Frequency Estimates from Prescription Drug Datasets

(Revision of #04-11-066A)

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

Running Head: Prescription Frequency

Corresponding author: Swu-Jane Lin, Ph.D.

Address: 833 S. Wood Street, Rm. 241 (M/C 871), Chicago, IL 60612-7231

Phone: 312-355-5666

Fax: 312-996-0868

Email: slin5@uic.edu

Abstract

Purpose: Accurate information about the number of times a drug is prescribed or dispensed annually is important to marketers, pharmacoepidemiologists, and patient safety researchers. Yet there is no standard reference for prescribing frequency data. The multiple sources that do exist vary in their sampling methods, target populations, nomenclature, and methods of tallying individual medications prescribed or dispensed. These differences are likely to create ambiguity and contradictions in the scientific literature, but they are not well understood.

Methods: We conducted a descriptive study to examine the similarities and differences between 5 well-known sources of prescribing frequency data: the National Ambulatory Medical Care Survey (NAMCS), the National Hospital Ambulatory Medical Care Survey (NHAMCS, emergency department and outpatient department), the IMS National Prescription Drug Audit, the Solucient outpatient dataset, and the Solucient inpatient dataset. We compared survey methods, costs, overall frequencies, number of unique names in each database, correlations between frequency estimates from different databases, the extent of overlap in the databases, and nomenclature differences between and within datasets.

Results: All the correlations between frequency estimates derived from different datasets were significant, but the frequency estimates differed considerably. The lowest correlation (0.17) was found between the IMS and emergency department of the

NHAMCS, and the highest correlation (0.93) was between IMS and Solucient outpatient data.

Conclusions: Although there were significant correlations between frequency estimates for comparable datasets, sampling methods and nomenclature choices resulted in important differences both for individual drug products and for overall frequency statistics. Researchers need to be aware of the differences when deriving drug frequency with these datasets.

Key Words: prescription, frequency, pharmacy, NAMCS, NHAMCS, IMS, Solucient

Word Count (main body text): 3620

Take-home messages:

1. Correlations between frequency estimates from public and proprietary drug datasets range from small to large.

2. High correlation was found between proprietary datasets. Moderate correlation was found between public and proprietary datasets.

3. Correlations between drug frequency datasets appear to depend on sampling frame, data collection method, and nomenclature differences.

4. Total prescribing frequency estimated from the five datasets differs significantly.

5. Each of the five prescription drug datasets provides 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.

Introduction

Knowing how frequently a drug product is prescribed or dispensed is important for market planning, evaluating market performance and sales, and designing strategies for research and development of new drugs. It is also an important factor to consider in pharmacoepidemiological research and in health policy decision-making. In the study of medication errors, combining prescribing frequency information with accurate estimates of error rates can produce estimates of the total number of errors committed per year. Frequency is also the single most important psycholinguistic variable affecting errors in word memory and perception,[1-10] and can play an important role in predicting look-alike and sound-alike medication errors. Generally speaking, common words are easier to recall and recognize than rare words. When two drugs are involved in a look-alike sound-alike mistake, it is often the case that the high-frequency drug was dispensed when a low-frequency drug was actually prescribed.

Our interest in drug frequency originated in medication errors involving drug name confusion, but soon we found that, unlike common English words, there is no consistent estimate on drug name frequency. Despite the relevance and usefulness of information about medication prescribing frequency, there are relatively few datasets with such information in the United States. Unlike countries with single-payer systems (e.g., Canada), the lack of a national healthcare system and a comprehensive healthcare dataset in the United States makes it almost impossible to obtain accurate and timely

5

information on prescription drugs used by the entire U.S. population at all healthcare settings.

Several public and proprietary datasets have been used by previous studies to approximately estimate prescribing frequency in the United States. The National Ambulatory Medical Care Survey (NAMCS) and National Hospital Ambulatory Medical Care Survey (NHAMCS) are two public datasets that collect information on the use of medications.[11] One of the most commonly used proprietary data sources is from the IMS Health,[12] which has been used by the pharmaceutical industry, researchers, and government agencies to estimate the national sales of pharmaceuticals. Solucient, another commercial source that claims to be maintaining the largest healthcare database in the United States, also provides prescription drug frequency and other healthcare information to thousands of clients.[13] Although these datasets all provide information on frequency of drug utilization, the specific methods used to collect the data might cause a difference in frequency estimates. This study was conducted to quantify and explain similarities and differences between these well-known and widely-used public and proprietary prescribing frequency datasets.

