DOCUMENT RESUME

ED 367 338 IR 054 869

AUTHOR Lavengood, Kathryn A.

TITLE Retrieving Online Information on Drugs: An Analysis of Four Databases.

PUB DATE Jul 93

NOTE 60p.; M.L.S. Research Paper, Kent State University.

PUB TYPE Dissertations/Theses - Masters Theses (042)

EDRS PRICE MF01/PC03 Plus Postage.

DESCRIPTORS Access to Information; Comparative Analysis; Databases; Health Personnel; *Indexing; Medical Libraries; *Online Searching; Online Systems; Online Vendors; *Pharmacology; Relevance (Information Retrieval); Scientific and Technical Information; *Search Strategies; Subject Index Terms; User Needs (Information)

IDENTIFIERS BIOSIS Previews; DIALOG; *Drugs

ABSTRACT

This study examines the indexing of drugs in the literature and compares actual drug indexing to stated indexing policies in selected databases. The goal is to aid health science information specialists, end-users, and/or non-subject experts to improve recall and comprehensiveness when searching for drug information by identifying the most useful search terms (or search term combinations) when seeking information about a drug. This study uses a comparative methodology to evaluate the indexing of 30 drugs within four online databases (Analytical Abstracts; BIOSIS PREVIEWS; Pharmaceutical News Index; and SCISEARCH). The results underscore the need for a basic understanding of pharmaceutical nomenclature; effective use of chemical dictionary files; awareness of indexing differences among databases; and a well-planned search strategy that includes flexibility to make changes as necessary in order to complete a successful search. A list of the drugs used is appended. (Contains 20 references.) (Author/JLB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it

☐ Minor changes have been made to improve reproduction quality

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

IR

ED 367 338

RETRIEVING ONLINE INFORMATION ON DRUGS: AN ANALYSIS OF FOUR DATABASES

A Master's Research Paper submitted to the
Kent State University School of Library and Information Science
in partial fulfillment of the requirements
for the degree Master of Library Science

by

Kathryn A. Lavengood

July, 1993

12054869

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Kathryn Lavengood

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ABSTRACT

Information specialists who work in any of the related fields of health science will often seek drug information. One of the biggest problems in retrieving drug information is the number of ways a drug can be described (variations of drug nomenclature). This study examined the indexing of thirty selected drugs in four online databases (Analytical Abstracts, BIOSIS PREVIEWS, Pharmaceutical News Index, and, SCISEARCH). The thirty drugs were first searched against two dictionary files (CHEMNAME and THE MERCK INDEX ONLINE) to identify all associated names and synonyms. Each term thus identified was then searched in each of the four databases. The search results are analyzed by indexed terms and compared between each database.

The study is intended to aid searchers improve recall and comprehensiveness when searching for drug information by identifying the most useful search terms or search term combinations. The study seeks to answer the questions, "Is recall improved by searching: all available nonproprietary names?; all available proprietary names?; and/or a combination of nonproprietary and proprietary names?"

The study results underscore the need for a basic understanding of pharmaceutical nomenclature, effective use of chemical dictionary files, awareness of indexing differences among databases, and a well-planned search strategy but flexibility to make changes as necessary.

Master's Research Paper by

Kathryn A. Lavengood

B.S., The Ohio State University, 1976

M.L.S., Kent State University, 1993

Approved by

Adviser _____ Date 7/15/93

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

1. The Problem

1.1 Introduction

Information specialists who work in any of the related fields of health science will often seek drug information. Depending on the nature of the search question a variety of databases may be consulted ranging from general science to chemistry, microbiology, physiology, various medical specialties, and business and legal regulatory sources. To ensure comprehensiveness, it is usually necessary to carefully search more than one database and the searcher must be fully cognizant of differences in indexing practices among them.

One of the biggest problems in seeking drug information is drug nomenclature because no single universal name exists for a chemical (Snow 1989). Each source will specify its preferred nomenclature in indexing chemical substances. But when the preferred term is not available, the searcher must determine what alternate terminology is used. To find information about a specific drug, it is usually necessary to search a range of possible names. The non-subject expert or inexperienced searcher can encounter numerous pitfalls such as inconsistent indexing, nomenclature variations, and varying indexing policies between databases.

Complicating the problem of variant terminology is that

clinical literature is indexed differently from chemical literature which is indexed differently than the pharmaceutical business literature. A search using only one or two common names may result in low recall and much missed information. User satisfaction is not a measure of adequate retrieval since many users will be unaware of what they are missing.

A careful analysis of the indexing practices of databases without controlled indexing (e.g., thesauri and other search aids) is necessary to aid in developing more effective search strategies. Particularly with the rising costs of online searching, a search involving too many terms (or too many unlikely terms) will add excessively to the cost of a search without improving recall.

1.2 Statement of the Problem

This study examines the indexing of drugs in the literature and compares actual drug indexing to stated indexing policies in selected databases. The goal is to aid health science information specialists, end-users, and/or non-subject experts to improve recall and comprehensiveness when searching for drug information by identifying the most useful search terms (or search term combinations) when seeking information about a drug.

The study seeks to answer the questions, "Is recall improved by searching:

- all available nonproprietary names?

- all available proprietary names?
- a combination of nonproprietary and proprietary names?

1.3 Limitations

This is a small-scale study which focuses on four representative databases (and those only on the Dialog Information Services System) and in no way can these results be considered definitive for all databases. Thirty drugs (of thousands available) have been tested in this study. It was hoped these thirty would be representative of drugs in general though may in fact present a biased set. Finally, results for comparison questions (one term vs. all terms) were calculated on an additive basis without removing duplicates and the single term used for comparison was selected arbitrarily (subjective assessment of "most common" term in U.S.).

1.4 Definitions

For this study, the term 'drug' has been defined as "any chemical compound that may be used as an aid in the diagnosis, treatment, or prevention of disease or for any other therapeutic purpose." The term 'drug nomenclature' refers to the system of terms used in the science of pharmacology.

Goodman and Gilman (1980) define pharmacology as encompassing "the knowledge of the history, source, physical and chemical properties, compounding, biochemical and physiological effects, mechanisms of action, absorption, distribution, biotransformation and excretion, and therapeutic and other uses

of drugs." This description makes it clear that the subject of pharmacology (the science of drugs) is quite extensive.

2. LITERATURE REVIEW

Many drugs, particularly drugs that have been available for several years, have numerous synonyms, trade names, and other means of identification. McGowan and Mater (1985/86) describe a method for identifying chemicals and drugs using a variety of tools such as the Physicians' Desk Reference, The Merck Index, and the Chemical Abstracts Index Guide in addition to online chemical dictionaries. They emphasize the usefulness of the Chemical Abstracts Service (CAS) registry number as a simple search strategy in databases that include registry numbers. They also point out differences in assigned chemical names and give an example of the same drug identified in Chemical Abstracts (parent group name followed by substituents in order of importance) and in The Merck Index (chemicals listed in non-inverted order; substituent groups followed by parent group name). They continue on to discuss effective use of online chemical dictionary files and search strategies when little is known about a substance.

Bronson (1992) elaborates further on using two chemical dictionary files, CHEMLINE and CHEMID, to find information about drugs. By providing more search terms, these files are useful to develop more comprehensive strategies. CHEMLINE and CHEMID provide current and superceded CAS registry numbers and up to

200 synonyms per chemical. These files also allow name fragment searches and molecular formula searches.

Deaves and Pache (1989) in their article, "Chemical and Numerical Indexing for the INSPEC Database," discuss difficulties of searching for chemical data online. Although they focus on inorganic substances, the difficulties apply equally to organic substances. For example, there can be different but very common formulae for the same substance. Hyphens can have different meanings in different contexts and formulae with subscripts can be difficult to search online.

Roth (1985) discusses pitfalls of chemical literature searching such as inconsistent indexing, nomenclature variations, language, and transliteration. These present problems for the non-subject expert or inexperienced searcher in particular. He presents an excellent literature review of search problems such as:

- indexing services covering the same subject (supposedly comprehensively) but varying in recall
- the expense of searching online nomenclature files
- the evolving nature of nomenclature

Comprehensive searching "depends on carefully searching a wide range of publications and/or databases." He concluded with examples of questions that are most dangerous for inexperienced searchers or those without subject expertise and most likely to consume hundreds of dollars yet yield unsatisfactory results.

John Barber, et al., (1988) in "Case Studies of the Indexing and Retrieval of Pharmacology Papers," presents a detailed analysis of the coverage and indexing of thirty papers on pharmacological topics and concluded there was considerable variation in the indexing applied with drug formulations. While the study did not focus specifically on drug nomenclature, examples of variant indexing of drugs were included.

Many papers describe and discuss problems of drug and/or chemical searching in the literature but do not systematically study the problem. Dwight Tousignaut (1982) in "Searching 'Pharmacy' Databases: Nomenclature Problems and Inconsistencies." points out such problems as: inconsistencies even in "standardized" nomenclature schemes; drug names that vary from country to country; errors in source articles that lead to errors in the secondary literature; and, confusion added when manufacturers use the same trade name in more than one country but with different formulations of the two. Such inconsistencies are not likely to be identified by the non-subject expert searcher.

In "Indexing: Old Methods, New Concepts," Tousignaut (1987) compares traditional indexing to a concept indexing scheme created as a result of developing Drug Information Fulltext. His conclusion was that "the future of fulltext will depend on controlled indexing approaches that offer easy access and dependable results."

Bonnie Snow (1982) in "Trade Names in Medicine: Searching

for Brand Name Comparisons and New Product News," discusses the need to consult search aid databases for references to alternate names for trade names.

Snow has written extensively about the general topic of searching for drug information. In her book, Drug Information: A Guide to Current Resources (1989), she devotes one chapter to "Drug Nomenclature" which is an excellent description and definition of the wide range of drug names that exist. In another chapter she describes "Identification and Nomenclature Sources" as aids in finding alternate names for particular drugs.

In the chapter on "Abstracting and Indexing Services," Snow discusses in detail several selected online bibliographic databases useful for the pharmaceutical searcher and describes the chemical indexing policies of each. As she points out, her guide cannot provide more than an overview of each database and general statements about the indexing policies. A typical example is the description of chemical indexing in DE HAEN DRUG DATA:

> USAN generic names are preferred nomenclature in the DE HAEN database. Trade names are searchable in many DE HAEN records. If the source author refers to a drug by trade names, the online record will include the names given. Chemical names, CAS registry numbers, and molecular formulas are indexed in many but not all...records.

This illustrates the multiple ways a drug may be indexed within a single database and makes clear the difficulty of searching more than one database when there is no standard method of

indexing drugs.

In her chapter on "Online Database Selection," Snow presents a chart showing the "Preference Hierarchy for Pharmaceutical Nomenclature in Selected Online Databases." This reinforces the differences between databases and the need for searching under a range of possible names for a particular drug.

For this study, the term 'drug' is used to describe a pharmacologically active chemical or compound. Each drug entity can be described in a variety of ways including:

- Chemical Name: In the United States this generally follows the American Chemical Society conventions for naming compounds.

- Molecular Formula: Describes a compound by atom count (e.g., $C_{16}H_{21}NO_2$).

- Chemical Abstracts Service (CAS) Registry Number: This is a unique identifying code for a substance.

- Nonproprietary Name or Generic Name: This is a simplified chemical name. The terms 'nonproprietary' and 'generic'are commonly used interchangeably although the terms are not synonymous. (The generic name refers to a class of drugs while a nonproprietary name refers to a specific compound.) In the U.S., this term is assigned by the United States Adopted Names Council and is referred to as a U.S. Adopted Name (USAN). Other agencies may assign different names to the same substance and variants may include the British Approved Name (BAN) or the International Nonproprietary Name (INN) assigned by the World Health Organization.

- Drug Investigational Code or Research Number: A code or abbreviation assigned to new products under investigation for ease of reference and for security of a new drug discovery. More than one research code may be used for the same drug.

- Trade Name or Proprietary Name: This is usually as registered trademark of the manufacturer to identify a specific product formulation and indication. Many products have multiple trade names and even when manufactured by one company may have different trade names in different countries.

3. METHODOLOGY

This study uses a comparative methodology to evaluate the indexing of drugs within four online databases. A comparative study will obtain comparable measures on different databases; that is, one searches the same terms in each of the four databases, tabulates the results, and examines the similarities or differences of the indexing of each database and its effect on retrieval performance.

Thirty drug entities were chosen to study. Fifteen of the drugs chosen have been commercially available in the United States for at least ten years. Fifteen of the drugs have only been available within the last five years, some of them still in the investigational process. Each drug entity was searched against two dictionary files (CHEMNAME and THE MERCK INDEX ONLINE) for all associated names including trade or proprietary names, chemical names, synonyms or acronyms, investigational codes or research numbers, Chemical Abstracts Service Registry Numbers, and, generic or nonproprietary names.

Next, each of the terms thus identified were searched in four online databases. The four databases selected for the study (Analytical Abstracts; BIOSIS PREVIEWS; Pharmaceutical News Index; and, SCISEARCH) offer broad coverage and subject

scope but none utilize controlled vocabulary indexing. Many alternate databases could have been selected for study. Since the drug information searcher is often required to find chemical, clinical, business, or general information about a particular drug, a database representing each type of search was selected.

SCISEARCH indexes 4,500 journal titles from more than fifty countries and offers broad coverage of the general science and technology literature. For this study, DIALOG File 34 (covering 1988 to the present) was studied. A journal issue is likely to be cited in SCISEARCH within two weeks of publication making it a good source of current information. Subject access is limited to title words and abstracts are not always included. Therefore, the searcher must use numerous alternate names to improve recall.

BIOSIS PREVIEWS is international in scope and encompasses research in the biological and biomedical sciences. For this study, DIALOG File 55 (covering 1985 to the present) was studied. About half of the citations are clinically oriented and the pharmaceutical search uses it to locate information on drug development, toxicity, and pharmacology. The preferred terms in indexing are the U.S. Adopted Names (USAN) but when mentioned in the original source, investigational codes and proprietary names are also included.

Analytical Abstracts is a chemically-oriented database published by the Royal Society of Chemistry and covering more than 1,300 journal titles, books, reports, national standards,

and conference proceedings. Drug information searchers utilize Analytical Abstracts for topics involving biochemistry or medicinal chemistry. The preferred nomenclature is the chemical name and CAS registry number.

