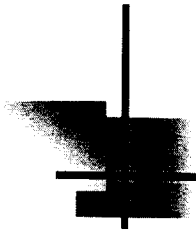
Computer Assisted Decision Analysis in Drug Naming

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



Overview

- How can computer resources be used to objectively measure differences between name pairs?
- How can computer resources be used to calculate weights for various elements in name similarity?
- Can computer assisted pattern recognition support the decision process to determine name/name similarities?

Preface: Need to Change Focus!

- Names are not enough: Must focus on drug *products* not drug *names*
- Similarity is not enough: Must focus on similarity and *frequency* (of prescribing)
- Error reduction is not enough: Must focus on harm reduction
- How to balance public risk against private/corporate benefit?

Objective Measures of Similarity

- N-gram: based on proportion of n-letter subsequences that two names have in common
  - Bigram: two-letter subsequences (e.g., Premarin=Pr, re, em, ma, ar, ri, in)
  - Trigram: three-letter subsequences (Pre, rem, etc.)
- Edit distance: the number of insertions deletions or substitutions needed to transform one name into another

Objective Measures of Similarity

- N-gram and edit distance measures can be used on any formal representation of the name (spelling or phonological)
  - Use phonetic alphabet (e.g., International Phonetic Alphabet or ARPAbet)
  - ARPAbet: *Zyprexa*: z ay p r eh k s ax
  - Phoneme bigrams: z ay, ay p, p r, r eh, eh k, k s, s ax

Objective Measures of Similarity

- There are many variations on these basic measures
  - Add spaces before or after to emphasize beginning or ending of name
  - Use different weights depending on position of letters
  - Use different equations to compute numerical similarity (Dice, Hamming, etc.)
  - Allow approximate matches between letters (e.g., m=n, a=e=i=o=u)

Objective Measures of Similarity

- Power of simple descriptive analyses
  - Ten most common three-letter prefixes in US brand names:
    - Pro-, Bio-, Car-, Tri-, Vit-, Pre-, Nut-, Ult-, Con-, Per-
  - Lambert BL, Chang KY, Lin SJ. Descriptive analysis of the drug name lexicon. *Drug Inf J.* 2001;35:163-172.

Objective Measures Do Predict Probability of Human Error

- Similarity accurately distinguishes between known error pairs and non-error pairs
- Greater objective similarity correlated with higher rates of recognition memory errors by laypeople and pharmacists
- Greater similarity correlated with lower rates of free recall errors
- Objective similarity correlated with subjective similarity (for experts and laypeople)
- Similarity neighborhoods predict visual perception errors