Methods

Data sources

Two publicly available datasets, NAMCS and NHAMCS, and three commercial datasets were included in this study. The commercial datasets were the IMS National Prescription Audit Plus (outpatient data, denoted as 'IMS' from here on), the Solucient Hospital Drug Utilization Database (inpatient data, denoted as 'Solucient inpatient' from

here on), and the Solucient Claims Data Warehouse (outpatient data, denoted as 'Solucient outpatient' from here on).[14-16] Both the NAMCS and NHAMCS are national probability sample surveys conducted annually by National Center for Health Statistics (NCHS), Center for Disease Control and Prevention.[11]

At the time the data were obtained (fall 2003), the latest data from NAMCS and NHAMCS were from the year 2000. NHAMCS dataset includes two subsets: one was based on visits made to outpatient department of hospitals (OPD), and the other included only visits made to emergency rooms of hospitals (ED). Although IMS had more current data, for consistency we used only the year 2000 data in our comparisons. Solucient drug frequency data were provided in aggregate over a three-year span (three year aggregate from 2000 to 2002 for Solucient outpatient and from July 1999 to June 2002 for Solucient inpatient).

In our preliminary analysis, we found that the OPD data from NHAMCS were more similar in nature to the NAMCS than to the ED data from NHAMCS. The OPD part of NHAMCS was therefore combined with the NAMCS data and referred to as "NAMCS+OPD". We treated ED as a standalone dataset for correlational analyses on drug frequency.

Generic and brand names were used inconsistently by the datasets. The 'drug mentions' of NAMCS and NHAMCS and drug names in IMS datasets contained a combination of generic and brand names. Solucient outpatient dataset listed both generic and brand names side by side, and Solucient inpatient dataset consisted mostly of generic names.

Comparisons

Projected national frequency estimates, data collection method, time lag for updates, number and types of drugs, and cost to use the datasets were compared. Pearson correlation among drug frequency was computed using the names listed in the five datasets. Due to the heterogeneity in nomenclature among the datasets, we performed two additional correlation studies as sensitivity analyses: one converting all drug names into generic names and the other using only the one-word drug names.

The conversion to generic names was carried out by linking drug names to the FDA Orange Book[17] and the generic-brand names lookup table compiled from NAMCS and NHAMCS documentation. Analyses based on these derived generic names presumably provide an estimate of maximum possible correlations when nomenclature differences are ignored. The exclusion of non-one-word drug names eliminates the variation between datasets in coding components of drug names (e.g., whether salt or strength was included as part of a drug name).

Correlations only provide information on the similarity between datasets in the *ranking* of drug frequencies, but not in the *estimates* of drug frequency. A list of 20 most frequently prescribed medications in NAMCS+OPD was used to compare the differences in frequencies estimated by the five datasets and to examine whether the differences followed any discernable pattern.

Sampling Frame

NAMCS and NHAMCS. NAMCS is an annual survey with a national probability multistage-sample design that involves samples from primary sampling units (PSUs), physician practices within PSUs, and patient visits within physician practices. The target universe of the NAMCS is visits made to the offices of non-federally employed physicians in the United States. The sampling method of NHAMCS is a four-stage probability design that involves samples of PSUs, hospitals within PSUs, clinics within hospitals, and patient visits within clinics. The NHAMCS includes an outpatient clinic dataset (OPD) and emergency room visits dataset (ED). Medications collected in the NAMCS and NHAMCS are called "drug mentions" which included up to 6 medications that "were ordered, supplied, administered or continued" during a visit. These mentions could be prescription drugs, OTC medications, immunizations, allergy shots, anesthetics, and non-pharmaceutical products.[11]

IMS. The sampling frame for IMS includes chain, independent, and food store pharmacies, as well as mass merchandisers and discount houses. Not included in the sampling frame are HMO pharmacies that serve HMO members only, dispensing physicians, hospital pharmacies, clinic pharmacies, and home healthcare. A stratified random sample of about 20,000 networked drug stores is taken from the 29,000 reporting stores in IMS Health's pharmacy database through stratifying pharmacies by region, type, and size. Every new and refilled prescription is collected each day from the sample

pharmacies. Altogether, the sample pharmacies constitute more than half of all retail

pharmacies in the US.[18]

Solucient Outpatient. The sampling frame of Solucient outpatient data is a

claims data warehouse. A sample of 1.2 million covered lives is drawn from the universe

of 170.47 million covered lives to produce a nationally representative sample. With

projection, the sample can reflect the utilization of health services by people covered

under the employment-based, private health insurance, which constituted about 60% of

the U.S. population in 2003.