The Pharmaceutical News Index database covers current and retrospective business news related to the pharmaceutical industry. It indexes information on legislation and regulations, research and development, and market analysis. There is no controlled vocabulary and trade names, generic names, chemical names, and manufacturer code names may all be indexed.

To compare actual drug indexing to stated indexing policies of the four databases, the search results for each drug were tabulated and graphically presented in terms of percentages of: CAS registry numbers indexed; nonproprietary names indexed; investigational codes indexed; and proprietary names indexed.

Data was also tabulated and graphically presented showing the results of the following searches:

A) One proprietary name vs. all proprietary names identified

B) One nonproprietary name vs. all nonproprietary names identified

C) All nonproprietary names identified vs. all nonproprietary names plus all proprietary names

Comparisons of these results demonstrate whether recall is increased by searching all available nonproprietary names; all available proprietary names; or, a combination of all nonproprietary names plus all proprietary names.

Finally, the search results were tabulated and graphically presented for each database (looking at the thirty drugs in aggregate) showing: percentage of nonproprietary terms indexed; percentage of proprietary terms indexed; percentage of CAS registry numbers indexed; and, percentage of investigational codes indexed.

4. RESULTS

4.1 Introduction

Thirty drug entities were first searched against two dictionary files to identify all associated terms. All of these terms were then searched in four online databases (Analytical Abstracts, BIOSIS PREVIEWS, Pharmaceutical News Index, and SCISEARCH) and results are discussed by database. Results for the thirty drug entities in aggregate (by type of term) are illustrated in Figure 1. The search results presented for each database include:

- Graphs showing percentage of:
  - CAS registry numbers indexed
  - nonproprietary terms indexed
  - investigational codes indexed
  - proprietary terms indexed
- Graphs showing comparison question results (i.e., one term vs. all terms)

4.2 Analytical Abstracts

The search results from Analytical Abstracts for percentage of Chemical Abstracts Service (CAS) registry numbers indexed are shown in Figure 2. Because Analytical Abstracts is chemically-oriented, it is not surprising that the majority of terms (70% of total) were retrievable using the CAS registry numbers. Each of the terms which were not retrievable using the registry number were from the recent drug group, again a reflection of less currency or narrower scope of the database.

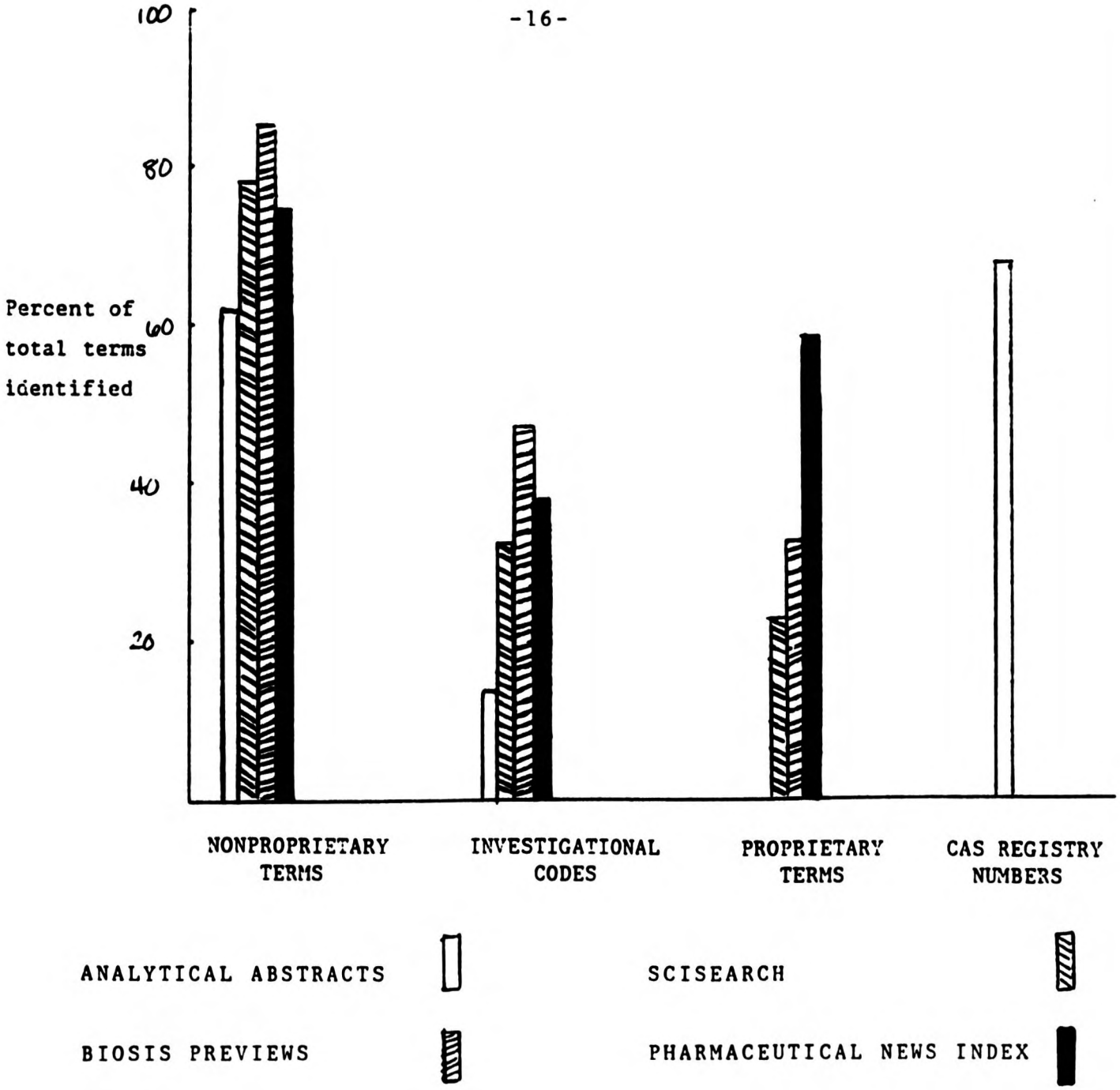

Fig. 1. Percent of Total Terms Indexed for Thirty Drugs in Aggregate

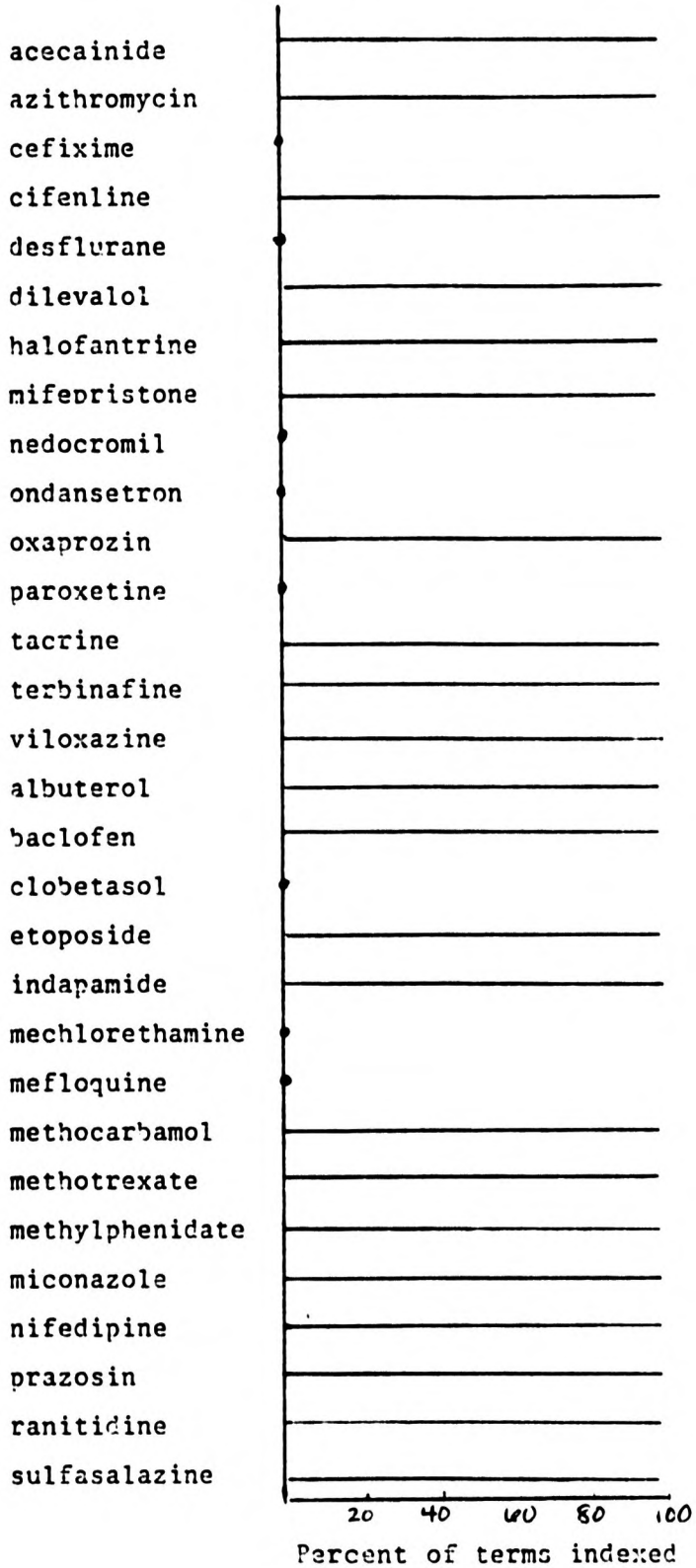

Fig. 2. Analytical Abstracts: Percent of Chemical Abstracts Registry Numbers Indexed

Unfortunately, this was the only database of the four studied that utilized registry numbers.

No proprietary terms retrieved any citations at all, consistent with the orientation specifically toward analytical chemistry. Investigational codes were not commonly searchable - only 13% of the total codes identified retrieved any citations (see Figure 3). Again this reflects the narrow subject scope and orientation to the needs of the analytical chemist.

Most of the drug entities (97%) were retrievable using the current nonproprietary term (USAN) as shown in Figure 4. However, very few of the synonyms retrieved any citations. This probably reflects the emphasis of the database on chemical names and CAS registry numbers. Two nonproprietary terms (cifenline and desflurane) retrieved no citations possibly reflecting the smaller scope of Analytical Abstracts or a less-current coverage as both of these drugs are of more recent discovery. The 'cifenline' entity was retrievable using its previous nonproprietary term of 'cibenzoline.' Also, Analytical Abstracts reflects its British influence (it is produced by the Royal Society of Chemistry) in its use of British spelling conventions. Thus 'sulfasalazine' retrieved no citations while 'sulphasalazine' retrieved several. Likewise, 'albuterol' retrieved eleven citations while the British term 'salbutamol' retrieved seventy-one citations.