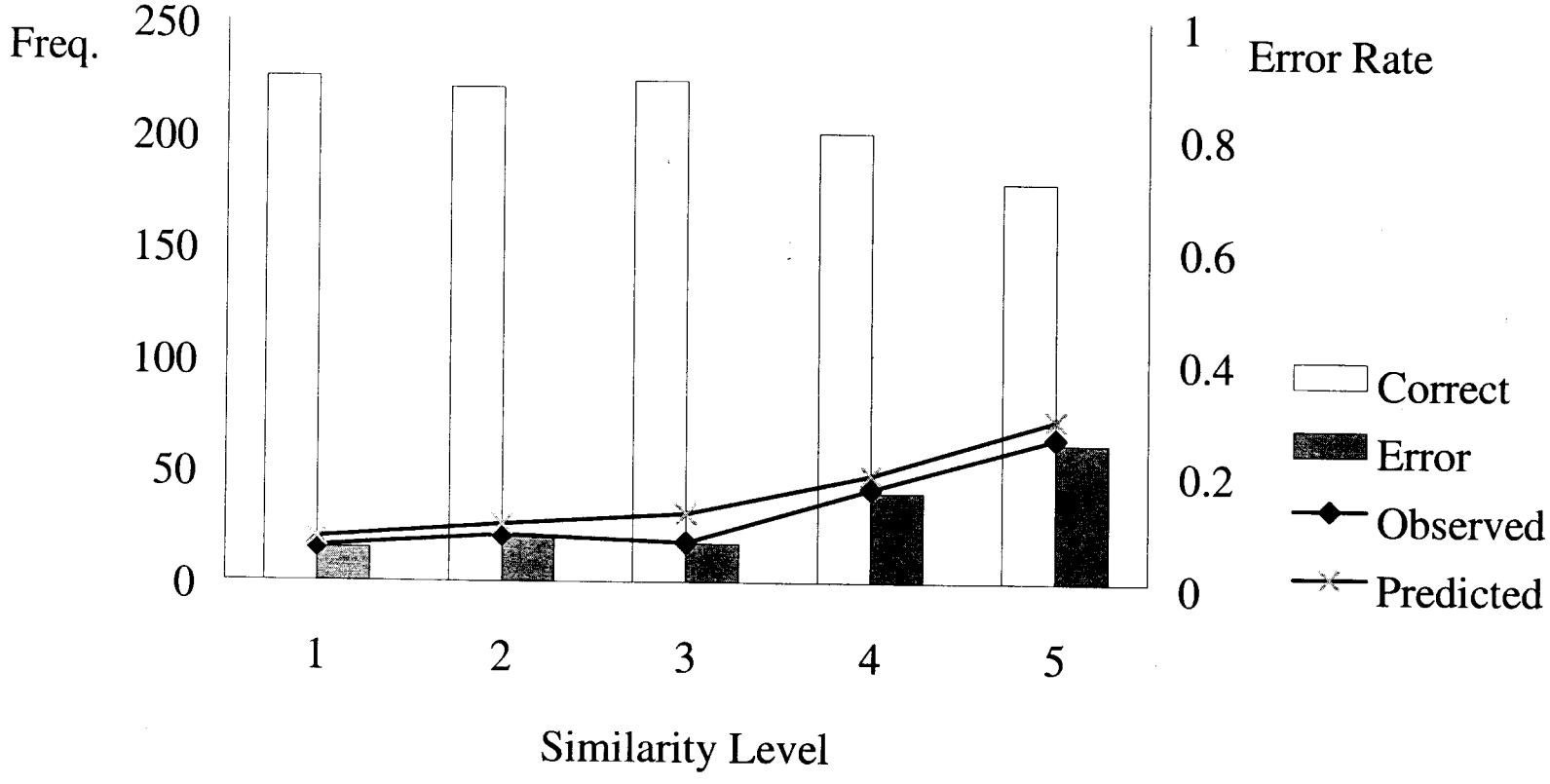


Figure 1. Effect of spelling similarity on pharmacists recognition memory errors

Lambert BL, Chang KY, Lin SJ. Effect of orthographic and phonological