Solucient Inpatient. The sampling frame in this dataset is all discharges from

short-term, general, non-federal hospitals in the US. In the year 2000, nearly 150

healthcare organizations and 2.1 million discharges from a universe of 35.4 million

discharges nationally were included in the dataset. The sampled discharge data are

projected to reflect approximately 20 million, or more than 50% of all-payer discharges

annually in the U.S.

Time Lag of Updates

Solucient datasets are updated quarterly, with a time lag of approximately 6

months. NAMCS and NHAMCS are released annually with a time lag of about 2 years

from the end of a survey year to the time the datasets are released to the public. There is

almost no time lag for IMS.

Reliability of Estimates

In both NAMCS and NHAMCS, estimates that are based on fewer than 30

records or have a Relative Standard Error (RSE) of greater than 30% are considered

unreliable. The RSE is calculated by dividing the standard error (sampling error) by the estimate itself. [19, 20] The IMS dataset considers any medication with fewer than five thousand projected prescriptions per month as unreliable. Reliability data for Solucient estimates were not provided in the documentation.

Costs

NAMCS and NHAMCS data can be freely downloaded from the web site of NCHS. One CD-ROM of the survey data can also be requested without charge. IMS and Solucient (includes both the inpatient and outpatient) datasets can easily cost tens of thousands of dollars each, depending on the amount and type of data requested and the availability of discounts.

Number and Type of Drugs

NAMCS 2000. There were 2187 unique drug mentions in the dataset, with 74.2% of them were prescription drug names, 18.2% were OTC medications, and 7.6% were neither (e.g., shampoo). After weighting, these names produced a total frequency of 1263.5 million 'drug' mentions for all visits made to physician's office in year 2000.

NHAMCS 2000. The ED subset had 1721 unique drug mentions and a weighted drug frequency of 173.5 million. The OPD subset had 2273 unique drug mentions, which, after weighting, gave a drug frequency of 129.9 million.

IMS 2000. In the year 2000, there were 6400 drug names listed, but only 3990 had a frequency greater than 0. The dataset estimates are by the thousands; therefore, drugs with fewer than a thousand prescriptions in the reporting period were assigned a frequency of zero. The weighted frequency gave an estimate of 3126.6 million

prescriptions, which represents the total number of prescriptions dispensed at pharmacies in the year.

       Solucient Outpatient. The Solucient outpatient dataset had 5572 unique brand names, which together produced a weighted prescription number of 3841.2 million from January 2000 to December 2002. Dividing the frequency by three, we obtained an estimate of 1280.4 million prescriptions, which represents the total number of prescriptions submitted every year for reimbursement purposes by the 60% of the US population who were covered under employment-based health insurance.

       Solucient Inpatient. The inpatient prescription drug dataset compiled data from July 1999 to June 2002. There were 2448 drug names (mostly generic names) in the dataset, which constituted a total frequency of 95.8 million over the three-year period. If divided by three, the estimated number of 31.9 million annual prescriptions represents the use of medications by the 20 million discharges included in the sampling frame.

Shared Drug Names Among the Datasets

The five datasets altogether listed 12,747 unique drug names. One-quarter of the drug names appeared only in the IMS 2000 dataset, one-fifth only in Solucient outpatient, 14% only in Solucient inpatient, 6% only in NAMCS+OPD and less than 2% only in ED. Only 84 names were common to all five datasets. If only considering the outpatient datasets (i.e., NAMCS+OPD, ED, IMS, Solucient outpatient), 702 drug names were found to be in all four datasets. Further removing the ED dataset from consideration, 1055 drug names were found to be present in NAMCS+OPD, IMS, and Solucient outpatient. IMS and NAMCS+OPD shared 1522 names, which represented only 23.8% of the 6400 IMS

names and 56.3% of the 2703 NAMCS+OPD names but they constituted 77.6% of IMS

and 81.9% of NAMCS+OPD total drug name frequency. A similar situation was

observed between drug names common to Solucient outpatient and NAMCS+OPD

datasets. Together, the statistics indicate that, despite the relatively small number of

names in NAMCS+OPD dataset, they have considerable representation in the two

commercial datasets in terms of the total number of prescriptions.