Of the comparison questions, only question B can be analy-

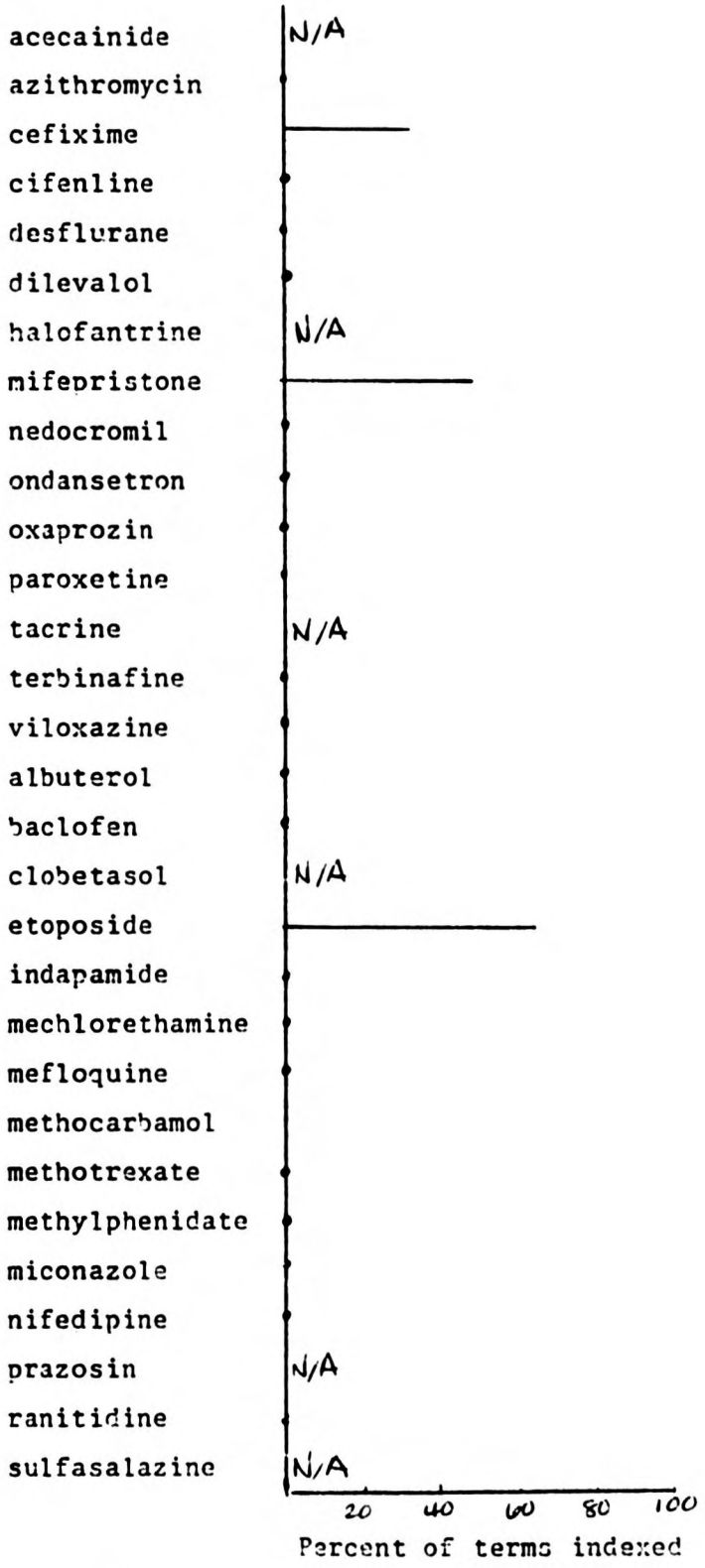

Fig. 3. Analytical Abstracts: Percent of Investigational Codes Indexed

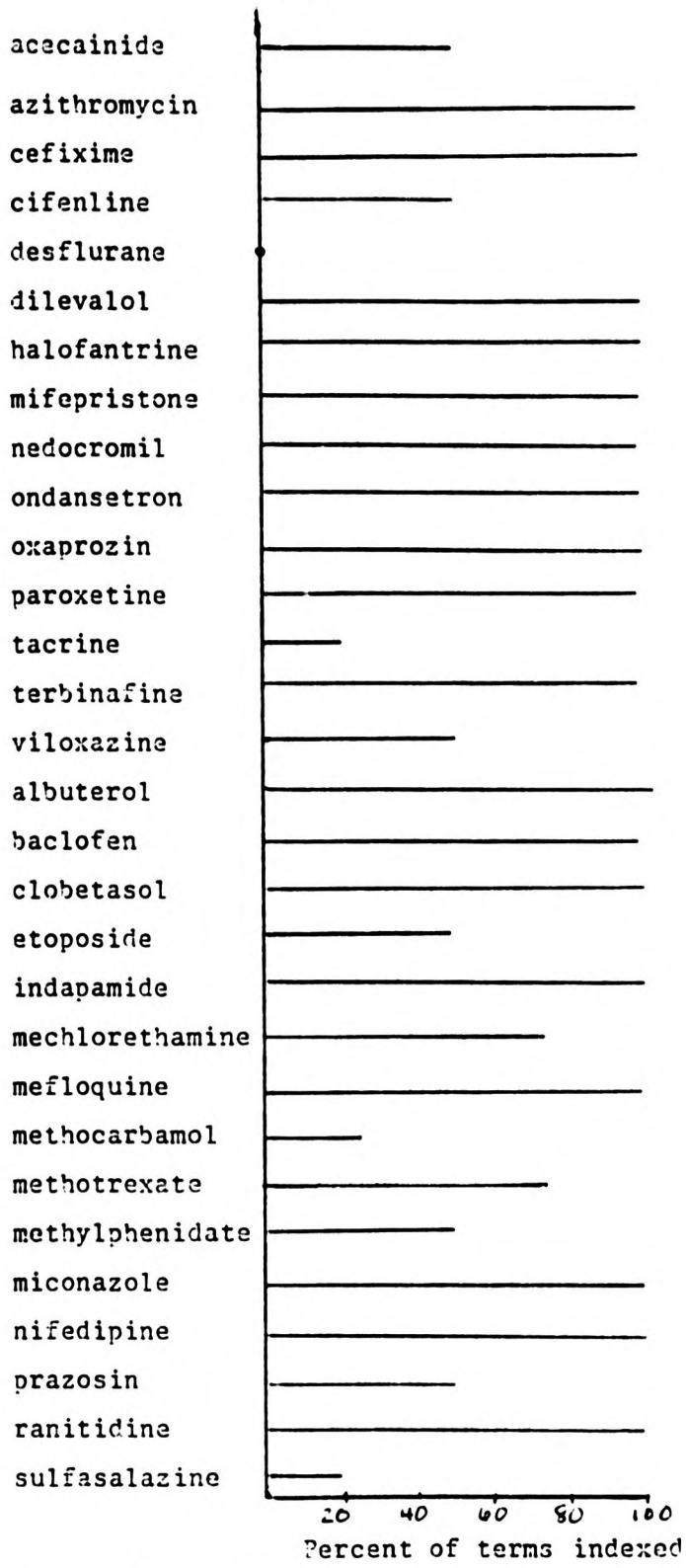

Fig. 4. Analytical Abstracts: Percent of Nonproprietary Terms Indexed

zed (see Figure 5) because no proprietary terms were searchable in Analytical Abstracts. As discussed previously, in general the current nonproprietary term (USAN) retrieved the majority of citations and addition of secondary terms did not improve recall. The exceptions involved spelling variants and use of older terminology.

4.3 SCISEARCH

The SCISEARCH database offers broad, multidisciplinary subject coverage of the literature of science and technology. In addition, a journal issue is likely to be cited in SCISEARCH within two weeks of publication and sources are indexed cover-to-cover. However, author abstracts are searchable only in records added since January 1991 when available. All terms in the database are derived from the author's language and this is a noticeable factor in which terms retrieve citations.

Therefore it is not surprising to find multiple hits when searching on nonproprietary names and alternate nonproprietary terms (see Figure 6). As in Analytical Abstracts, there is a significant increase in recall when using older or previous nonproprietary term names compared to the newer or current term (e.g., cifenline - 5 hits, cibenzoline - 65 hits; albuterol - 273 hits, salbutamol - 952 hits).

Curiously, there was very low recall (16 hits) on the term 'methocarbamol' and zero postings on any of its alternate terms,

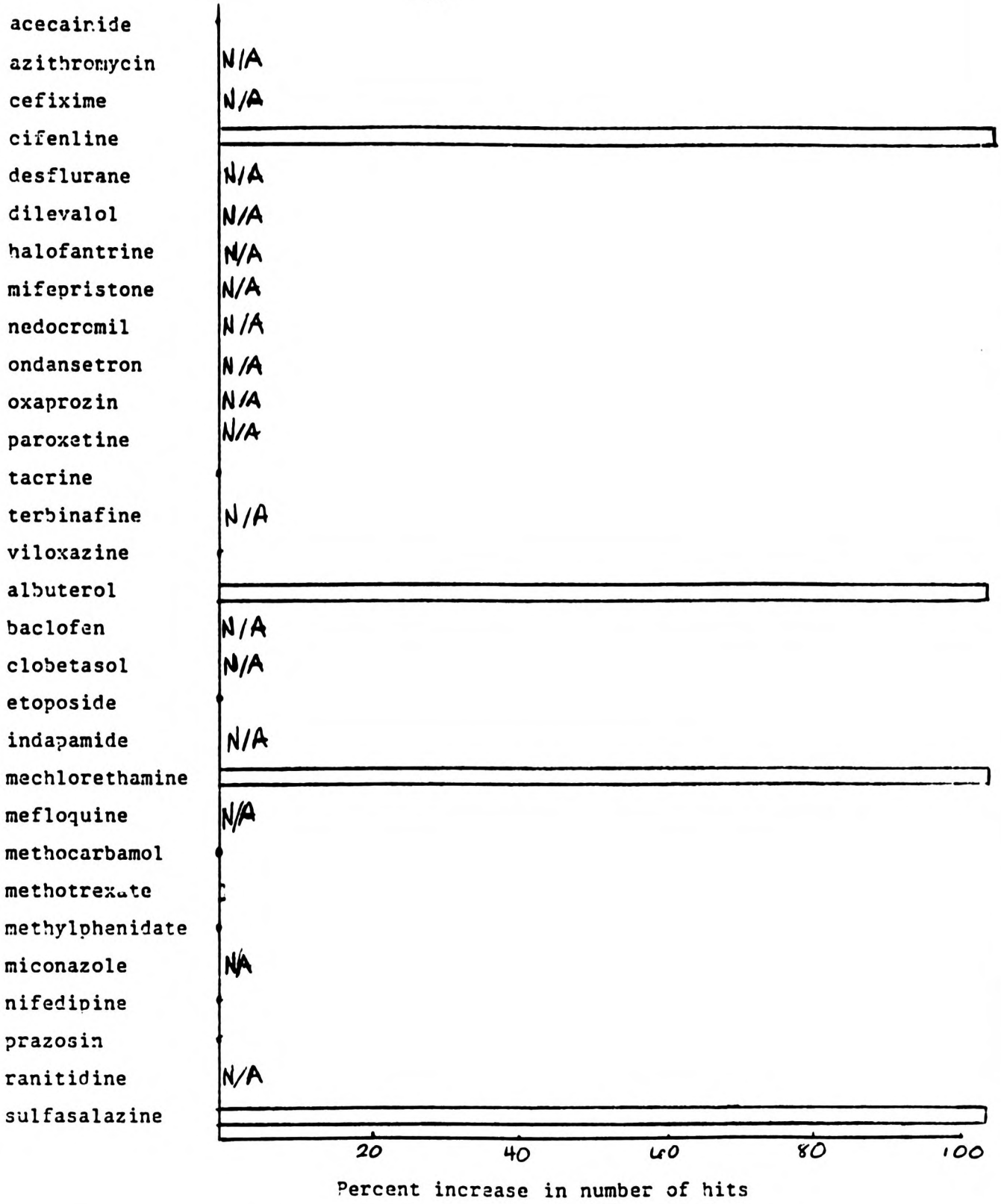

Fig. 5. Analytical Abstracts: One Nonproprietary Term vs. All Nonproprietary Terms

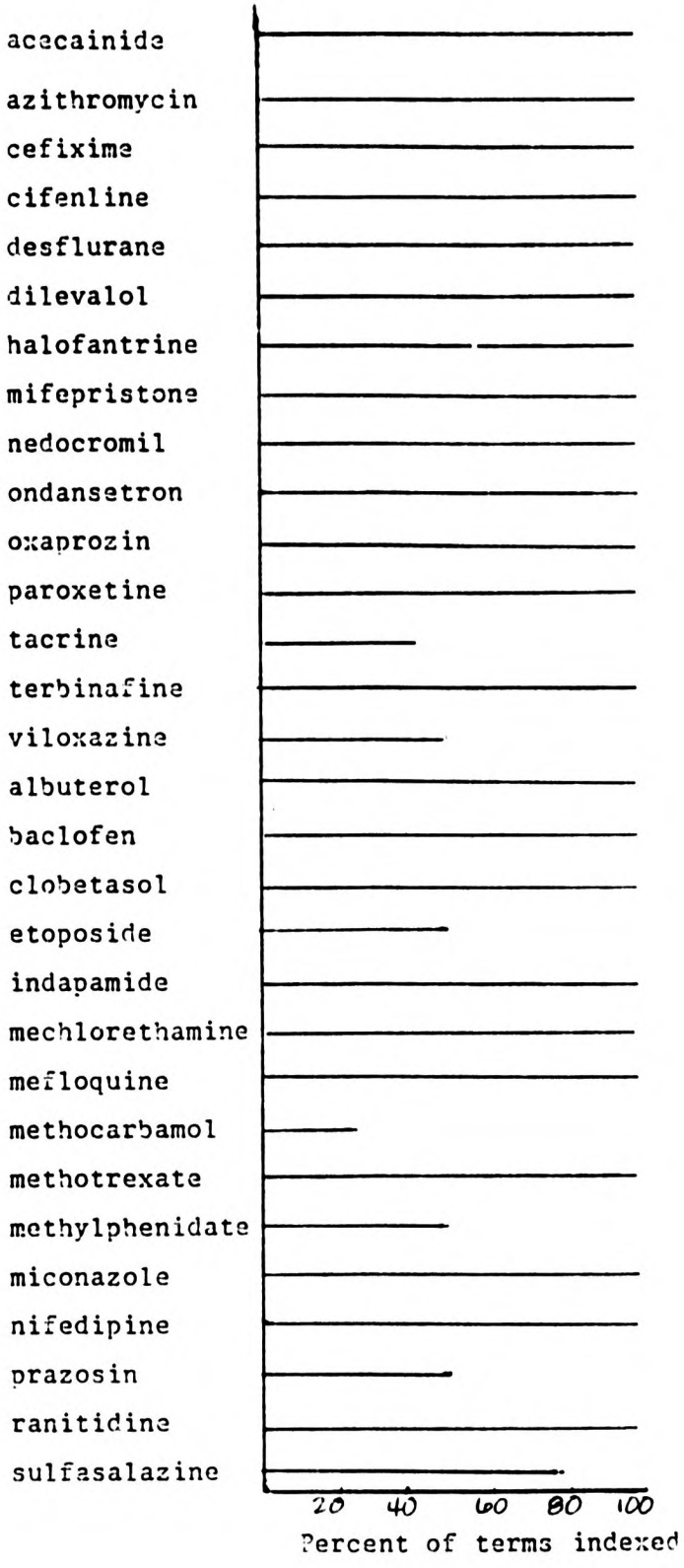

Fig. 6. SCISEARCH: Percent of Nonproprietary Terms Indexed

including both nonproprietary and proprietary terms. Possibly, this reflects the fact that this drug has long been available and has not been studied in recent years. In contrast, the older drug 'sulfasalazine' had numerous postings both on that term specifically (634) and on its alternate nonproprietary terms (867). This may indicate ongoing research on this entity.

Only about a third of the investigational codes were indexed (see Figure 7). This is not surprising since authors would more likely refer to a drug being studied by the nonproprietary term. An exception here are entities which are better or equally known by their investigational codes (e.g., mifepristone - 377 hits vs. RU 486 - 692 hits).

Most of the entities generated hits on at least one of the proprietary terms identified (see Figure 8). Therefore entities with only one proprietary term are likely to show 100% retrieval. But where multiple proprietary terms are available, very few additional postings were gained. Of a total of 184 proprietary terms identified, only 24% were retrieved in SCISEARCH. When comparing the total number of postings for each term (nonproprietary vs. proprietary) nonproprietary terms yielded the highest recall by far with very few postings for any of the proprietary terms.

Figure 9 illustrates the search results of comparison question A which compares one proprietary term to all proprietary terms. The percent increase in number of hits is misleading because of the low number of hits on any one term.