similarity on false recognition of drug names. *Soc Sci Med.* 2001;52:1843-1857.

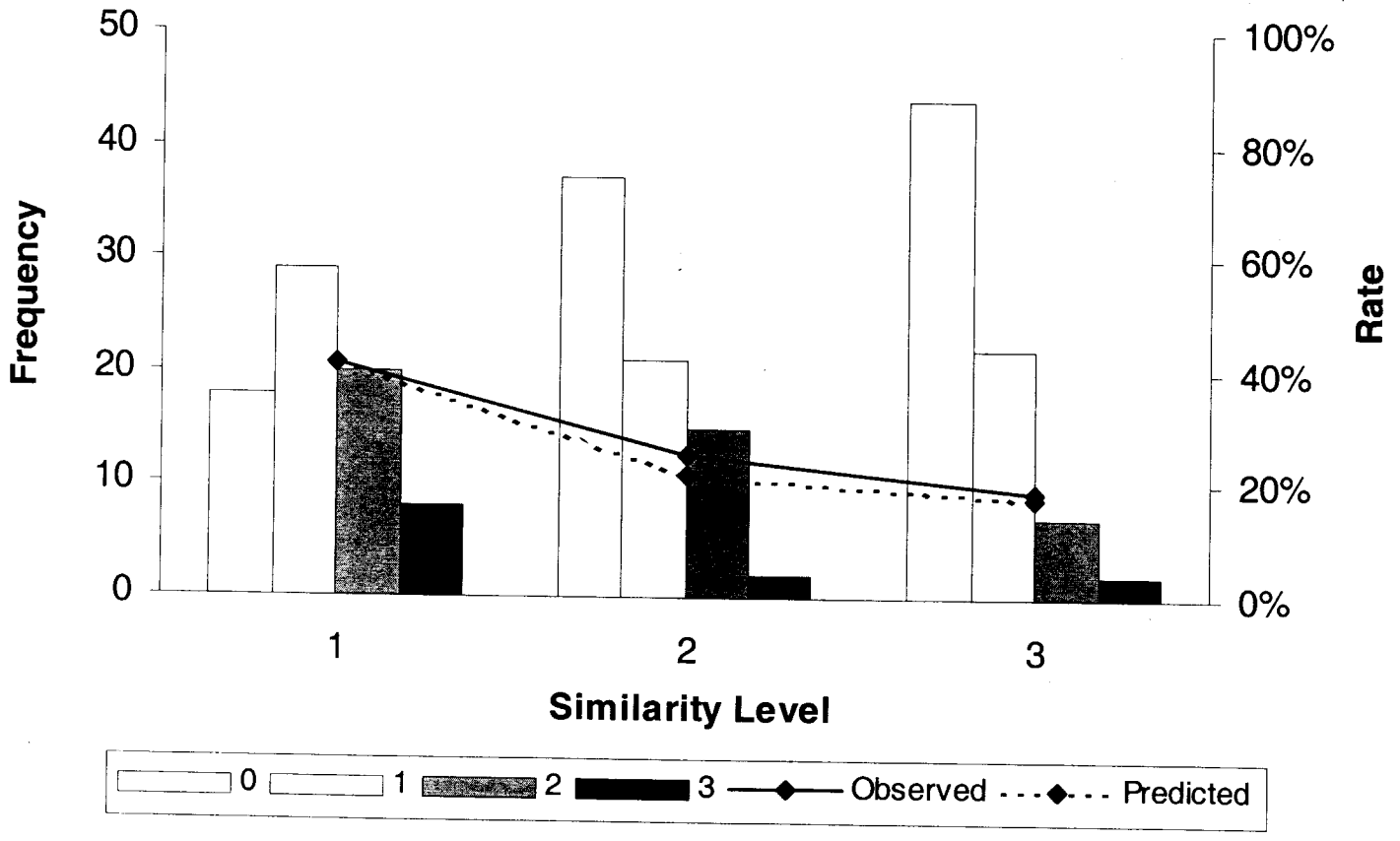


Figure 2. Effect of spelling similarity on pharmacists free recall errors.

