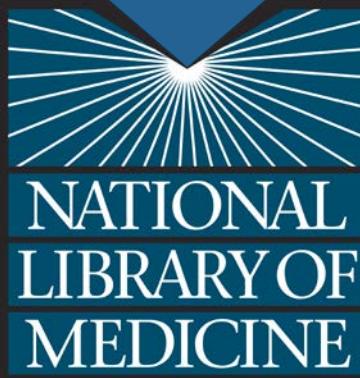


An Analysis of RxNORM Clinical and Branded Drug Similarity



Maria Yang

Lister Hill National Center for
Biomedical Communications



Outline

- What is RxNorm?
- Key Components of RxNorm Drugs
- What makes two drugs similar
- Gathering Data
- Transforming Raw Data to Usable Data
- Developing a Similarity Function
- Results
- So What?



String



diphenhydrAMINE Hydrochloride 12.5 MG Chewable Tablet [RxCUI = 1049900]

RxNorm Graph RxNorm Properties NDC RxTerms NDF-RT Pill Images Class View Interaction View

Views

- Classic
- Simple

Filters

- Human
- Vet
- Rx Pres

- Group
- Form

Links

Drug Label

MedlinePlus

Drug Portal

Legend

Multi Ingredient

Pack

IN/MIN

Ingredient (1)

diphenhydrAMINE

PIN

Precise Ingredient (1)

diphenhydrAMINE Hydrochloride

BN

Brand Name (2)

Benadryl

Kids-Eeze Allergy

SCDC

Clinical Drug Component (1)

diphenhydrAMINE Hydrochloride 12.5 MG

SBDC

Branded Drug Component (2)

diphenhydrAMINE Hydrochloride 12.5 MG [Benadryl]

diphenhydrAMINE Hydrochloride 12.5 MG [Kids-Eeze Allergy]



SCD/GPCK

Clinical Drug or Pack (1)

diphenhydrAMINE Hydrochloride 12.5 MG Chewable Tablet

SBD/BPCK

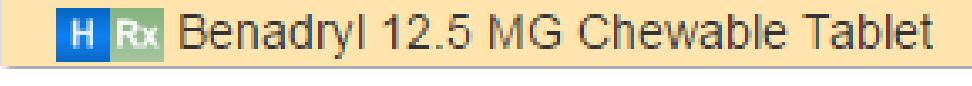
Branded Drug or Pack (2)

Benadryl 12.5 MG Chewable Tablet

Kids-Eeze Allergy 12.5 MG Chewable Tablet

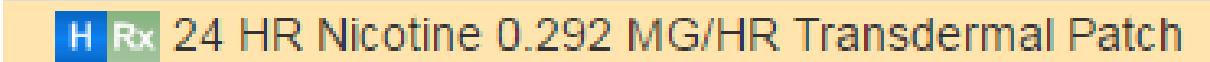
What is RxNorm?

- ▶ RxNorm is a database containing information on all medications available on the US Market
- ▶ We use it to extract data on all clinical and branded drugs to conduct comparisons
 - ▶ Semantic Clinical Drug (SCD)
A yellow rectangular box containing the text "H Rx diphenhydramine hydrochloride 12.5 MG Chewable Tablet". The "H" and "Rx" are in blue boxes, while the rest of the text is in a brown font.

H Rx diphenhydramine hydrochloride 12.5 MG Chewable Tablet
 - ▶ Semantic Branded Drug (SBD)
A yellow rectangular box containing the text "H Rx Benadryl 12.5 MG Chewable Tablet". The "H" and "Rx" are in blue boxes, while the rest of the text is in a brown font.

H Rx Benadryl 12.5 MG Chewable Tablet

Key Components of RxNorm Drugs

- ▶ Ingredient
 - ▶ Some drugs can contain multiple ingredients
- ▶ Dose Form
- ▶ Strength
- ▶ Quantity H Rx 24 HR Nicotine 0.292 MG/HR Transdermal Patch
- ▶ Quality
 - ▶ Very new factor and is available for a few select drugs
- ▶ Brand Name vs. Generic

What makes two drugs similar?