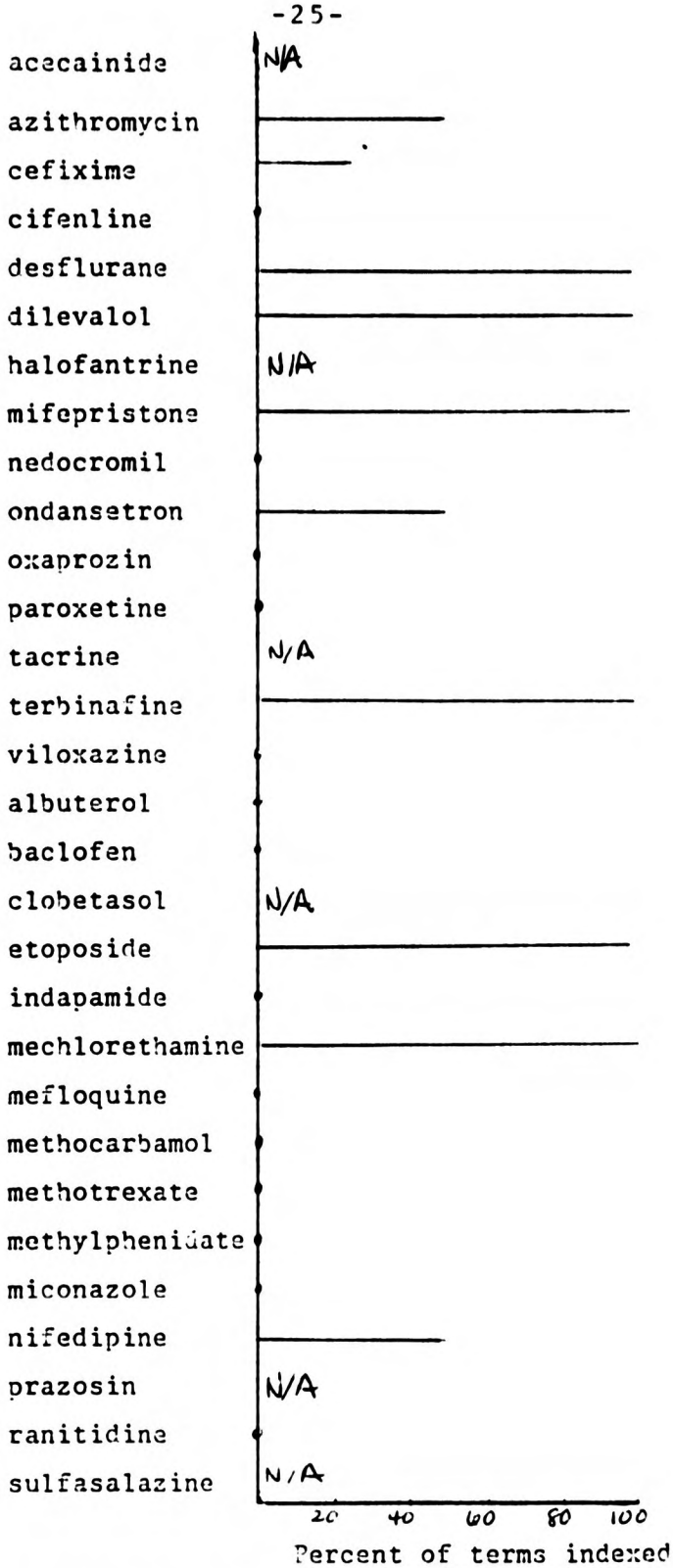

Fig. 7. SCISEARCH: Percent of Investigational Codes Indexed

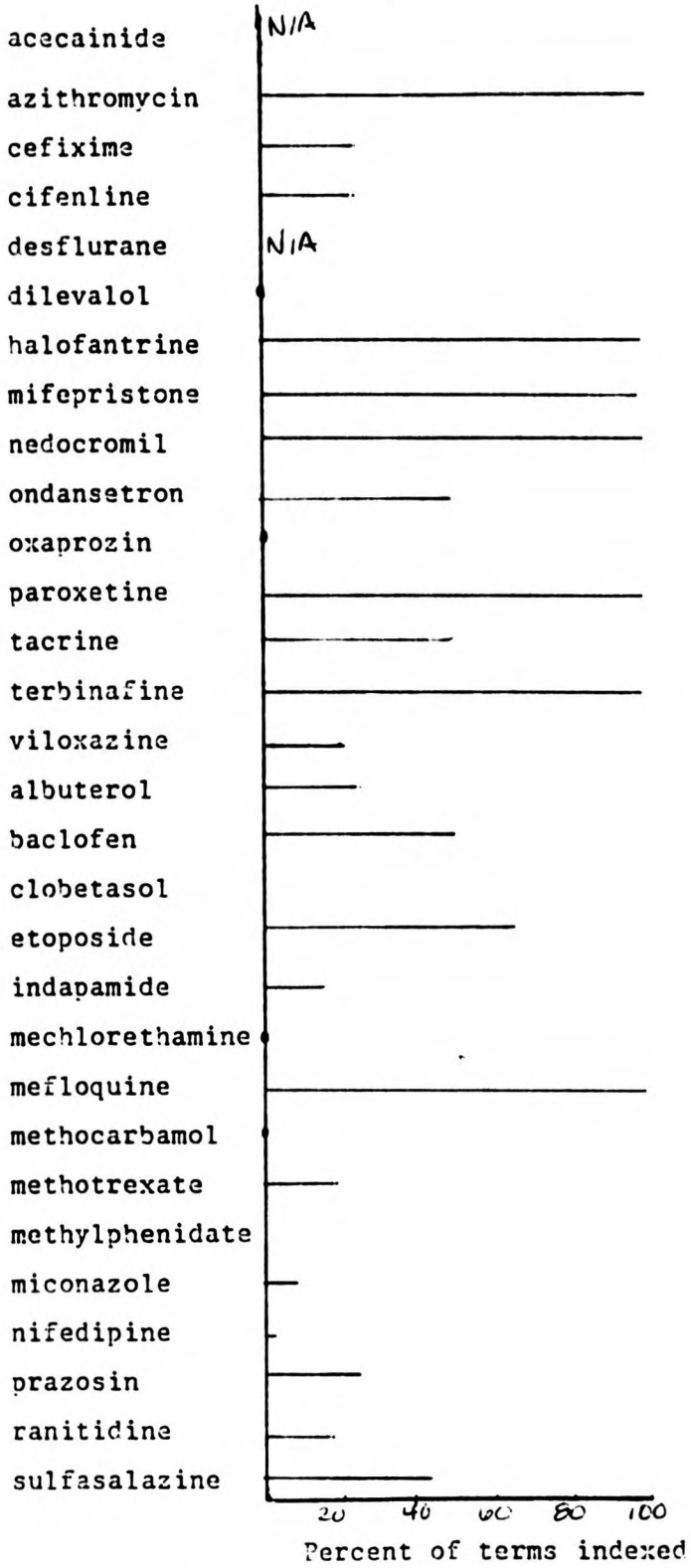

Fig. 8. SCISEARCH: Percent of Proprietary Terms Indexed

Fig. 9. SCISEARCH: One Proprietary Term vs. All Proprietary Terms

Thus, picking up only a few additional hits may appear as a 100% or greater increase in hits.

In question B which compares one nonproprietary term to all nonproprietary terms, a significant increase in recall is observed (see Figure 10). This is particularly true for nonproprietary terms with older synonyms where a search on all identified nonproprietary terms is necessary for more complete recall (e.g., mechlorethamine - 277 hits vs. mechlorethamine + synonyms - 544 hits). However, this is not always the case because even some older terms show little or no increase in recall when combined with other synonyms (e.g., methylphenidate - 308 hits vs. methylphenidate + synonyms - 308 hits).

As discussed earlier, the result of searching on all nonproprietary terms identified yields the greatest number of postings. In general, the addition of proprietary terms does not significantly increase recall (see Figure 11).

4.4. BIOSIS PREVIEWS

The results from BIOSIS PREVIEWS are very similar to the search results in SCISEARCH including highest number of postings for the nonproprietary term; higher postings for older synonyms; and, relatively few postings for proprietary terms. An interesting exception occurs with two terms where acronyms yield significantly greater hits than the nonproprietary term (tacrine - 183 hits vs. THA - 1216 hits; acecainide - 12 hits vs. NAPA - 122 hits). This is consistent with the coverage of BIO-

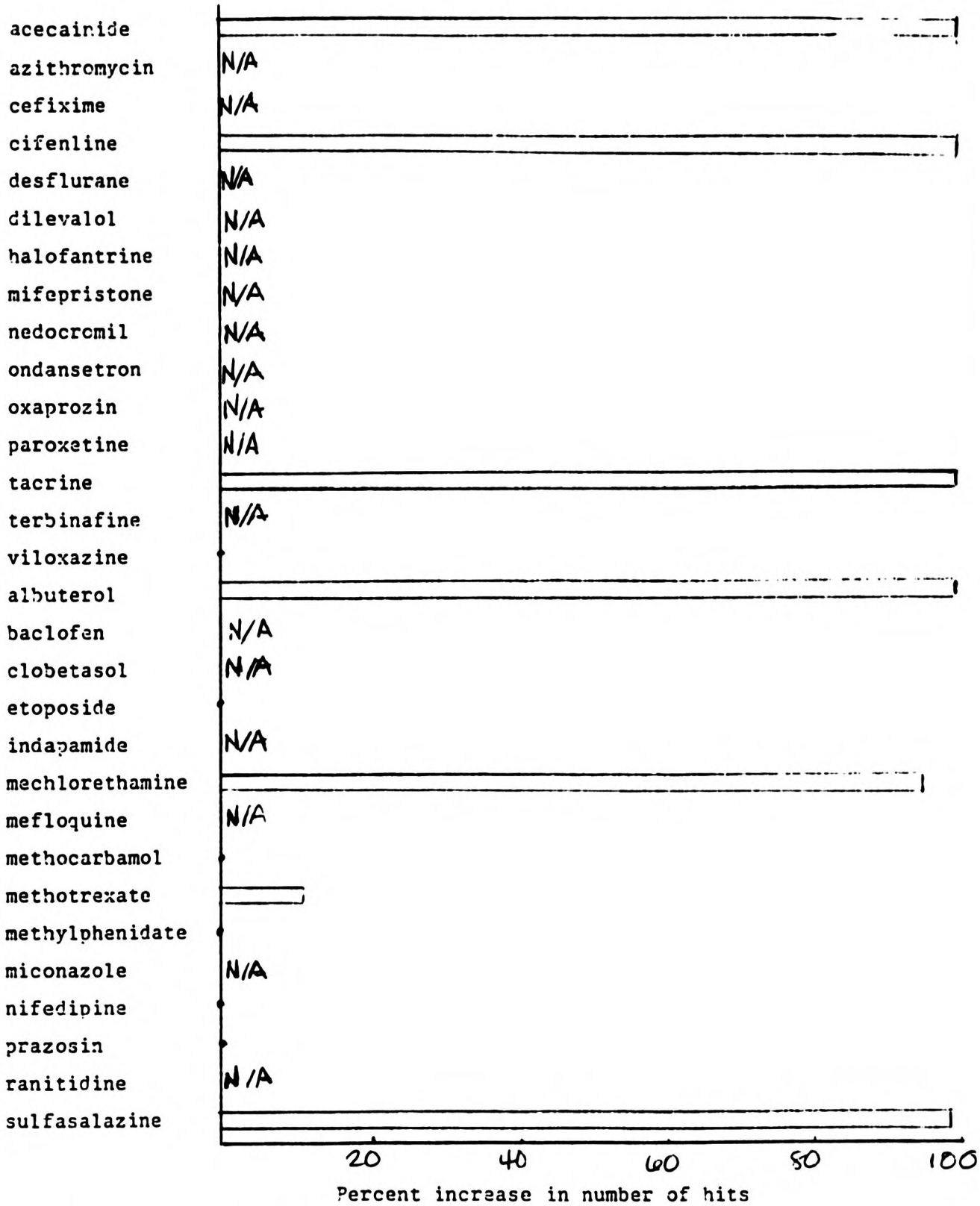

Fig. 10. SCISEARCH: One Nonproprietary Term vs. All Nonproprietary Terms

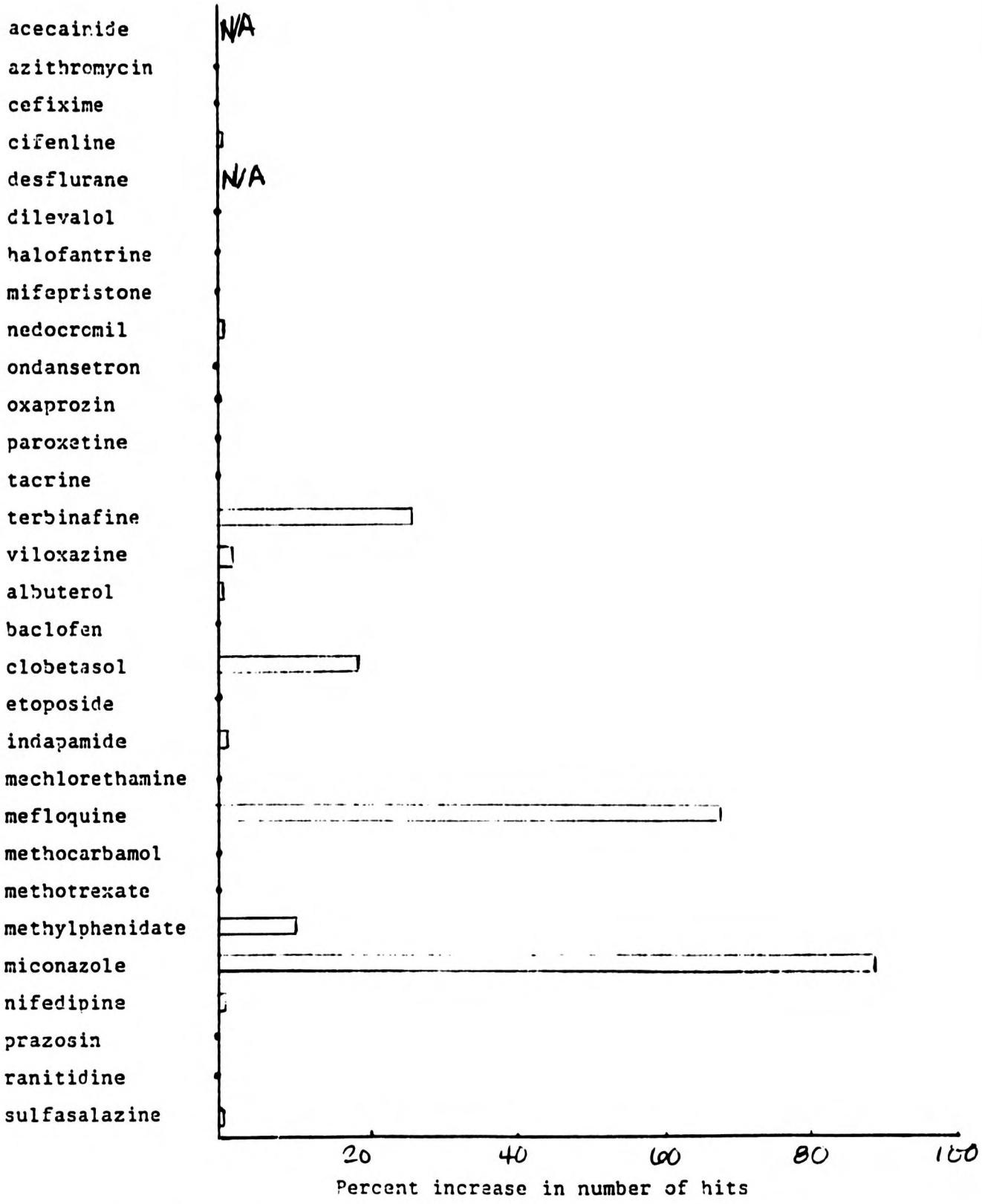

Fig. 11. SCISEARCH: All Nonproprietary Terms vs.
All Nonproprietary Terms + All Proprietary Terms

SIS PREVIEWS of original research and availability of abstracts in all records added since July 1976. Therefore, the more commonly known acronyms are likely to be included as terms.