Lambert BL, Chang K-Y, Lin S-J. Immediate free recall of drug names: effects of similarity and availability. *Am J Health-Syst Pharm.* 2003;60:156-168.

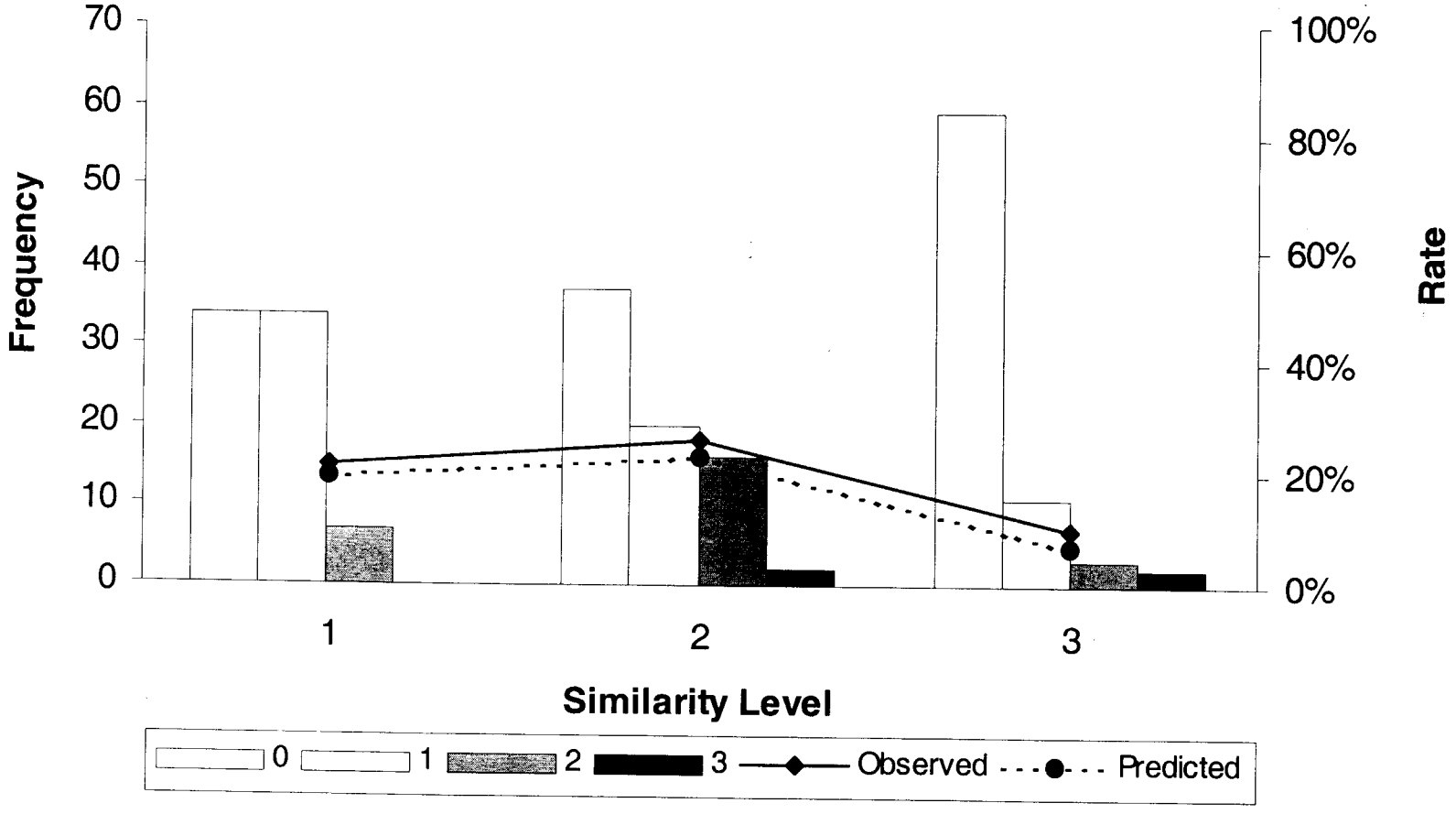


Figure 3. Effect of phonological similarity on pharmacists' free recall errors.

Lambert BL, Chang K-Y, Lin S-J. Immediate free recall of drug names: effects of similarity and availability. *Am J Health-Syst Pharm.* 2003;60:156-168.