Correlation of Drug Frequency by Listed Drug Names in the Five Datasets

Table 1 gives the correlation coefficients and the corresponding number of names

shared by datasets. All the correlations were significant at the $p < 0.001$ level.

Insert Table 1 about here.

NAMCS+OPD has a moderate correlation with the other three outpatient datasets

(i.e., ED, IMS, and Solucient outpatient), which ranged from 0.635 (ED) to 0.720 (IMS).

The ED dataset derived from the NHAMCS had a moderate correlation with

NAMCS+OPD, but weak correlations with all the other datasets (ranging from 0.133

with Solucient outpatient to 0.237 with Solucient inpatient). The two commercial

outpatient datasets, IMS and Solucient, were highly correlated ($r = 0.926$). As expected,

the only inpatient dataset (Solucient inpatient) did not correlate well with any of the other

datasets. The highest correlation was with IMS ($r=0.372$), and the lowest was with ED

($r=0.237$).

Correlations of Drug Frequency by Generic Names

IMS 2000 and NAMCS+OPD 2000. After converting all the drug names into generic names, 701 generic names were common in both the IMS and NAMCS+OPD datasets, which altogether represented 67.2% of total IMS prescriptions and 96.4% of total NAMCS+OPD drug mentions. The correlation of these generic names was 0.789. After removing the non-prescription drugs and vaccines from both datasets, 574 common drug names remained, yielding a correlation of 0.915 in their frequencies.

Solucient Outpatient and NAMCS+OPD 2000. There were 433 generic names that appeared in both the Solucient outpatient and NAMCS+OPD datasets. The correlation between the shared generic names of the two datasets was 0.802; after the removal of non-prescription drugs and vaccines from the NAMCS, the correlation increased to 0.871. Although the number of the matched generic names was only 372, it represents 43.7% of total Solucient outpatient prescription frequency and 50.0% of the NAMCS+OPD drug mentions.

Correlation Between Frequencies of One-Word Drug Names

One-word drug names constituted more than 70% of the drug names in NAMCS+OPD 2000, ED 2000, and IMS 2000 datasets, but only around 38% in Solucient outpatient and 26% in Solucient inpatient. In terms of frequency, these one-word drug names constituted more than 86 to 89% in the NAMCS+OPD 2000, ED 2000, and IMS 2000 datasets; 65% in Solucient outpatient; and 31% in Solucient inpatient.

Although one-word drug names constituted only a portion of total number of drug names and prescription frequency in the original datasets, the magnitude of the

correlations between these subsets differ only slightly from the previous analysis using complete datasets. The results suggest the representativeness of one-word drug names in the frequency database.

Comparisons of Drug Frequency

Table 2 lists the 20 most frequently mentioned drug names in the NAMCS+OPD 2000 and frequency estimates from the five datasets. All of the 20 drugs appeared in ED 2000, 19 appeared in both IMS and Solucient outpatient datasets, but only 8 appeared in Solucient inpatient dataset. The only drug missing from the IMS and Solucient outpatient datasets is the influenza vaccine, which is an injectable product often given at clinics but rarely dispensed directly to patients. Two significant differences in the frequency of *Tylenol*[®] and *Lasix*[®] were found between NAMCS+OPD, IMS, and Solucient outpatient datasets. Although *Tylenol*[®], a brand name of an OTC analgesic, is frequently mentioned in visits made to the physician's office, it does not necessarily lead to a prescription. It is also very possible that when *Tylenol*[®] was prescribed, a generic version could be dispensed, which causes a discrepancy between prescription and dispensing frequency. A similar reason could also lead to the discrepancy observed in *Lasix*[®] between datasets. Both *Lasix*[®] and furosemide (generic name of *Lasix*[®]) were listed in all the five datasets; however, while 85% of the products were listed as *Lasix*[®] in NAMCS+OPD, only 5% and 6% were listed as such in IMS and Solucient outpatient.

Insert Table 2 about here.