Also noticeable was the large number of postings per term (particularly for the nonproprietary term) compared to the other databases. This may reflect broader scope and/or coverage or the value of abstracts being available to search because the total number of records in BIOSIS PREVIEWS is similar to the number in SCISEARCH.

BIOSIS PREVIEWS includes more of the investigational codes than the other databases (47% yielded postings) (see Figure 12). Again, because it covers original research and includes abstracts it is more likely that investigational codes will be indexed.

As in SCISEARCH, proprietary terms generally do not result in many hits (see Figure 13) especially compared to nonproprietary terms (see Figure 14). However BIOSIS PREVIEWS indexed 32% of the total number identified which is slightly greater than the total indexed by SCISEARCH (24%).

The comparison question results once again are not significantly different than SCISEARCH results. Small numbers of postings with an additional term may appear to greatly increase recall (e.g., 1 hits vs. 2 hits shows a 100% increase). In general, a search on multiple proprietary terms will add citations but the increase in total number of hits will not be great (see Figure 15).

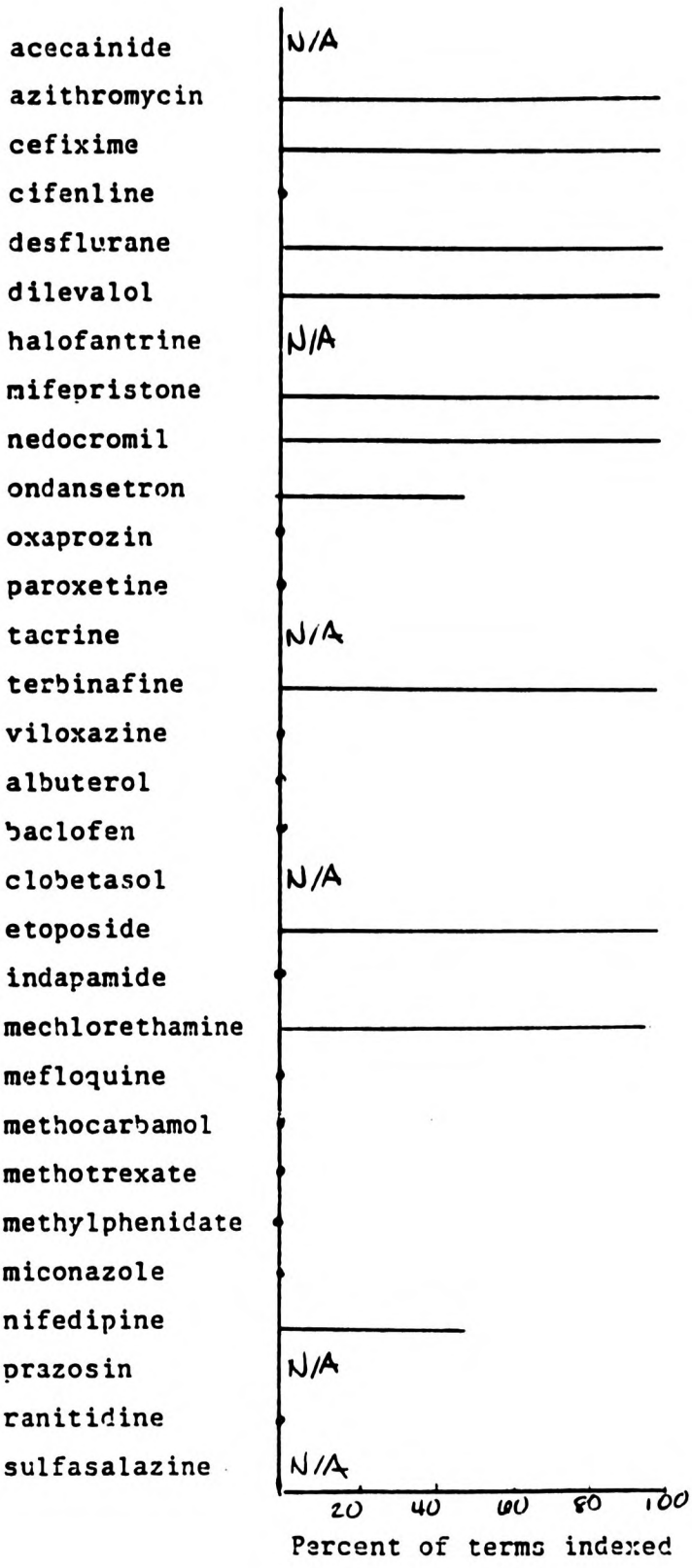

Fig. 12. BIOSIS PREVIEWS: Percent of Investigational Codes Indexed

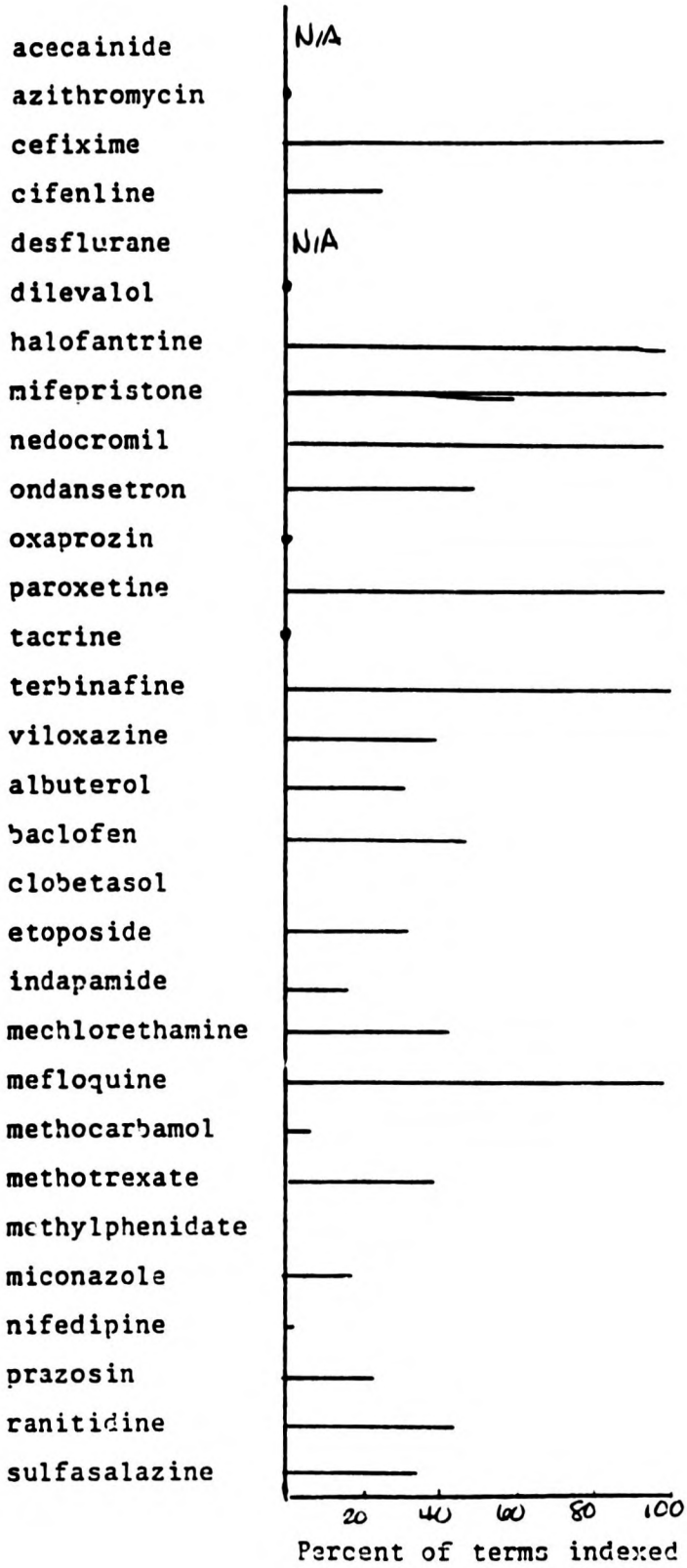

Fig. 13. BIOSIS PREVIEWS: Percent of Proprietary Terms Indexed

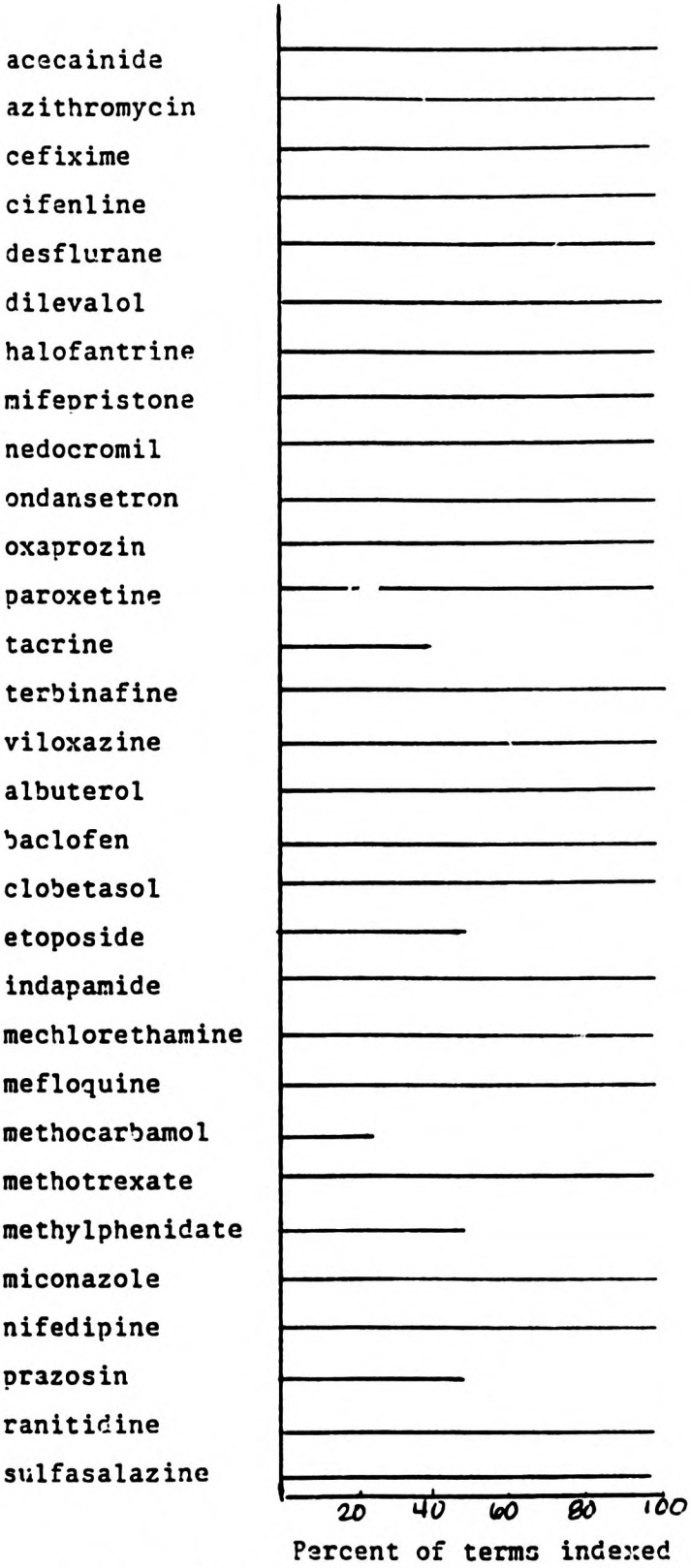

Fig. 14. BIOSIS PREVIEWS: Percent of Nonproprietary Terms Indexed

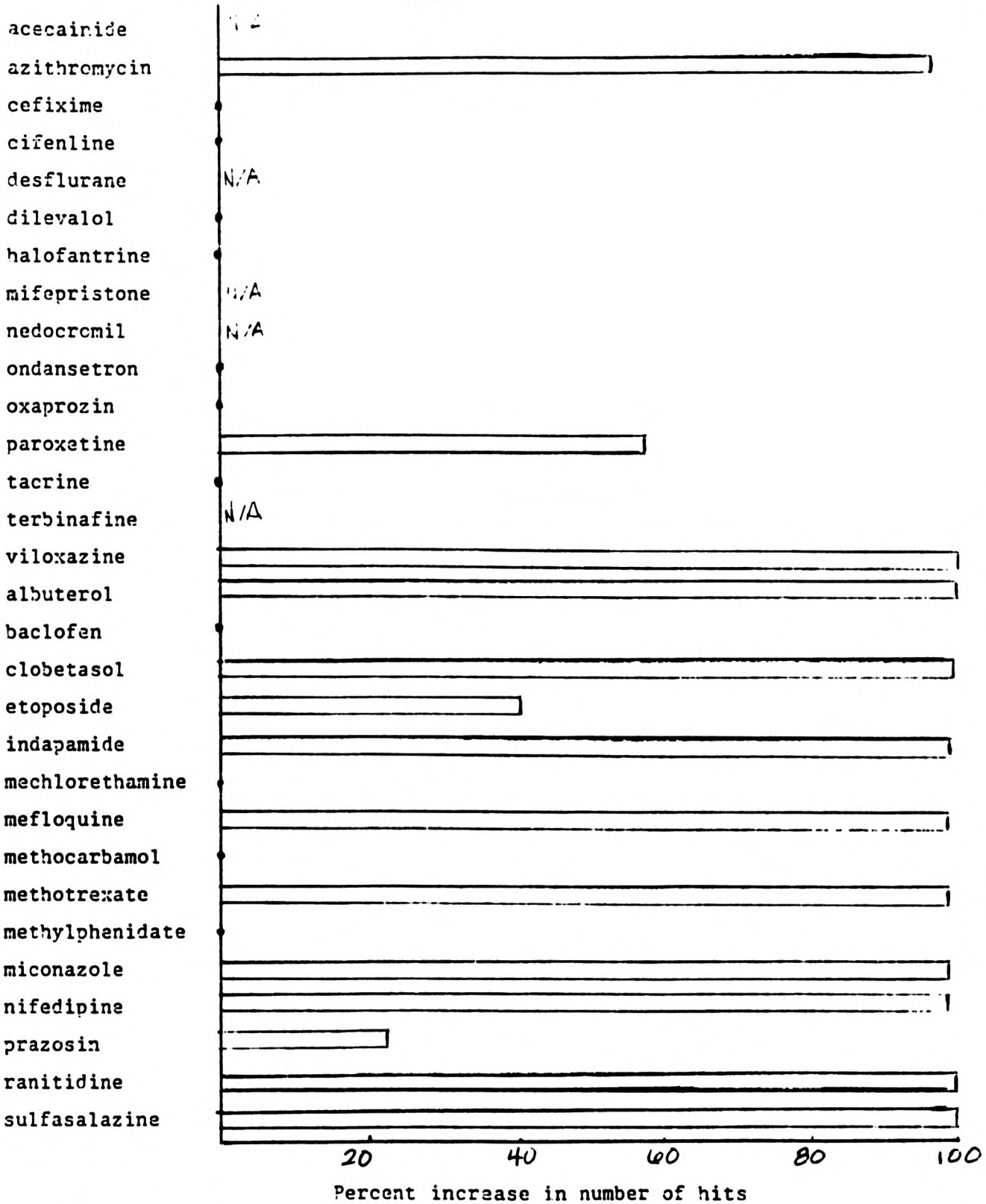

Fig. 15. BIOSIS PREVIEWS: One Proprietary Term vs. All Proprietary Terms

Questions B and C results are similar to SCISEARCH results. That is, searching on multiple nonproprietary terms can significantly increase recall although not in every case (see Figure 16). And, adding all proprietary terms to all nonproprietary terms does not significantly improve recall (see Figure 17).