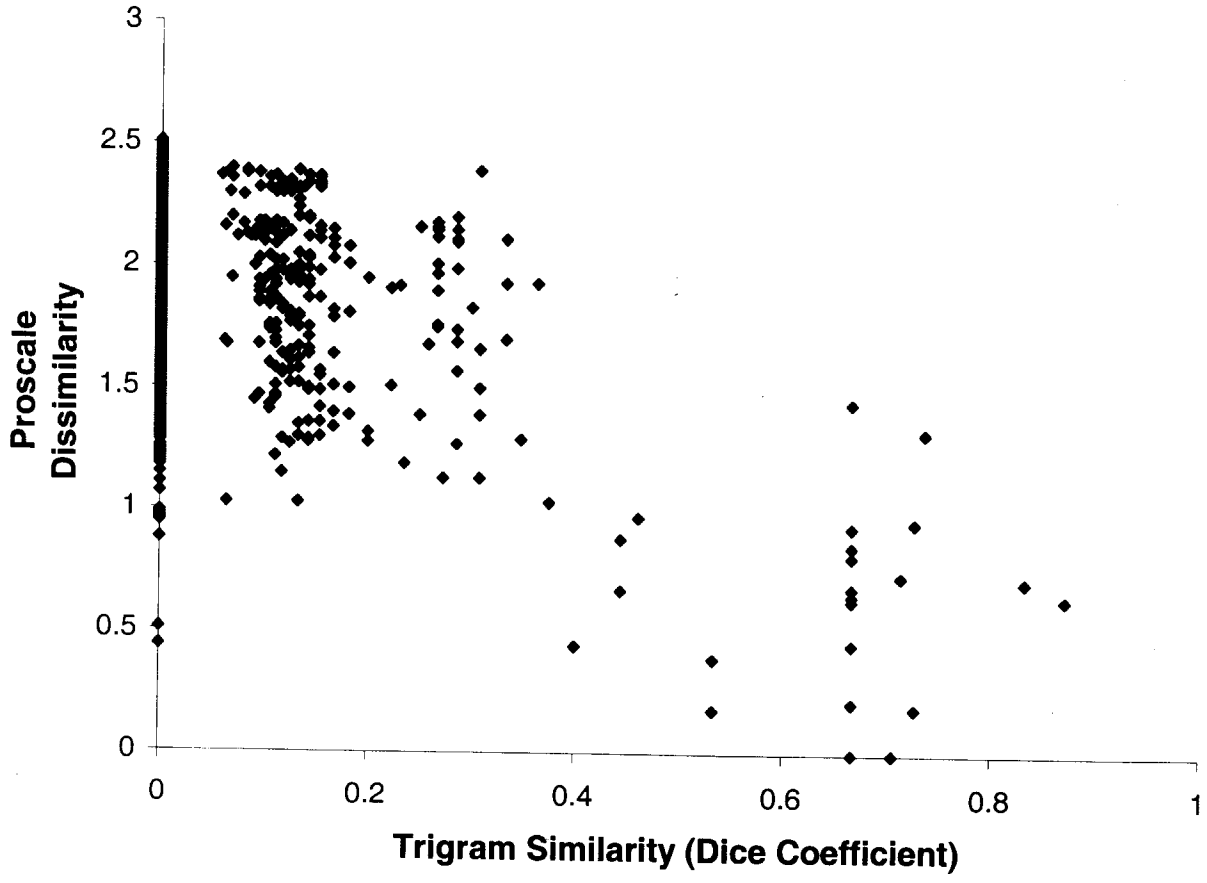


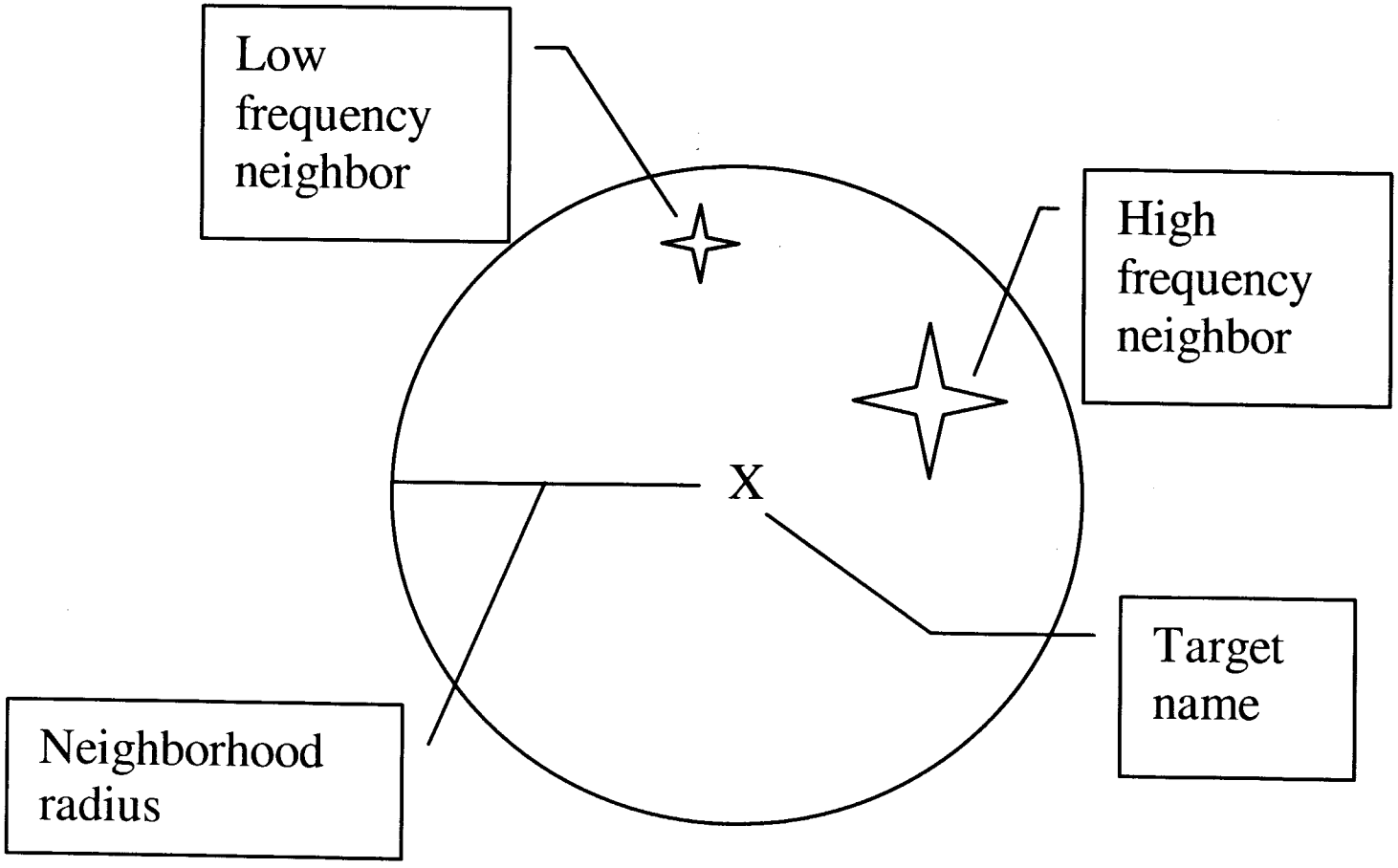
Figure 4. Relationship between objective similarity and lay people's subjective dissimilarity.

Lambert BL, Donderi D, Senders J. Similarity of drug names: Objective and subjective measures. *Psychology and Marketing.* 2002;19(7-8):641-661.