- ▶ An Ingredient Match
- ▶ Topical vs. Systemic
- ▶ Close Strengths
- ▶ Whether or not the drug contains multiple ingredients
- ▶ Close quantity values
- ▶ The same quality

Gathering Raw Data/Implementation

- ▶ Wrote Java code using the RxNorm RESTful API to extract SCDs and SBDs, access attributes and properties of each drug
- ▶ Isolate each component
 - ▶ RxCUI #
 - ▶ Multiple ingredient
 - ▶ Strength (keeping units)
 - ▶ Dose form RxCUI/Dose Form
 - ▶ Quantity
 - ▶ Quality
 - ▶ Brand Name RxCUI/Brand Name
- ▶ Write to text file delimited by tabs
- ▶ Export to Excel to have filtering functionality

Transforming Raw Data to Usable Data

- ▶ Ingredient - no transformation, exact match
- ▶ Dose form - binarization and Boolean evaluation
 - ▶ The dose form could either be topical or systemic
- ▶ Strength - quantitative measure
 - ▶ Ensure that units are the same
- ▶ Multiple ingredient - binarization and Boolean evaluation
 - ▶ Creation of Boolean value that says whether there are multiple ingredients or not
- ▶ Quantity - quantitative measure
- ▶ Quality - no transformation, exact match

Calculating the Components based on Exact Match

- ▶ If the Strings are exact matches, we give it the full weight
 - ▶ $ingredientScore = 1$
 - ▶ $qualityScore = 1$
- ▶ Otherwise, it doesn't get any of the weight
 - ▶ $ingredientScore = 0$
 - ▶ $qualityScore = 0$

Calculating the Components Based on Binarization and Boolean Evaluation

- ▶ To get binary dose form types, a hashmap was created with each Dose Form RxCUI associated with whether it was systemic or topical
- ▶ If the binary dose form types are the same, we give it the full weight
 - ▶ $doseFormScore = 1$
- ▶ Otherwise, it doesn't get any of the weight
 - ▶ $doseFormScore = 0$

Calculating the Components Based on Binarization and Boolean Evaluation

- ▶ The multiple ingredient Boolean is determined on if more than one ingredient were present or not. The exact number didn't matter.
- ▶ If the drugs were either both multiple ingredients or both single ingredients, we give it the full weight
 - ▶ $multipleIngredientScore = 1$
- ▶ Otherwise, it doesn't get any
 - ▶ $multipleIngredientScore = 0$

Calculating the Components Based on Binarization and Boolean Evaluation

- ▶ The binarization is whether the drug is branded or generic
- ▶ If the drugs were either both branded or both generic, we give it the full weight
 - ▶ $brandScore = 1$
- ▶ Otherwise, it doesn't get any
 - ▶ $brandScore = 0$

Calculating the Components Quantitatively (Strength, Quantity)

- ▶ We want a quantitative measure instead of an all-or-nothing method
- ▶ Compare the units on each drug's strength to determine if there's a match. If not, we set the strength factor to 0.
 - ▶ $strengthScore = 0$
- ▶ Otherwise, we want a fraction to measure how close two numbers are. If they're the same we want the function to return 1. The closer the numbers are, the closer to 1 the factor should be.

$$\frac{MIN(d_1, d_2)}{MAX(d_1, d_2)}$$

- ▶ $strengthScore = \frac{MIN(s_1, s_2)}{MAX(s_1, s_2)}$
- ▶ $quantityScore = \frac{MIN(q_1, q_2)}{MAX(q_1, q_2)}$

Developing a Similarity Function

- ▶ Determining weights for each component
 - ▶ Different weights for various use cases
 - ▶ Built separate weight file for user input
 - ▶ The weights that we went with:
 - ▶ Ingredient (50%)
 - ▶ Binary Dose Form (15%)
 - ▶ Strength (10%)
 - ▶ Multiple ingredient boolean (10%)
 - ▶ Quantity (5%)
 - ▶ Quality (5%)
 - ▶ Brand (5%)