4.5 Pharmaceutical News Index

Pharmaceutical News Index (PNI) focuses on international pharmaceutical business information. PNI often adds synonyms, proprietary names, acronyms, and abbreviations as descriptors. For this reason, the searcher may benefit greatly by identifying and searching as many terms as are available for greatest recall.

Each of the primary nonproprietary terms was indexed (see Figure 18) which is consistent with the coverage of this database. Once again, there were a greater number of postings for the older terms (e.g., 'cifenline' vs. 'cibenzoline') and for acronyms (e.g., tacrine vs. THA and acecainide vs. NAPA) although not as great an increase as in SCISEARCH or BIOSIS PREVIEWS.

About 37% of the total identified codes resulted in postings (see Figure 19). This is consistent with the database coverage of such areas as R & D in progress, pharmaceutical research, and New Drug Application (NDA) approvals.

PNI significantly differed from the other databases studied in the coverage of proprietary terms (see Figure 20).

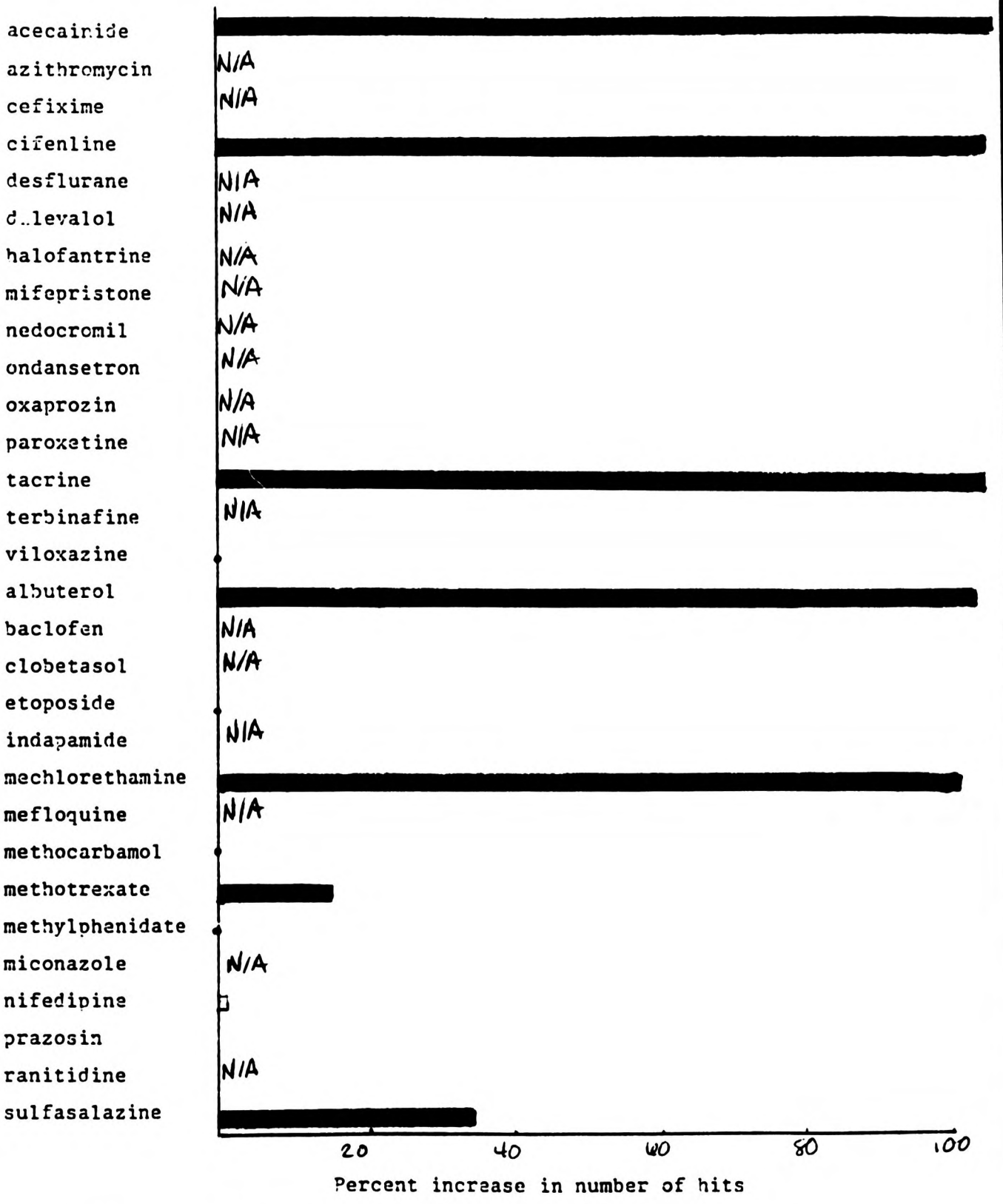

Fig. 16. BIOSIS PREVIEWS: One Nonproprietary Term vs. All Nonproprietary Terms

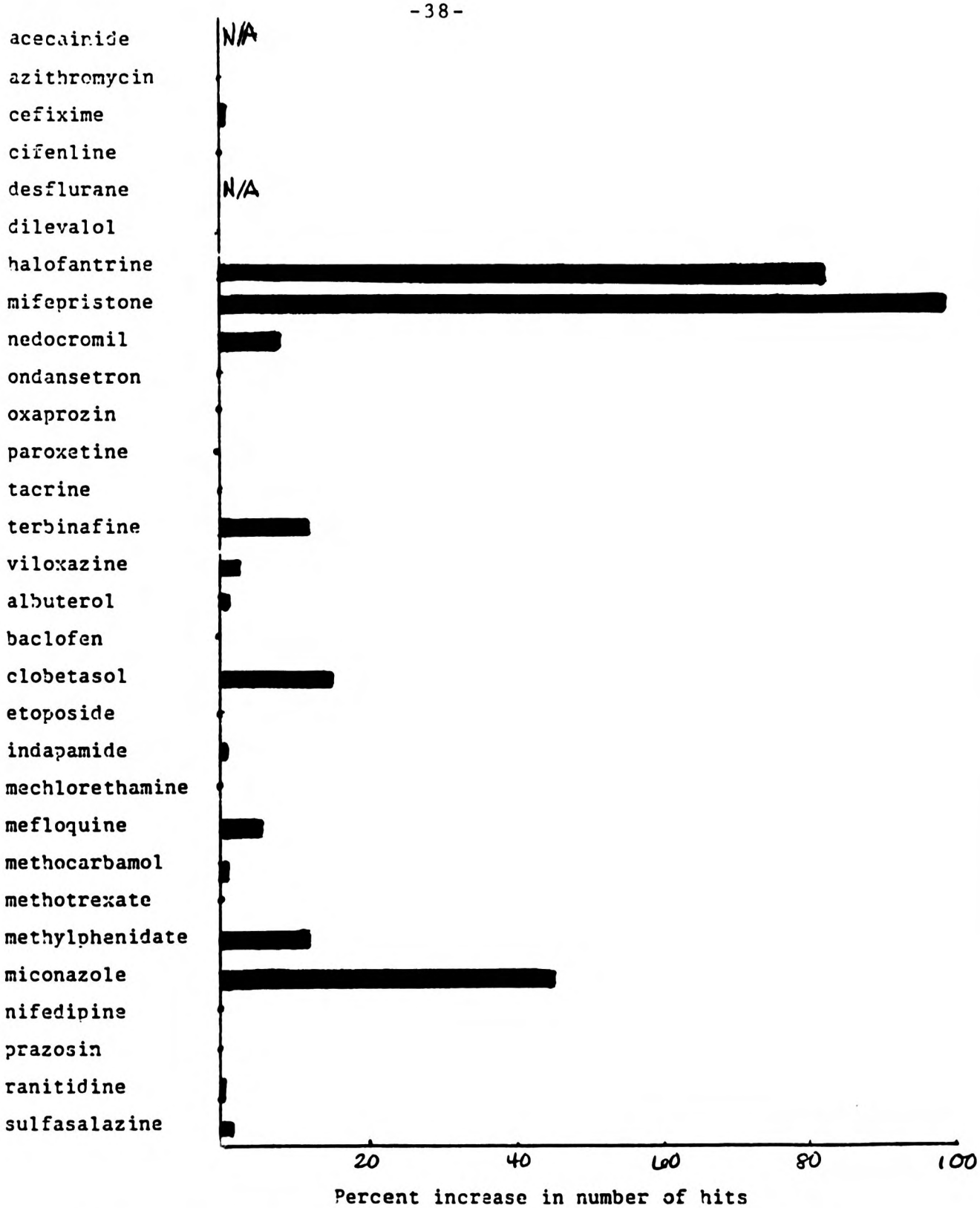

Fig. 17. BIOSIS PREVIEWS: All Nonproprietary Terms vs. All Nonproprietary Terms + All Proprietary Terms