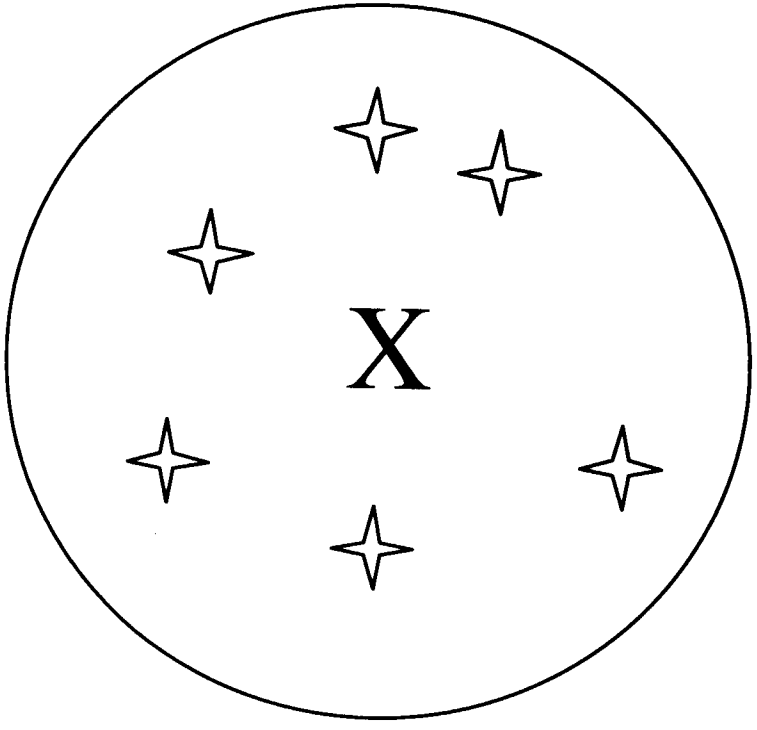
Neighborhoods Matter

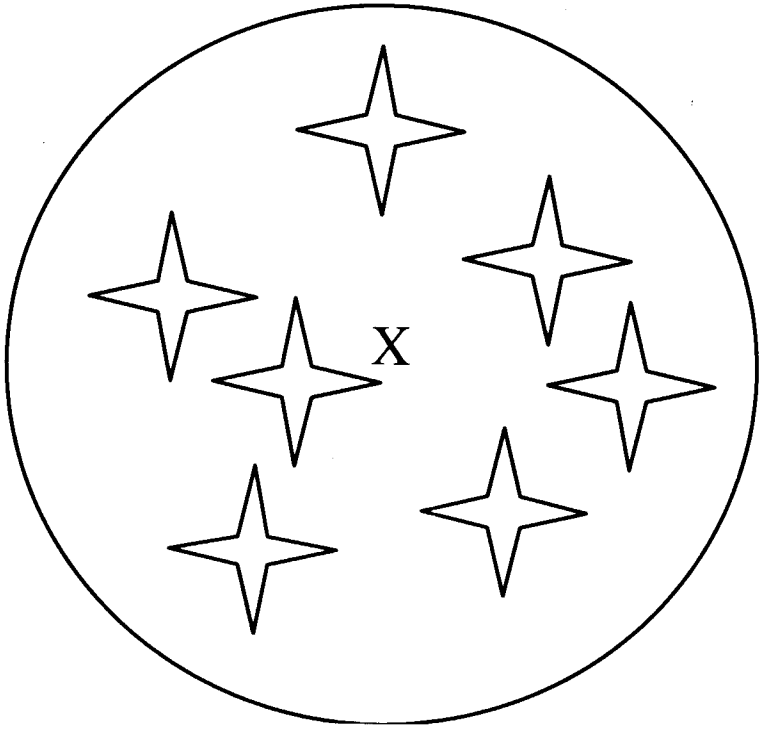
- Concept of similarity "neighborhood" is key part of modern theories of visual and auditory perception
- Neighborhood characteristics
  - Frequency (of prescribing)
  - Density
  - Radius (i.e., how close does a name have to be to be in another name's neighborhood?)

Neighborhood Illustration

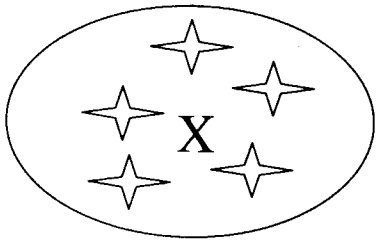


Dense Neighborhoods: High and Low Frequency

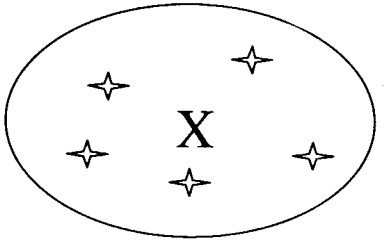




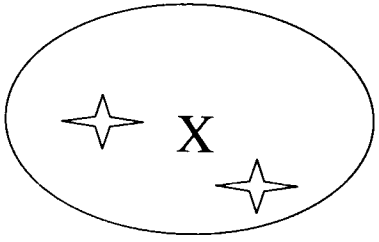
1) High SF, High NF, High ND



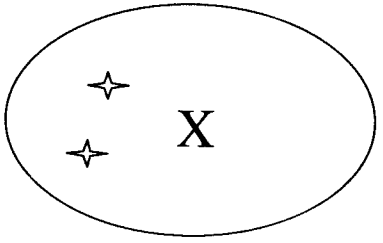
2) High SF, Low NF, High ND



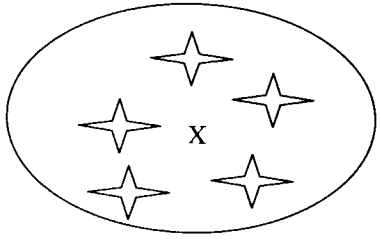
3) High SF, High NF, Low ND



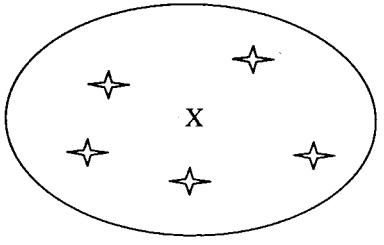
4) High SF, Low NF, Low ND



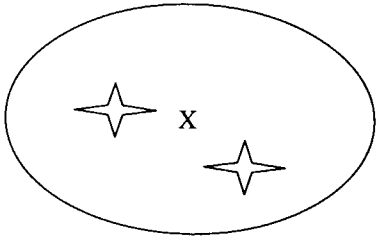
5) Low SF, High NF, High ND



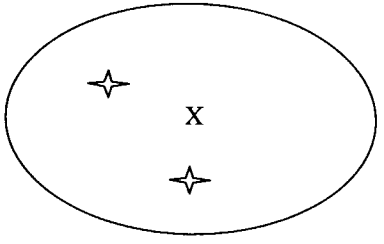
6) Low SF, Low NF, High ND



7) Low SF, High NF, Low ND



8) Low SF, Low NF, Low ND



Examples

- High log SF names (log SF > 7): Ventolin®, Dyazide®, Provera®
- Low log SF names (log SF < 3): Vistazine®, Antispas®, Protaphane®
- Name from a sparse neighborhood: Flexeril® (no neighbors in NAMCS/NHAMCS)
- Name from a dense neighborhood: Dynabac®, Synalar®, Rynatan®, Dynapen®, Dynacirc®, Dynacin®, Cynobac®