Getting the Total Similarity Score

- ▶ Add up all the component scores that were multiplied by their weights to get the final similarity score
- ▶ $score = 100 * (ingredientScore * ingredientWeight + doseFormScore * doseFormWeight + strengthScore * strengthWeight + multipleIngredientScore * multipleIngredientWeight + quantityScore * quantityWeight + qualityScore * qualityWeight + brandScore * brandWeight)$

Dealing with Multiple Ingredients

- Decided to run the similarity test for each individual ingredient and take the highest match from all the comparisons.

RxCI	MultipleIngredient	Ingredient RxC	Ingredient	Strength	DoseForm	DoseFormBina	Quantity	Quality	BrandName
1485531	TRUE	82003	Docusate	55 MG	Oral Capsule		1		
1485531	TRUE	47628	tricalcium phosphate	160 MG	Oral Capsule		1		
1485531	TRUE	1151	Ascorbic Acid	25 MG	Oral Capsule		1		
1485531	TRUE	237099	alpha Tocopherol	30 UNT	Oral Capsule		1		
1485531	TRUE	684879	pyridoxine	25 MG	Oral Capsule		1		
1485531	TRUE	24941	Ferrous fumarate	30 MG	Oral Capsule		1		
1485531	TRUE	1006469	doconexent	265 MG	Oral Capsule		1		
1485531	TRUE	4511	Folic Acid	1.2 MG	Oral Capsule		1		
802792	TRUE	1151	Ascorbic Acid	50 MG	Oral Tablet		1		
802792	TRUE	2418	Cholecalciferol	400 UNT	Oral Tablet		1		
802792	TRUE	11248	Vitamin B 12	0.002 MG	Oral Tablet		1		
802792	TRUE	4511	Folic Acid	0.8 MG	Oral Tablet		1		
802792	TRUE	11246	Vitamin A	4000 UNT	Oral Tablet		1		
802792	TRUE	7405	Niacinamide	10 MG	Oral Tablet		1		
802792	TRUE	1925	Calcium Sulfate	250 MG	Oral Tablet		1		
802792	TRUE	21579	Copper Sulfate	0.15 MG	Oral Tablet		1		
802792	TRUE	24941	Ferrous fumarate	27 MG	Oral Tablet		1		
802792	TRUE	285155	Kelp preparation	0.01 MG	Oral Tablet		1		
802792	TRUE	6582	Magnesium Oxide	0.15 MG	Oral Tablet		1		
802792	TRUE	6585	Magnesium Sulfate	0.05 MG	Oral Tablet		1		
802792	TRUE	34323	potassium sulfate	0.84 MG	Oral Tablet		1		
802792	TRUE	684879	pyridoxine	1 MG	Oral Tablet		1		
802792	TRUE	9346	Riboflavin	2 MG	Oral Tablet		1		
802792	TRUE	39954	Zinc Sulfate	0.085 MG	Oral Tablet		1		