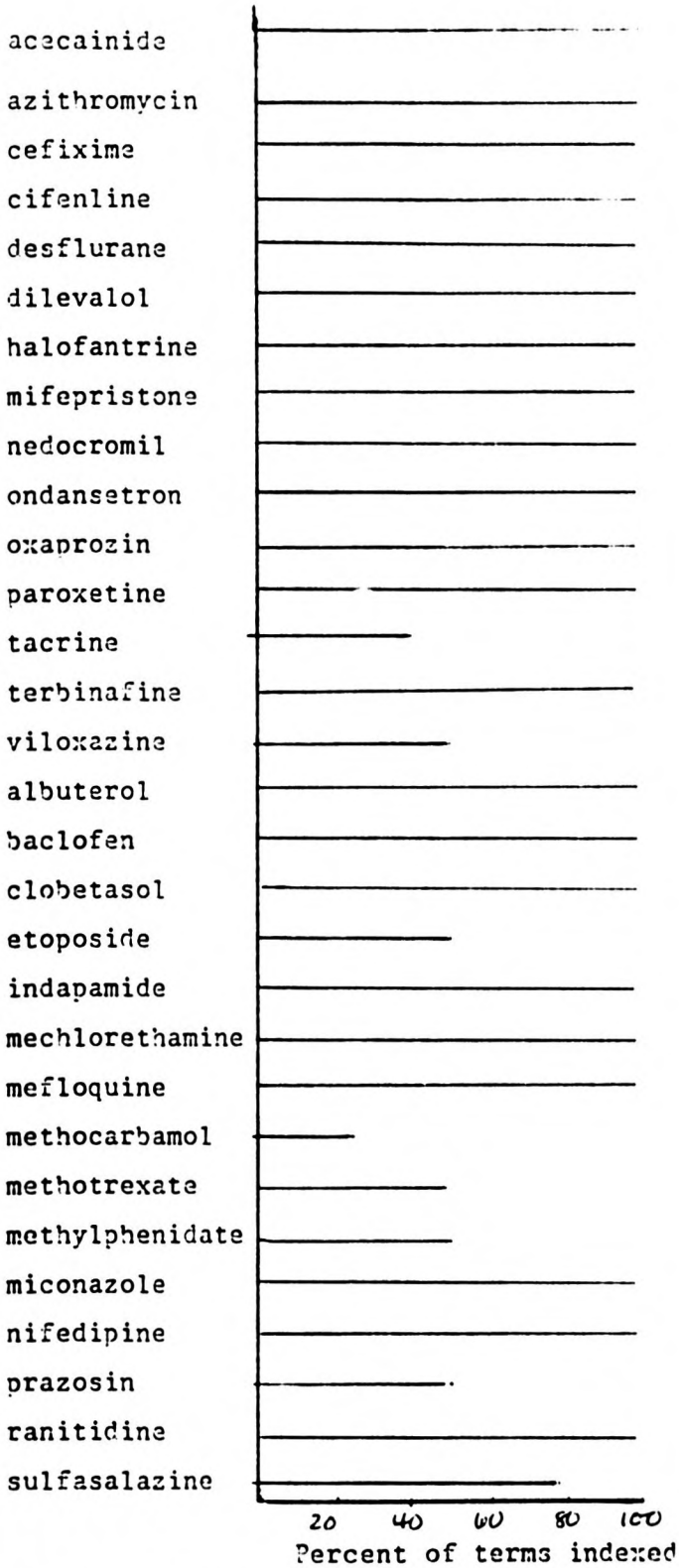

Fig. 18. Pharmaceutical News Index: Percent of Nonproprietary Terms Indexed

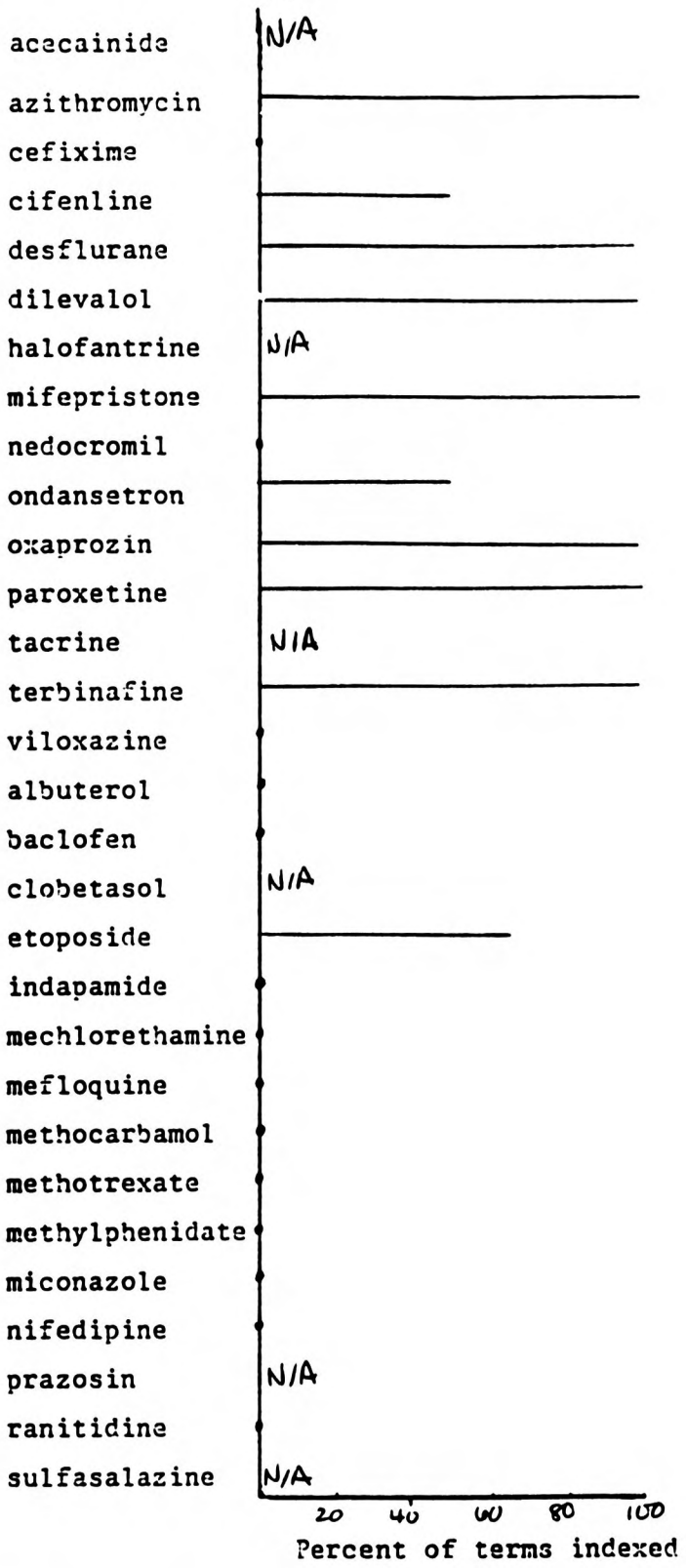

Fig. 19. Pharmaceutical News Index: Percent of Investigational Coded Indexed

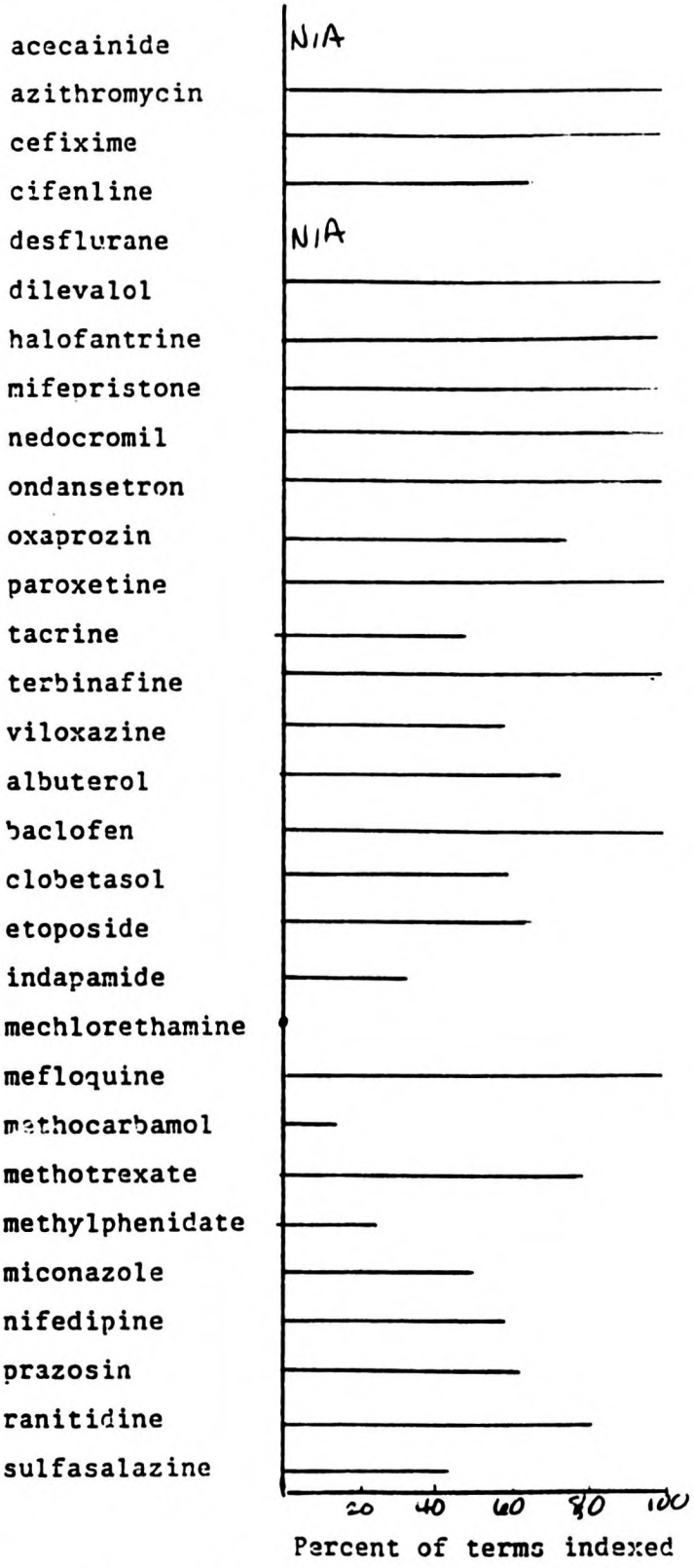

Fig. 20. Pharmaceutical News Index: Percent of Proprietary Terms Indexed

Of the 184 terms identified, 58% of them were indexed. Again, this is consistent with the business focus of the database and its coverage of advertising campaigns, market analyses, and prescription markets where proprietary terms would predominate. Also, the number of postings per proprietary term was significantly greater than the number in the two previous databases. Therefore in this database it is useful to identify multiple terms prior to searching as no one type of term predominates (e.g., nonproprietary term).

The difference in coverage of proprietary terms is noticeable in the results of the comparison questions. In question A, which compares one proprietary term to all proprietary terms, larger increases in recall are observed (see Figure 21). Similarly, in question B, moderate increases in recall are observed when searching all identified nonproprietary terms vs. a single term (see Figure 22).

The most notable change is in question C which compares recall for all known nonproprietary terms vs. all nonproprietary terms plus all proprietary terms (see Figure 23). Because recall among proprietary terms is greater (in general) in PNI, it is not surprising that searching on all nonproprietary terms plus proprietary terms significantly increases recall. These search results underscore the importance of searching multiple terms in this database.

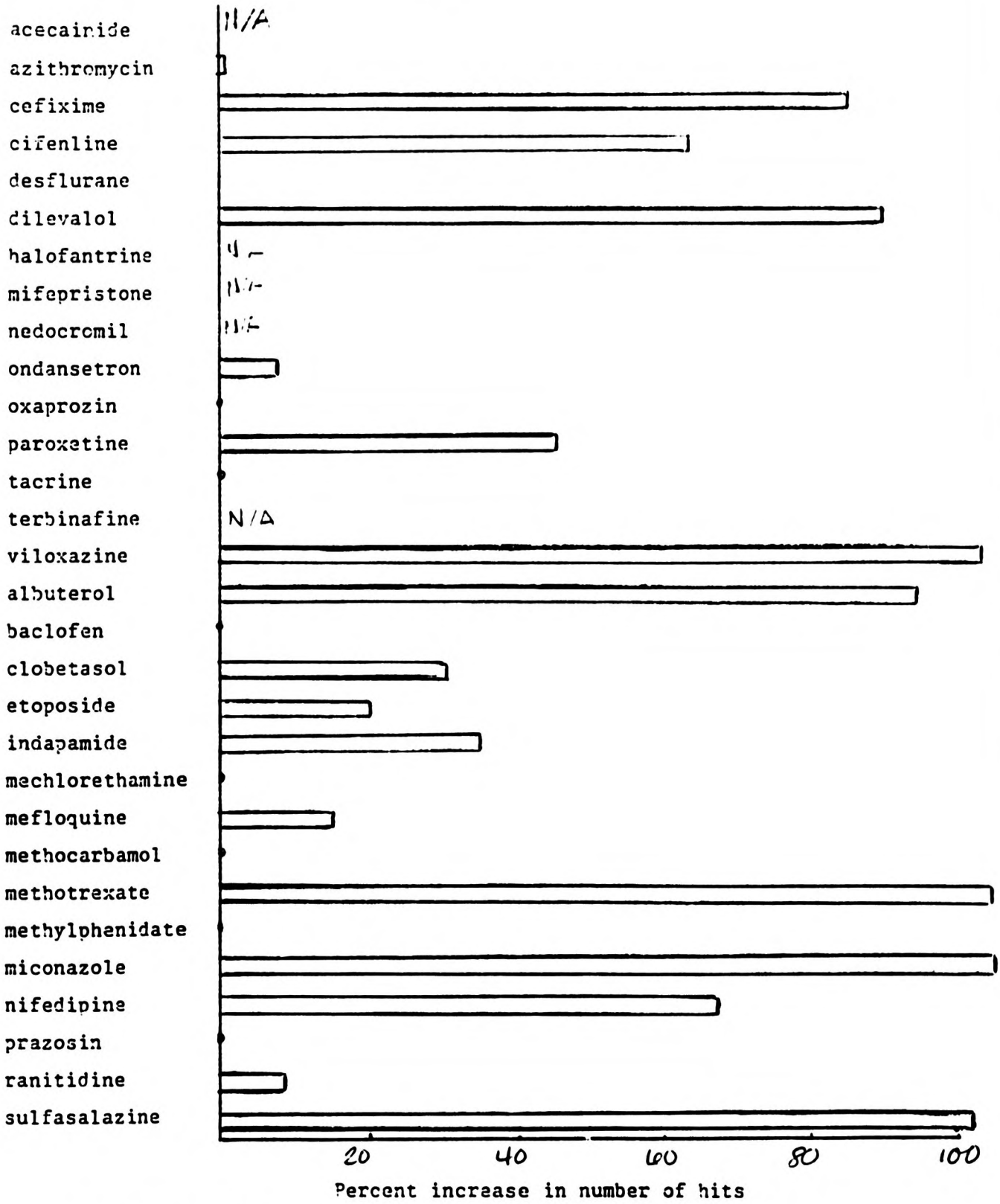

Fig. 21. Pharmaceutical News Index: One Proprietary Term vs. All Proprietary Terms

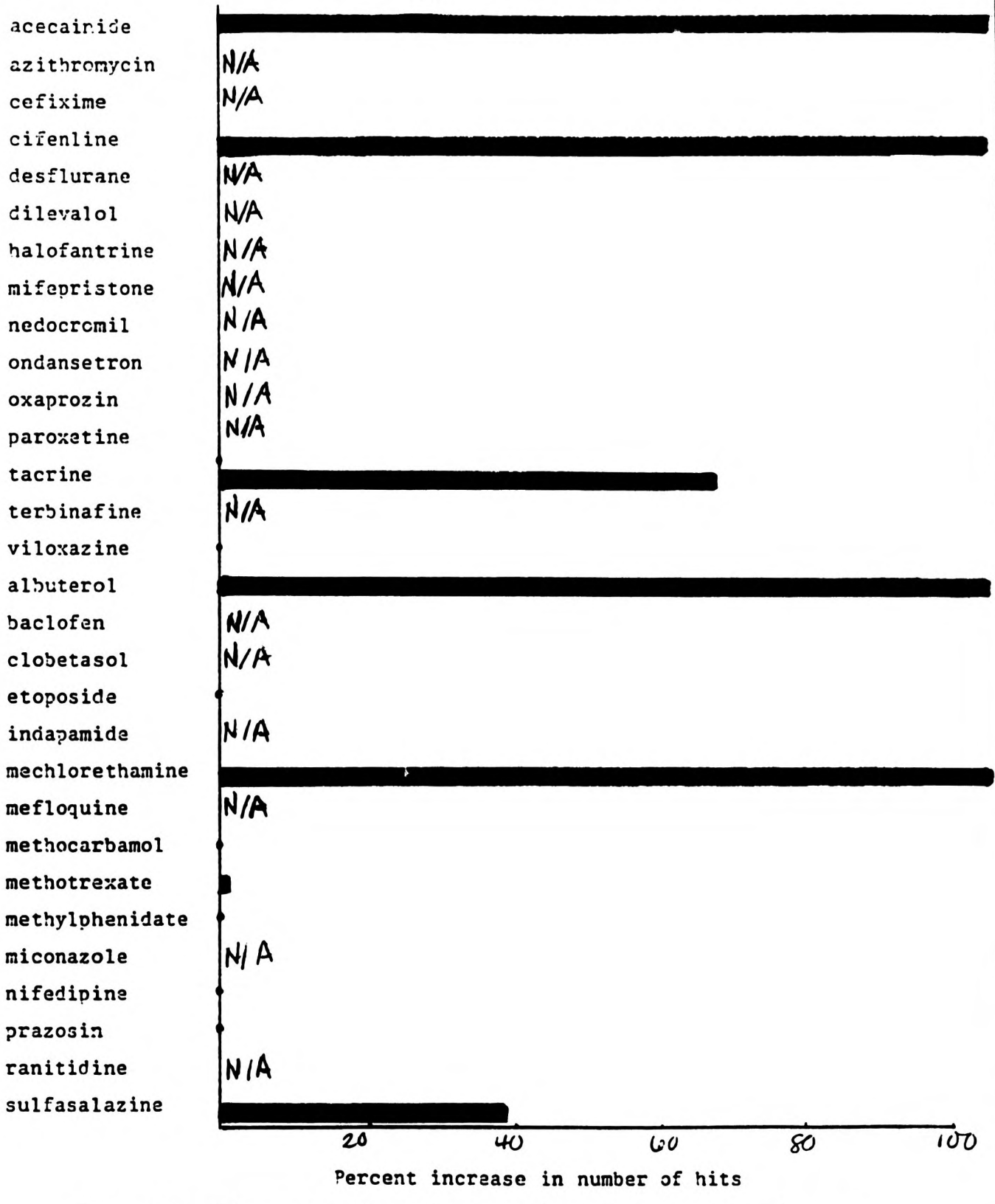

Fig. 22. Pharmaceutical News Index: One Nonproprietary Term vs. All Nonproprietary Terms