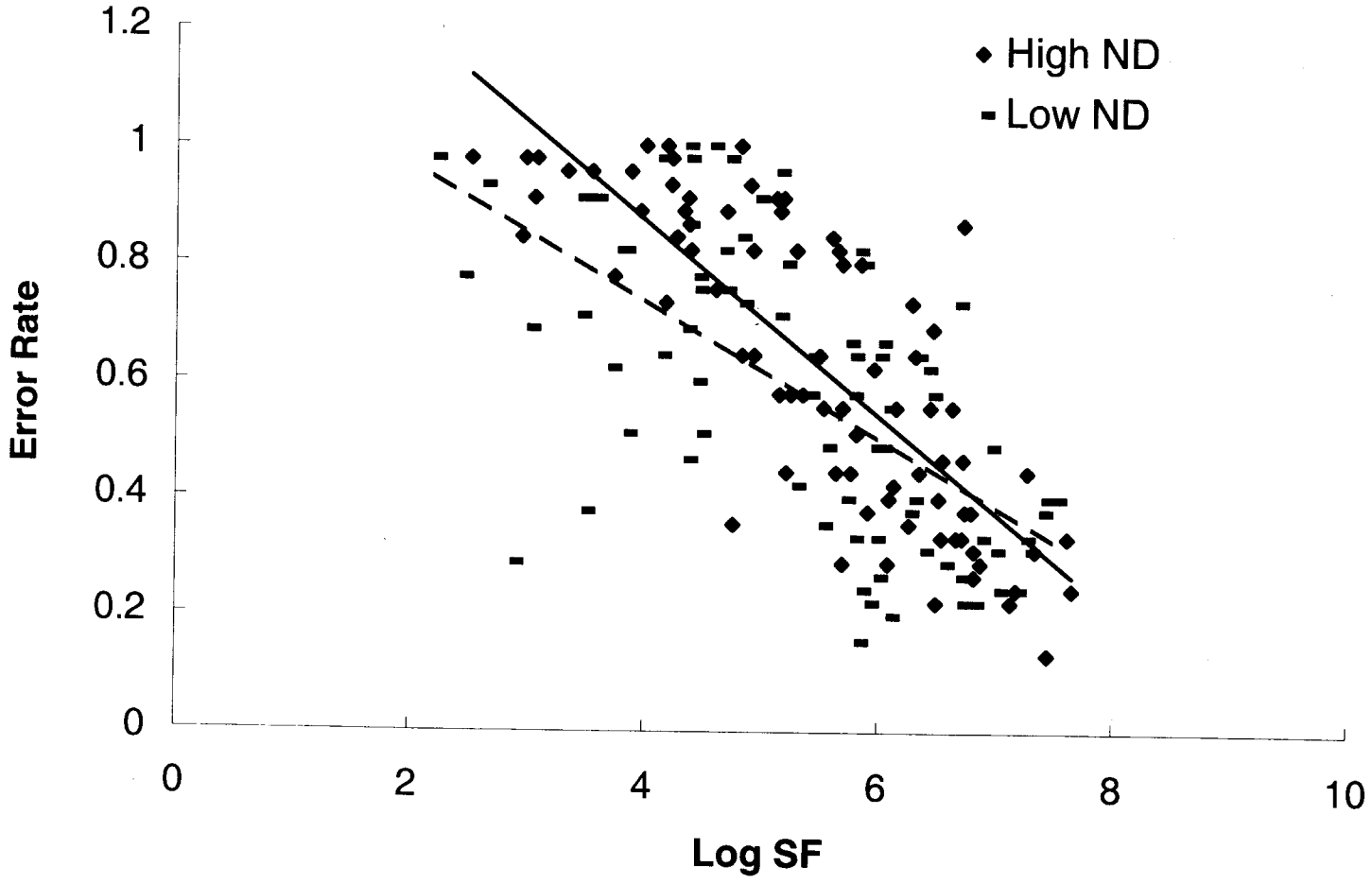


Figure 5. Effect of similarity neighborhood on RPh visual perception of drug names.

Lambert BL, Chang K-Y, Gupta P. Effects of frequency and similarity neighborhoods on pharmacists' visual perception of drug names. *Soc Sci Med.* in press.

Objective Measures: Conclusions

- They work.
- They are not perfect.
- Better on population basis than on individual basis
- Better for public health than for legal wrangling
- We should be using them.

Software Demonstration (time permitting)

- Name searching
  - Spelling
    - N-gram, edit distance
  - Pronunciation
    - N-gram, edit distance
- Product Searching
  - Name, dosage form, strength, route
  - Each weighted for importance
- Lambert BL, Yu C, Thirumalai M. A system for multi-attribute drug product comparison. *Journal of Medical Systems.* in press.
- Lambert BL, inventor. Apparatus, method, and product for multi-attribute drug comparison. US patent 6,529,892. March 4, 2003.

How can computer resources be used to calculate weights for various elements in name similarity?

- One way is to calculate a composite similarity score using multiple distinct similarity measures
- Use multiple measures to predict error probability or some other outcome
- Expert = 0.69 - 0.01*Editex - 0.30*NED + 0.22*Trigram2b - 0.02*EditSoundex

