting data from lay people (undergrads) this Summer/Fall
- Looking for collaborators/consultants for future grant proposals
- Need better handle on word/prescribing frequency

How Can We Prevent LASA Errors?

- Engineer the drug lexicon to make it ‘error resistant’
- Engineer the work environment to make it ‘error resistant’

Engineer the Drug Lexicon

- Each drug is a point in a multi-dimensional space
- Dimensions of this space include orthographic and phonological representation as well as dose, schedule, route of administration, color, shape, etc.
- Errors occur when drug products are ‘too close’ to one another in this space

Engineering the Lexicon

- Use automated measures of similarity to screen new drug products
- Only approve new products that are a ‘safe distance’ from existing products
- Something like this is routinely done as part of the legal screening of new trademark names

Engineering the Drug Lexicon

- We need a reference standard database of drug information against which new drug products would be screened
- More research is needed to determine how close is ‘too close’
- Must decide what to do when new name is ‘too close’ to an old name

Engineering the Drug Lexicon

- Automated searches should be part of failure mode and effects analysis (FMEA)

- FMEA should be a routine part of the FDA approval process

Engineering the Work Environment

- No talking while prescribing, dispensing, or administering drugs
- No alphabetical storage of drugs
- No handwritten prescriptions
- No faxed prescriptions
- Use of bar codes where feasible
- Use of additional retrieval cues where feasible (e.g., dose, indication, scheduling, etc.)

Summary

- Basic facts about look-alike/sound-alike (LASA) errors
- Why do they occur?
- What can be done to predict or prevent them?

lambertb@uic.edu