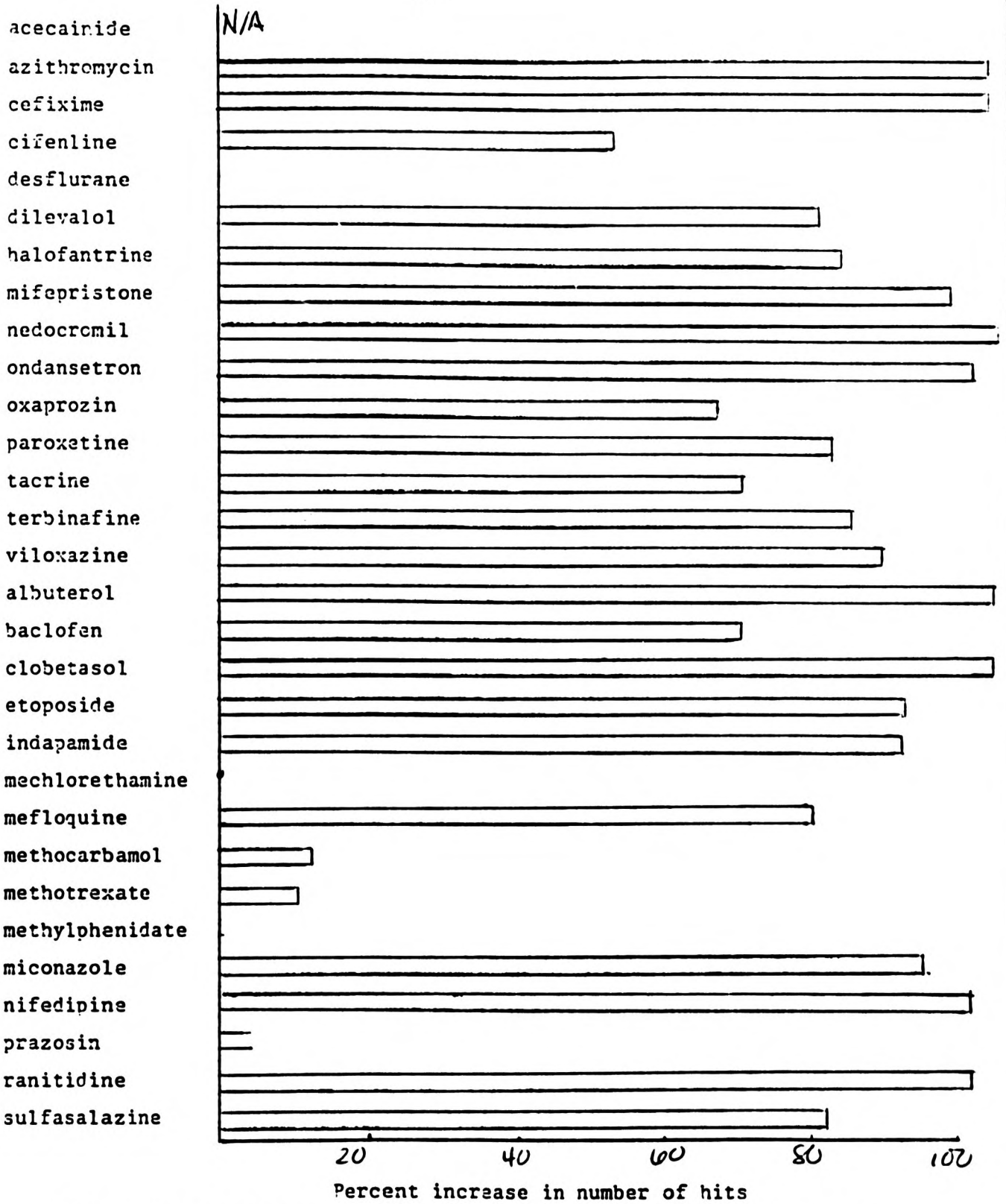

Fig. 23. Pharmaceutical News Index: All Nonproprietary Terms vs. All Nonproprietary Terms + All Proprietary Terms

5. CONCLUSIONS

'Drug information' is a very broad term that often draws on the literature of chemistry, various medical specialties, related sciences such as biochemistry and microbiology, and business events of the pharmaceutical industry. No single resource will contain all types of drug information and there is no one standard form of indexing 'drug information.' This study looked closely at the drug indexing of four distinctly different online databases, none of which uses controlled indexing.

Results showed the highest recall for each of the databases to be nonproprietary terms where recall ranged from 62% to 84% of the total nonproprietary terms identified. However use of the one current nonproprietary term does not consistently provide the best recall. In several cases, an older synonym or more commonly used acronym retrieved more postings than the current accepted term. Also, differences in recall were observed when searching the British produced database where the British spellings predominate and must be considered.

Analytical Abstracts indexes CAS Registry Numbers and a search on RNs yielded 70% retrieval. Combining nonproprietary terms with RNs yields very high recall. Of the thirty entities studied, this strategy retrieved information on twenty-nine of them. The RN offers a unique identifying code for each sub-

stance and is particularly useful in a chemically-oriented database such as this. However, the same drug can be assigned more than one RN depending on how it is described by the author. Thus the searcher must consider current RNs as well as previous RNs.

SCISEARCH and BIOSIS PREVIEWS had similar profiles of search results with the highest recall on nonproprietary terms (78% and 84%) and some recall on proprietary terms (24% and 32%). SCISEARCH derives its terms from the author's title and abstract (since January 1991 for some records). Thus, drug names appear in the form used in the original title (or author abstract when available) whether spelled out, hyphenated, nonproprietary, acronym, or proprietary. The complexities of "natural language" must be considered when searching this database.

Searching investigational codes yielded the highest recall (47%) in BIOSIS PREVIEWS and therefore is a useful concept to consider in building a search strategy for this database. As in SCISEARCH, searching on multiple nonproprietary terms can significantly increase recall (although not always) and adding proprietary terms does not significantly improve recall. Also noted was a significant increase in number of hits per term, possibly due to the availability of abstracts. The increased number of hits per term may reflect the broad coverage of all life sciences and the large size of the database. It would be useful to examine the retrieved citations and evaluate precision which would

require a more defined search question than was used in this study.

PNI also yielded high recall on nonproprietary search terms (75% of total) and as in SCISEARCH and BIOSIS PREVIEWS, there were more postings for older nonproprietary terms and acronyms, . PNI covered significantly more proprietary terms (58%) the highest of any of the four databases. Thus it is useful to consider all proprietary terms when searching PNI. A combination search strategy using multiple nonproprietary and proprietary terms yields the highest recall in PNI.

This study looked at only four representative databases from among hundreds. Likewise, this study used two dictionary files for term identification, where many other files could have been used. Also, only thirty drug entities (chosen arbitrarily) were studied from among thousands of drug entities. Still. some general conclusions can be drawn from this study.

First, the study underscores the need for identification and use of variant terminology when searching for information on drug entities. No single term can be relied on to retrieve complete information about a particular drug. The searcher must consider many things including:

- orientation of the database to be searched
- whether indexing is added (via descriptors or abstracts) or relies on author language for its terms
- where the database is produced (e.g., Great Britain vs,
- preferred nomenclature of the database to be searched

Secondly, the study revealed some potential pitfalls. One such problem to consider is possible errors in the databases, either in the dictionary files or the bibliographic files. In this study, one of the proprietary terms identified in the Marck Index was incorrectly spelled 'zantic' which should be 'zantac.' Without knowledge of this, many citations would be missed particularly since 'zantac' is the name of the product as distributed in the United States.

Another problem involves false drops when searching certain acronyms (which may also be common abbreviations for other terms) and certain proprietary names which are ambiguous. For example, one of the proprietary terms identified for 'ranitidine' was 'trigger' and this term retrieved primarily false drops such as "FDA Supplemental Appropriations Bill to 'Trigger' User Fees" and "Versatile 'Trigger' and Time-Delay Generator for Laser-Enhanced Time-of-Flight Mass Spectrometry." For this reason, the term 'trigger' was eliminated from this study. Similarly, the disproportionately higher number of postings for 'THA' in BIOSIS PREVIEWS might indicate several false drops.

This study showed that use of the nonproprietary term and its synonyms or acronyms yields the highest recall and at minimum the searcher must identify alternate nonproprietary terms. To improve recall, it is useful to identify other types of terms for searching depending on the database being searched. For example, use of CAS RNs in Analytical Abstracts or proprietary

terms in PNI together with nonproprietary terms improves recall. It is important to also consider that a term which yields few postings may yield the best information result and therefore looking only at quantative results does not indicate anything about the quality of results.

Future research might examine four or more databases of the same orientation (e.g., four pharmaceutical business databases) to compare indexing of drugs. It would also be useful to look at more than thirty drugs and/or do a more in-depth comparison of search results for newer drugs vs. older drugs.

In conclusion, searchers seeking information involving drug entities must begin with a basic understanding of the diverse nomenclature used in the pharmaceutical literature. The searcher must effectively utilize chemical dictionary files to determine likely alternate terms and must consider the orientation and indexing policies of the databases to be searched. Finally, the searcher would be well-served to plan the search strategy prior to going online but to observe closely initial search results and make changes in search strategy if necessary. The complexities of drug nomenclature makes searching for drug information somewhat of an art and the searcher must be creative when seeking such information.

6. APPENDIX

List of Drugs Studied

Newer Drugs	Older Drugs
acecainide	albuterol
azithromycin	baclofen
cefixime	clobetasol
cifenline	etoposide
desflurane	indapamide
dilevalol	mechlorethamine
halofantrine	mefloquine
mifepristone	methocarbamol
nedocromil	methotrexate
ondansetron	methylphenidate
oxaprozin	miconazole
paroxetine	nifedipine
tacrine	prazosin
terbinifine	ranitidine
viloxazine	sulfasalazine

BIBLIOGRAPHY

Barber, John, Sheena Moffat, Frances Wood, and David Bawden. "Case Studies of the Indexing and Retrieval of Pharmacology Papers." Information Processing & Management 24, no. 2 (1988): 141-150.

Brandsma, R., H.W.J. Deurenberg-Vos, S. Bakker, D.L. Brand-de Heer, R.H.J. Otten, and A. Pinatsis. "A Comparison of the Coverage of Clinical Medicine Provided by BIOSIS PREVIEWS and MEDLINE." Online Review 14, no. 6 (1990): 367-375.

Bronson, Roberta J. "Two Chemical Dictionary Files: CHEMLINE and CHEMID." Medical Reference Services Quarterly 11, no. 1 (Spring 1992): 1-15.

Buntrock, Robert E. "Name Searching of Chemical Compounds: A Rose by any Other Name..." Database 16 (February 1993): 102-104.

Deaves, J.C., and J.E. Pache. "Chemical and Numerical Indexing for the INSPEC Database." The Indexer 16, no. 3 (April 1989): 163-167.

Dorland's Illustrated Medical Dictionary, 1985 ed.

Feinberg, H. ed. Indexing Specialized Formats and Subjects. Metuchen, N.J.: Scarecrow, 1983.

Gilman, Alfred Goodman, Louis S. Goodman, and Alfred Gilman, eds. Goodman and Gilman's The Pharmacological Basis of Therapeutics. 6th ed. New York: MacMillan, 1980.

McCain, Katherine W., Howard D. White, and Belver C. Griffith. 'Comparing Retrieval Performance in Online Data Bases." Information Processing & Management 23, no. 6 (1987): 539-553.

McGowan, Anna Therese, and Dee A. Mater. "Identifying Drugs in the Chemical Literature: A Proposed Strategy." Medical Reference Services Quarterly 4, no. 4 (Winter 1985/86): 1-16.

Roth, Dana Lincoln. "The Role of Subject Expertise in Searching the Chemical Literature...and Pitfalls That Await the Inexperienced Searcher." Database 8 (February 1985): 43-45.

Snow, Bonnie. Drug Information: A Guide to Current Resources. Chicago: Medical Library Association, 1989.

Snow, Bonnie. "MARTINDALE ONLINE: Drug Info in a Detective's Toolkit." Database 11 (June 1988): 90-98.

Snow, Bonnie. "SCISEARCH Changes: Abstracts and Added Indexing." Online 15 (September 1991): 102-106.

Snow, Bonnie. "Trade Names in Medicine: Searching For Brand Name Comparisons and New Product News." Database 15 June 1992): 99-105.

Snow, Bonnie. "Why Use a Database Thesaurus?" Online 9 (November 1985): 92-96.

Soremark, Gun. "MEDLINE Versus EMBASE: Comparing Search Quality." Database 13 (December 1990): 66-67.

Tousignaut, Dwight R. "Indexing: Old Methods, New Concepts." The Indexer 15 (October 1987): 197-204.

Van Camp, Ann J. "The MeSH Vocabulary File and CHEMLINE: Online Sources of Chemical Nomenclature." Medical Reference Services Quarterly 3, no. 1 (Spring 1984): 1-17.

White, Howard D., and Belver C. Griffith. "Quality of Indexing in Online Data Bases." Information Processing & Management 23, no. 3 (1987): 211-234.