Frequency Estimates from Prescription Drug Datasets

205

Swu-Jane Lin, Ph.D.[1], Bruce Lambert, Ph.D.[1,2,3], Hiangkiat Tan, B. Pharm, M.S.[1], Sengwee Toh, B. Pharm.[1]
[1]Department of Pharmacy Administration, [2]Department of Pharmacy Practice, [3]Department of Communication, University of Illinois at Chicago

Background

- Accurate information about prescribing and dispensing frequency is important to marketers, policymakers, pharmacoepidemiologists, and patient safety researchers.
- There is no standard reference for drug frequency in the United States.
- Multiple data sources are available for use in estimating drug frequency.
- No previous study compares and contrasts the estimates derived from various prescribing frequency datasets.

Objective

- To quantify and explain similarities and differences between several well-known prescribing frequency data sets.

Methods

- Data sources -
  - National Ambulatory Medical Care Survey (NAMCS, year 2000),
  - National Hospital Ambulatory Medical Care Survey (NHAMCS, emergency department, ED, and outpatient department, OPD, year 2000),
  - IMS National Prescription Drug Audit (year 2000)
  - Solucient outpatient database (average between 2000 to 2002), and
  - Solucient inpatient database (average between July 1999 and June 2000).
- NAMCS and OPD of the NHAMCS were combined for all the analyses due to their similar nature.
- Comparisons – data collection methods, costs, overall drug frequencies, number of unique names, Pearson correlations between frequency estimates, extent of overlap in drug names, and nomenclature differences.
- Sensitivity analyses – frequency correlations were also computed by
  - Using only the one-word drug names
  - Converting all drug names into generic names
- NAMCS and NHAMCS can be downloaded for free from the National Center for Health Statistics. IMS and Solucient datasets cost tens of thousands dollars.

Results

- A total of 13,802 unique drug names were found in the 5 datasets.
- Overlap in drug names - Only 84 names appeared in all the 5 datasets; 702 appeared in all 4 outpatient datasets (i.e., NAMCS+OPD, ED, IMS, Solucient outpatient), and 1055 appeared in the three non-ER outpatient datasets (I.e., NAMCS+OPD, IMS and Solucient outpatient).

Results

Table 1. Annual drug frequency by dataset

Dataset	NAMCS*	NHAMCS* (OPD)	NHAMCS* (ED)	IMS*	Solucient Outpatient#	Solucient Inpatient+
Total weighted frequency (in million)	1263.5	129.9	173.5	3126.6	1280.4	31.9
Number of unique drug names	2187 (drug mentions, brand or generic names)	2273 (drug mentions, brand or generic names)	1721 (drug mentions, brand or generic names)	6400 (medications dispensed, brand or generic names)	5775 (medication claims, mostly brand names)	2448 (hospital administrative data, mostly generic names)
Projection	All U.S. visits to office-based physicians	All U.S. visits to outpatient departments of hospitals	All U.S. visits to emergency rooms of hospitals	All U.S. pharmacies	60% of the U.S. population covered by employment-based insurance	About 50% of all discharges in the U.S.

Table 2. Correlation between brand name frequencies for 5 datasets

	NAMCS +OPD 2000	ED 2000	IMS 2000	Solucient Outpatient	Solucient Inpatient
NAMCS+ OPD 2000	1.000 (n=2703)				
ED 2000	.635 (n=1349)	1.000 (n=1721)			
IMS 2000	.720 (n=1522)	.170 (n=1026)	1.000 (n=6400*)		
Solucient Outpatient	.708 (n=1166)	.133 (n=866)	.926 (n=2333)	1.000 (n=5572)	
Solucient Inpatient	.334 (n=211)	.237 (n=242)	.372 (n=295)	.241 (n=473)	1.000 (n=2448)

Table 3. Correlation between one-word drug name frequencies for five datasets

	NAMCS +OPD 2000	ED 2000	IMS 2000	Solucient Outpatient	Solucient Inpatient
NAMCS+ OPD 2000	1.000 (n=1958)				
ED 2000	.631 (n=1179)	1.000 (n=1233)			
IMS 2000	.720 (n=1447)	.171 (n=991)	1.000 (n=4902)		
Solucient Outpatient	.707 (n=1030)	.129 (n=723)	.924 (n=1931)	1.000 (n=2190)	
Solucient Inpatient	.332 (n=204)	.258 (n=179)	.374 (n=288)	.289 (n=211)	1.000 (n=633)

All the Pearson correlations were significant at the 0.001 level (2-tailed)

Correlation Between Frequencies for Generic Names

- **IMS 2000 and NAMCS+OPD 2000**
  - 701 common generic names yielded a Pearson correlation of 0.789.
  - After the removal of non-prescription drugs and vaccines, the remaining 574 common drug names gave a correlation of 0.915
- **Solucient Outpatient and NAMCS+OPD 20**
  - 433 common generic names yielded a Pearson correlation of 0.802
  - After the removal of non-prescription drugs and vaccines, the remaining 372 common drug names gave a Pearson correlation of 0.871

Results

Table 4. Ten most frequently mentioned drugs in NAMCS+OPD 2000 and their frequency (in thousands) in respective datasets

Drug	NAMCS +OPD 2000	ED2000	IMS 2000	Solucient outpatient average#	Solucient inpatient average^	Minimum	Maximum
Claritin	17145	453	30157	14474	N/A*	453	30157
Lipitor	16267	395	48826	24819	78	78	48826
Synthroid	15999	481	43971	19245	N/A	481	43971
Premarin	14775	268	46939	19641	N/A	268	46939
Amoxicillin	13068	2234	23053	12318	N/A	2234	23053
Tylenol	12789	9991	1738	0**	N/A	0	12789
Lasix	12577	1717	2115	395	N/A	395	12577
Celebrex	12161	537	24911	9976	44	44	24911
Glucophage	11468	268	27578	8021	N/A	268	27578
Albuterol	10862	3529	41947	11556	69	69	41947

Discussion

- Considerable differences in drug frequency estimates were found across the five commonly used datasets.
  - The meaning of frequency is different among the datasets.
    - ✓NAMCS and NHAMCS – frequency of a drug mentioned, prescribed, or given by physician
    - ✓IMS – frequency of a drug dispensed by pharmacy
    - ✓Solucient outpatient – frequency of a drug submitted for reimbursement
    - ✓Solucient inpatient – frequency of a drug given to patients
- Estimates of correlations between the datasets varied significantly.
  - Differences in the characteristics of datasets made it difficult to corroborate results or construct a more comprehensive picture of medication usage from multiple sources.
- Nomenclature inconsistency is a major obstacle for corroborating estimates among datasets.
  - Methods (e.g., RxNorm) to link across various decentralized databases with vastly different naming conventions are required.

Conclusions

- Estimated *absolute* frequency differs significantly between datasets due to differences in nomenclature, sampling and measurement methods.
- *Relative* frequency is quite consistent between the three outpatients datasets of NAMCS+OPD, IMS and Solucient outpatient.
- Researchers must understand precisely which quantity they are interested in and how that compares to the specific quantities estimated by these different datasets.

UIC University of Illinois at Chicago College of Pharmacy

Contact Information: Swu-Jane Lin, Ph.D.
Address: 833 S. Wood Street, Rm. 241 (M/C 871), Chicago, IL 60612-7231
Email: slin5@uic.edu Phone: 312-355-5666 Fax: 312-996-0868