Overall Advantages and Disadvantages

The major characteristics as well as advantages and disadvantages of the five drug frequency datasets are given in Table 3. The most significant advantages of NAMCS and NHAMCS are the low cost, and the opportunity to study medications being mentioned in physicians' offices and the co-prescribing of medications in a patient visit. Timeliness and accurate sales data are the primary reasons for using IMS data, and Solucient inpatient is the only dataset in our study that provided information on inpatient medication use.

Discussion

We purposefully selected and examined five datasets that have been commonly used by researchers in deriving prescribing patterns, drug utilization trends, total sales, prescription rate, etc., despite their vast difference in sampling frame, sample size, data collection methods. To our knowledge, there is no published study that compares and contrasts drug frequency estimated from these datasets. We found that although the ranking of drug frequency is relatively similar across datasets, there were considerable differences in estimates of drug frequency and inclusion of drug names. The frequencies, types, and names of drugs included in each prescription dataset are mostly determined by where and how the prescription information was collected.

So which dataset provides the most accurate estimates of drug frequency? From our results, it appears that the IMS dataset is more close to the commonly agreed upon estimate of roughly 3 billion outpatient prescriptions per year in the United States. However, it should be noted that this estimate of 3 billion was in fact derived from the

IMS data. There is no independent source to confirm whether this is a correct estimate of total prescribing frequency in the United States. Moreover, there is also the problem of defining what is meant by "prescribing frequency" or "total prescriptions".

NAMCS and NHAMCS measure drug mentions, which are drugs prescribed, continued, supplied or administered in a visit. IMS measures prescriptions actually dispensed at outpatient pharmacies. Solucient outpatient measures insurance claims for prescriptions. Discrepancy in frequencies estimated by these different data sources can happen when a prescription written for a patient in a visit was not filled, or a prescription filled was not claimed for reimbursement. Also, a prescription drug 'mentioned' in a patient visit might be filled for three times in a pharmacy and lead to three claims. Vaccines administered to patient in physician's office might be captured by the NAMCS but most likely would not be included in IMS dispensing data. Not to mention that Solucient outpatient dataset is based only on information from people covered by employment-based health insurance and does not provide valid means to derive national statistics on drug utilization. When these datasets refer to prescribing frequency or total prescriptions, they are not referring to the same underlying quantity.

Furthermore, in any given year, a significant portion of NAMCS drug mentions were coded as 'illegible', for example, 7.5% of drug mentions in the year 2000 NAMCS were illegible and could not be categorized by any drug characteristics (e.g., brand name or generic drug, prescription drug or OTC). Similarly, while IMS data can accurately capture prescriptions dispensed at outpatient pharmacy, it could not capture the drug samples supplied to patients by physicians. According to a report in 2004, the retail value

of free samples given to physicians for promotion purpose in the United States could amount to $16.4 billion,[21] when the total expenditure on prescription drugs of the year was about $164 billion.[22]

Another cause of the divergent estimates of drug frequency was the different naming conventions used by the five datasets that we compared. Different datasets adopt different methods of abbreviating drug names and include different components (salt, strength, formulation) in drug names. These differences will continue to pose a problem when researchers attempt to corroborate results or construct a more comprehensive picture of medication usage from multiple sources. A possible solution is to use tools that can link across and map major proprietary and public healthcare datasets. The clinical drug nomenclature of RxNorm database developed by the U.S. National Library of Medicine is one such effort to tackle the problem of different naming and coding conventions.[23]

There are still other dimensions of prescribing frequency that these databases do not address. For example, how many unique patients were exposed to a given drug in a given year? How many new prescriptions were written? How many refills? How many milligrams of the active ingredient were consumed per person per year? How many units of the drug product were manufactured or shipped to wholesalers in a given year? How many different physicians prescribed the drug? Hence, prescribing frequency can mean different things to different people and in different datasets. With this study, we attempted to draw attention to the diversity of meanings, methods, and measures associated with this basic concept and to caution researchers to keep this diversity in

mind when analyzing and interpreting results that incorporate one or another measure of prescribing frequency.

There are several limitations to this study. The two Solucient datasets we acquired are aggregated frequencies over a three-year period. This limitation makes the datasets less comparable with other datasets in terms of time horizon. Comparability problems also may result from differences in reliability between and within datasets. When estimating frequency and performing correlations, all the available drugs in each dataset, regardless their degree of reliability or lack of reliability data (as in the case of Solucient datasets), were included in analyses. The readers should also note that the results from the commercial sources were based on the specific data "cuts" we obtained, which may be different from what the owners of the datasets can potentially provide. More detailed information, including information related to the drug products, prescribers or pharmacies, may be available from the data providers. Most of the analyses were based on year 2000 data, which are almost five years old now and may not reflect the more recent trends.