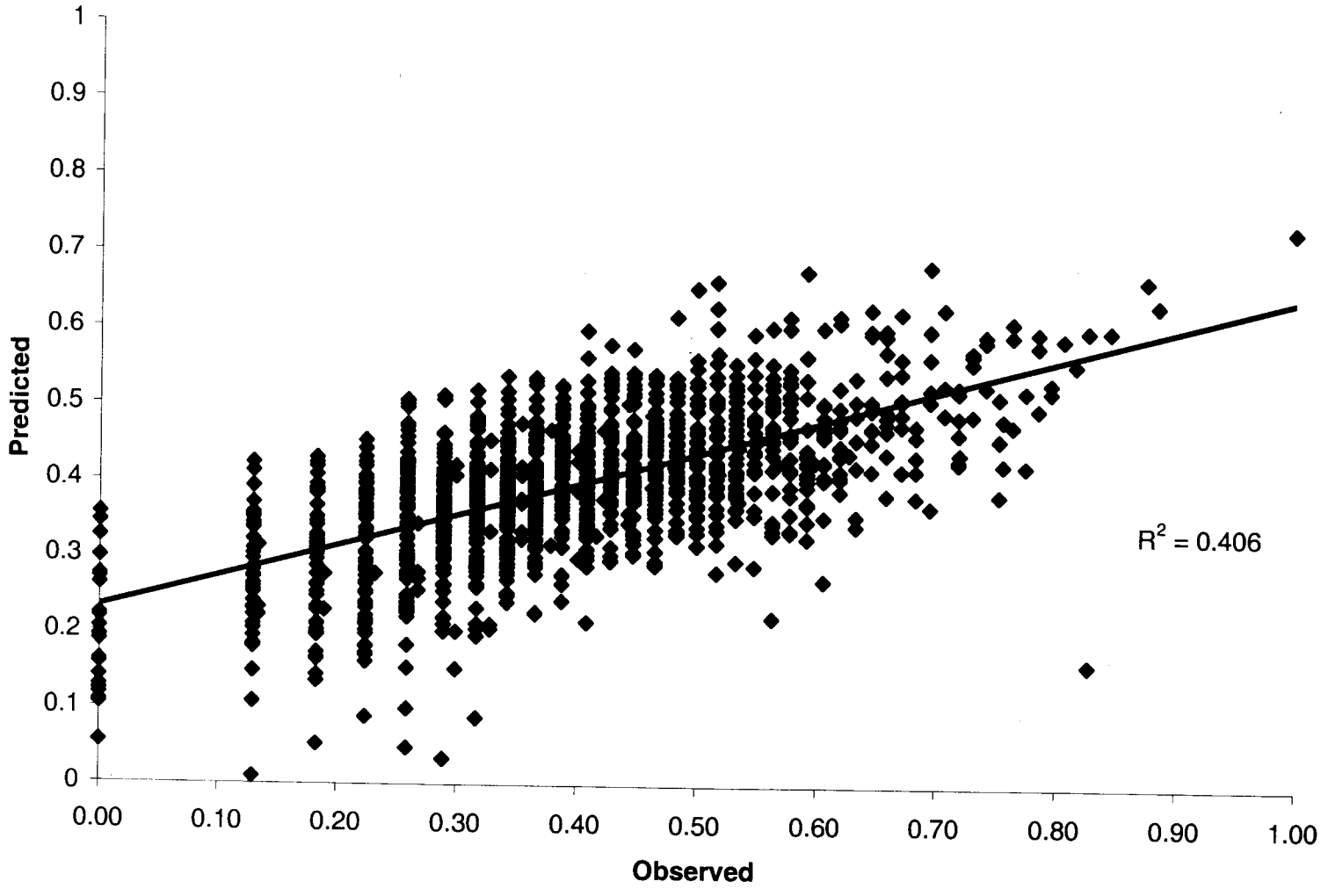


Figure 6. Using multi-measure regression model to predict expert similarity judgments.

Lambert BL, Yu C, Thirumalai M. A system for multi-attribute drug product comparison. *Journal of Medical Systems.* in press.

How can computer resources be used to calculate weights for various elements in name similarity?

- Compute distinct similarity score for each product attribute
  - Name, Dosage Form, Strength, Route of Administration, Schedule
  - Indication, Shape, Color, etc.
  - Use equivalence classes for approximate matching of attribute values (e.g., tablet and capsule)
- Use regression or other modeling techniques to assign weights to various attributes
- Use ISMP/MERP error data to estimate the importance of various attributes

Can computer assisted pattern recognition support the decision process to determine name/name similarities?

- Yes.
- General problem can be framed as a prediction problem
  - Given some predictors, predict some outcome
  - Predictors include product attributes and attribute similarity scores, as well as prescribing frequency
  - Outcome is probability of error or expert judgment of similarity or probability of FDA approval, etc.
- Many techniques exist for tackling these sorts of problems

Can computer assisted pattern recognition support the decision process to determine name/name similarities?

- Statistical modeling techniques
  - Regression
  - Discriminant analysis
- Machine Learning
  - Decision trees
  - Neural networks
  - Instance-based methods
  - Support vector machines, etc.

Can computer assisted pattern recognition support the decision process to determine name/name similarities?

- Problems
  - False positives
  - False negatives
  - Reliability of data for modeling (too often based on voluntary reports)
  - Determination of a threshold beyond which a name is "too confusing to approve"
  - Validation of predictive models

Summary

- How can computer resources be used to objectively measure differences between name pairs?
- How can computer resources be used to calculate weights for various elements in name similarity?
- Can computer assisted pattern recognition support the decision process to determine name/name similarities?
- Please contact me to discuss further. Thanks.