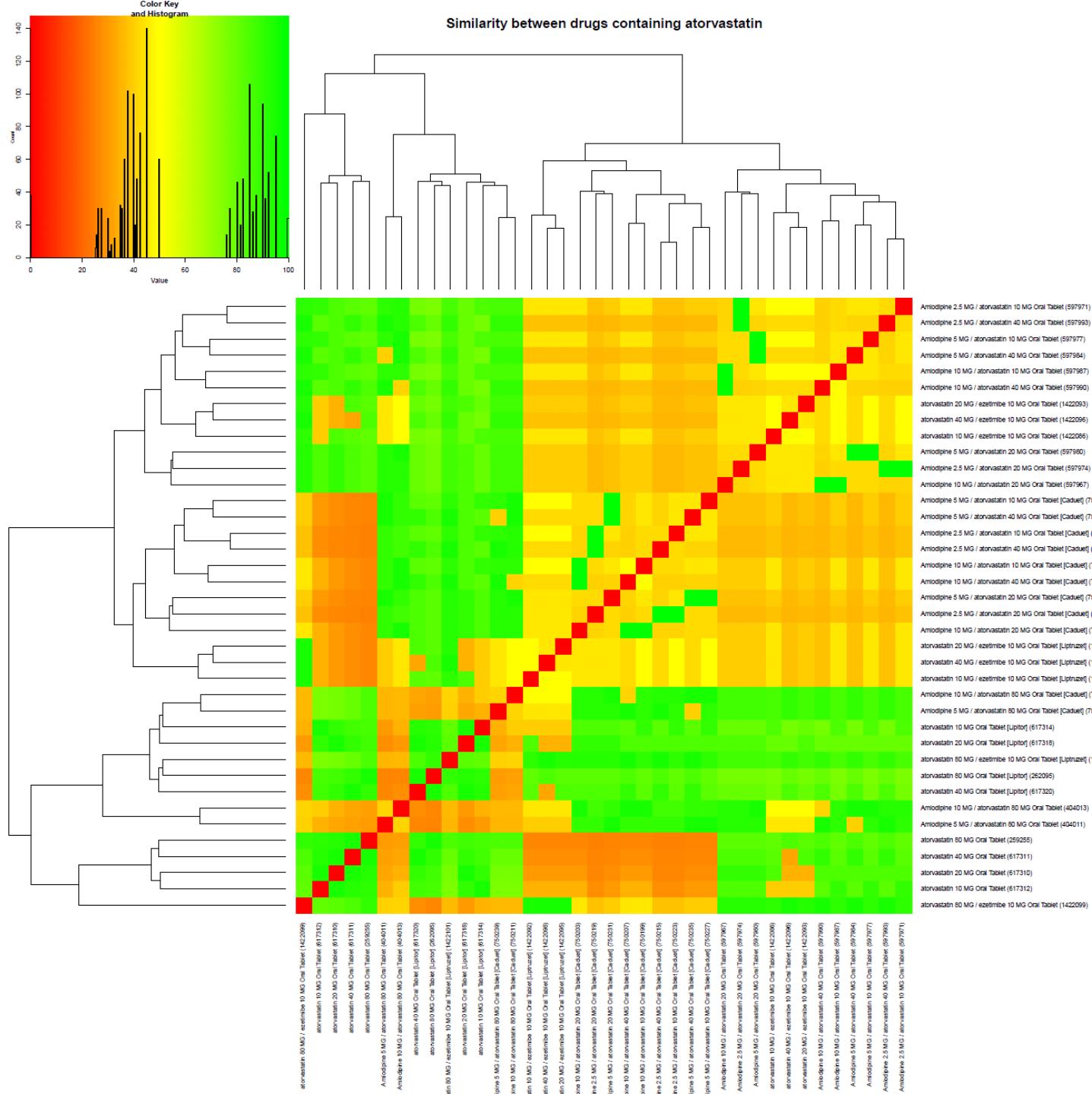
Results

Examples:

Drug 1	Drug2	Similarity Score
0.05 ML afibercept 40 MG/ML Injection	0.05 ML ranibizumab 10 MG/ML Injection	42.5%
0.2 ML Bemiparin sodium 12500 UNT/ML Prefilled Syringe	0.2 ML Bemiparin sodium 17500 UNT/ML Prefilled Syringe	90%
0.4 ML Iufenuron 100 MG/ML Prefilled Syringe	0.4 Methotrexate 18.8 MG/ML Auto-Injector	40%

So What?

- We were able to develop a proof-of-concept for getting similarity scores for drugs



Limitations

- ▶ We used arbitrary weights for each component
- ▶ Similarity function is application-dependent
 - ▶ Each use case may need different sets of weights
 - ▶ In the Java code, the weights are in a property file, not in the code

Acknowledgements

- ▶ Dr. Olivier Bodenreider
- ▶ Ms. Keyla Cooper
- ▶ ORISE



Questions?