Conclusion

Although there are significant correlations between datasets in their estimates of drug frequency, the estimates differ considerably due to the different methods employed to sample and measure drug frequency. Researchers need to be aware of the differences when using the datasets to derive drug related statistics.

References

1.   Glanzer M, Bowles N. Analysis of the word-frequency effect in recognition memory. *J Exp Psychol [Hum Learn]* 1976; **2**:21-31.

2.   Lambert B, Chang K, Gupta P. Effects of frequency and similarity neighborhoods on pharmacists' visual perception of drug names. *Soc Sci Med* 2003; **57**:1939-1955.

3.   Solomon R, Postman L. Frequency of usage as a determinant of recognition threshold for words. *J Exp Psychol* 1952; **43**:195-210.

4.   Hall J. Learning as a function of word frequency. *Am J Psychol* 1954; **67**:138-140.

5.   Monsell S. The nature and locus of word frequency effects in reading. In: Besner D, Humphreys G, eds. Basic processes in reading: Visual word recognition. Hillsdale, NJ: Erlbaum, 1991:148-197.

6.   Howes D. On the relation between the intelligibility and frequency of occurrence of English words. *J Acoust Soc Am* 1957; **29**:296-305.

7.   Grainger J. Word frequency and neighborhood frequency effects in lexical decision and naming. *J Mem Lang* 1990; **29**:228-244.

8.   Gregg V. Word frequency, recognition, and recall. In: Brown J, ed. Recall and recognition. New York: John Wiley & Sons, 1976:183-216.

9.   Savin H. Word-frequency effect and errors in the perception of speech. *J Acoust Soc Am* 1963; **35**:200-206.

10. Hulme C, Roodenrys S, Schweickert R, Brown G, Martin S, Stuart G. Word-frequency effects on short-term memory tasks: Evidence for a redintegration process in immediate serial recall. *J Exp Psychol Learn Mem Cogn* 1997; **23**:1217-1232.

11. National Center for Health Statistics. Ambulatory Health Care Data: Survey Methodology. Available at http://www.cdc.gov/nchs/about/major/ahcd/ahcd1.htm [Accessed 8 November 2004].

12. IMS Health. IMS. Available at http://www.imshealth.com/ [Accessed 8 November 2004]

13. Solucient LLC. About Solucient. Available at http://www.hcia.com/aboutus/aboutus.shtml [Accessed 8 November 2004]

14. IMS Health. NPA Plus: National prescription audit. IMS Health, Inc., Westport, CT, 2003.

15. Solucient Inc. Hospital drug utilization database. Available at http://www.solucient.com/solutions/Solucients_Databases.shtml [Accessed 8 November 2004]

16. Solucient Inc. Claims Data Warehouse. Available at http://www.solucient.com/solutions/Solucients_Databases.shtml [Accessed 8 November 2004]

17. Food and Drug Administration. Approved Drug Products with Therapeutic Equivalence Evaluations. Rockville, MD: FDA, CENTER for Drug Evaluation and Research, 2004. Available at

http://www.fda.gov/cder/ob/docs/preface/ectablec.htm [Accessed 8 November 2004]

18. IMS Health. IMS HEALTH Information Services Manual. IMS Health, Inc., Westport, CT, 2002.

19. Cherry D, Woodwell D. National Ambulatory Medical Care Survey: 2000 summary. *Adv Data* 2002:1-32.

20. Ly N, McCaig L. National Hospital Ambulatory Medical Care Survey: 2000 outpatient department summary. *Adv Data* 2002:1-27.

21. Anonymous. Doctors not influenced by pharmaceutical marketing tactics. Medical News Today, Dec 8, 2004. Available at http://www.medicalnewstoday.com/newssearch.php?newsid=17469 [Accessed June 10, 2005]

22. IMS Health. Tracking 13 key global Pharma markets. IMS retail Drug Monitor. Available at http://www.imshealth.com/vgn/images/portal/cit_40000873/56/40/73661312IMS _Retail_Drug_Monitor_Feb05.pdf [Accessed 10 June 2005]

23. National Library of Medicine. A guide to RxNorm. Available at http://www.nlm.nih.gov/research/umls/rxnorm_main.html [Accessed March 15, 2005]

Table 1. Frequency correlations of brand names among the five 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)

* Including those with zero frequency.
All the Pearson correlations were significant at the 0.001 level (2-tailed).

Table 2. The 20 most mentioned drugs in NAMCS+OPD 2000 and their frequency (in

thousands) in respective datasets

Drug	NAMCS+OPD 2000	ED2000	IMS 2000	Solucient outpatient average [a]	Solucient inpatient average [b]	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
Vioxx	10801	373	20630	9838	45	45	20630
Prilosec	10751	754	32021	14497	N/A	754	32021
Norvasc	10635	401	31108	11090	N/A	401	31108
Atenolol	10372	349	38888	14438	110	110	38888
Influenza Vaccine	10197	26	N/A	N/A	N/A	26	10197
Prednisone	10049	1873	20719	8356	124	124	20719
Amoxil	9719	1170	13746	5678	N/A	1170	13746
Prevacid	9268	415	25299	11739	N/A	415	25299
Zocor	9202	197	22348	10694	112	112	22348
Zoloft	9183	289	25699	12847	66	66	25699

[a] Three-year average of frequency for people covered by employment-based insurance
(about 60% of total US population)
[b] Three-year average of frequency for about 50% of total discharges in the US
* N/A: not in the dataset
** Less than 1000

Table 3. Advantages and disadvantages of the five drug frequency datasets

Characteristics	NAMCS and NHAMCS	IMS	Solucient outpatient	Solucient inpatient
Total weighted frequency (in millions)	NAMCS 1263.5 NHAMCS OPD 129.9 NHAMCS ED 173.5	3126.6	1280.4 [a]	31.9 [b]
Number of unique drug names	NAMCS 2187 NHAMCS OPD 2273 NHAMCS ED 1721	6400	5775	2448
Population	All U.S. visits to office-based physicians, hospital clinics and ED - good for understanding the drug prescribed or given at the particular setting	All U.S. pharmacies – good for studying all medications dispensed by pharmacy	People with employment-based insurance – good for studying claims data for medications	Hospitalizations – good for understanding inpatient drug use
Costs	Free	Expensive	Expensive	Expensive
Acceptance by research community	Widely used by researchers—good for corroborating study results with previous research	Widely used by the industry, some government agencies, and researchers	Widely used by the industry, some government agencies, and researchers	Widely used by the industry, some government agencies, and researchers
Special category of drugs	More comprehensive inclusion of vaccines and OTC drugs	Very low frequency of vaccines and OTC drugs	Extremely low frequency of vaccines and OTC drugs	No vaccines and OTC drugs; only inpatient drug-use data
Time lag	About 2 years	Almost no time lag	About 6 months	About 6 months
Reliability in estimates	Only about 5-8% of drug names (depending on the datasets) are within the acceptable reliability. Only 6 medications per	3528 names (55.1%) had a frequency equal to or more than 5000	Data not available	Data not available

	visit are captured.			
Completeness of information	No information about strength, form, or route of drug administration	Complete information	Complete information	Complete information
Number of names	About 1200 names each year	About 6000 names	Unknown due to the aggregated nature of the data, but will be less than 5600 names	Unknown due to the aggregated nature of the data, but will be less than 2500 names
Unique characteristics	Frequency is based on drug mentions, which could be generic or brand names and may not be the drugs dispensed at the pharmacies	Dispensing data, closely linked to sales data	Claims data, closely linked to dispensing and sales data	Utilization data, closely linked to dispensing. Uses mostly generic names; unable to isolate frequency for any specific brand
Recommendation to use	• When restricted funds is a major consideration • When the focus is on the "prescribing" behavior of office-based physicians • When timeliness is not a major concern • When the interest is to study co-prescribing of several medications	• When funding is not a major constraint • When detailed product information is vital to the research • When timeliness is a concern • When sales data are required • When more drug names and/or more accurate estimates are required	• When funding is not a major constraint • When detailed product information is vital to the research • When employment-based insurees are the focus of research	• When funding is not a major constraint • When inpatient drug use is required for the research

[a] Three-year average of frequency for people covered by employment-based insurance (about 60% of total US population)
[b] Three-year average of frequency for about 50% of total discharges in the US