Health Information and Libraries Journal

Health Libraries Group | cilip Chartered Institute of Library and Information Professionals

# Beyond relevance and recall: testing new user-centred measures of database performance



| | |
|---|---|
| Journal: | *Health Information and Libraries Journal* |
| Manuscript ID: | draft |
| Manuscript Type: | Original Article |
| Keywords: | Databases, Bibliographic, Information Storage and Retrieval, Medline, Students, Nursing |
| | |

## ABSTRACT

**Background** Measures of the effectiveness of databases have traditionally focused on recall, precision, with some debate on how relevance can be assessed, and by whom. New measures of database performance are required when users are familiar with search engines, and expect full text availability.

**Objectives** This research ascertained which of four bibliographic databases (BNI, CINAHL, MEDLINE and EMBASE) could be considered most useful to nursing and midwifery students searching for information for an undergraduate dissertation.

**Methods** Searches on title were performed for dissertation topics supplied by nursing students (n=9), who made the relevance judgements. Measures of Recall and Precision were combined with additional factors to provide measures of Effectiveness, while Efficiency combined measures of Novelty and Originality, and Accessibility combined measures for Availability and Retrievability, based on Obtainability.

**Results** There were significant differences among the databases in Precision, Originality, and Availability, but other differences were not significant (Friedman test). Odds ratio tests indicated that BNI, followed by CINAHL were the most effective, CINAHL the most efficient, and BNI the most accessible.

**Conclusions** The methodology could help library services in purchase decisions as the measure for accessibility, and odds ratio testing helped to differentiate performance.


## INTRODUCTION

Although information literacy programmes for undergraduate nursing programmes appear effective in improving database searching skills, [1 2 3 4] databases remain more difficult to search successfully for many students or professional practitioners who may search infrequently. Most may prefer to search using search engines such as Google, but that may not lead to the best evidence, although Google is efficient in the sense that some information is supplied quickly (and it is often good information although not necessarily the best). Libraries purchasing databases for use by nursing students need to be assured that the databases provide value for money. The questions for service providers and users concern the ease of use, the unique content provided by a particular database, the relevance of items retrieved, and the local availability of the full text for the references retrieved in a search.

The traditional measures used in evaluation of database have been the measurement of Precision and Recall. Precision and Recall which according to Van Rijsbergen[5] "attempt to measure what is now known as the effectiveness of the retrieval system. In other words it is a measure of the ability of the system to retrieve relevant documents while at the same time holding back non-relevant one) are the traditional measures used in

evaluation. Precision is relevant and retrieved documents divided by all retrieved documents; Recall is relevant and retrieved documents divided by all relevant documents. Korfhage[6] raises three problems with Recall and Precision: 1) Precision can be determined exactly, recall cannot; 2) Recall and Precision are not necessarily significant to the user and 3) Recall and Precision are related and so each alone provides an incomplete picture of the system's effectiveness. Schamber et al [7] reviewed the definition of relevance and concluded that relevance is a multi-dimensional cognitive concept, and dependent on users' perceptions of information and their own information needs, dynamic (as quality judgements may change), and complex, but measurable if the user's perspective is the focus of investigation. Other researchers have considered multiple levels of relevance, [8] partial relevance, [9 1011] and changing frameworks for relevance depending on the stage of problem solving.[12] Other researchers have focused more on the usefulness to the user; the value of the information retrieved and have examined quality attributes such as goodness, usefulness, currency, accuracy, and trustworthiness.[13] Clearly difficulties exist with the subjective and dynamic nature of relevancy.

Various studies have examined the performance of databases themselves, examining their capacity to provide unique information, good coverage, and references to material that is easily available. Many of the studies have focused on the needs of systematic reviewing for the Cochrane Collaboration. McDonald et al.[14] evaluated the coverage of MEDLINE, EMBASE, BIOSIS and PsycLIT for psychiatry journals, Brettle and Long[15] attempted to locate research papers useful for a systematic review across six selected databases. Subirana et al. [16] compared CINAHL, MEDLINE and EMBASE for their effectiveness in contributing studies for a systematic review, and in another study compared the recall and overlap of articles from MEDLINE and CINAHL.[17] Watson and Perrin[18] compared CINAHL and MEDLINE for coverage of allied health journals and the relevance of items retrieved, Yonker et al[19] examined four databases for their coverage of forensic medicine, and Suarez-Almazor et al[20] examined coverage of controlled clinical trials. The methods used in some studies focus more on the functionality of the databases to help users with precision and recall. Watson and Richardson[21] compared the effectiveness of broad and narrow search strategies, Marson and Chadwick[22] compared basic, comprehensive and hand searches, and Jenuwine and Floyd[23] compared textword searches with subject searches within one journal. Keyword searching was used by Gehanno et al.[24] and Okuma.[25] Some studies test using single topics [17 19 21 22 23] Other studies have used multiple topics, for example Brown[26] on medical topics, Brown[27] on retrieval of pharmaceutical information, Abd Brand-de Heer[28] for clinical medicine, and McCain et al. [29] for medical behavioural sciences. Studies that have focused on nursing students include Brazier and Begley's comparison of MEDLINE and CINAHL that used topics selected by nursing students, searched on titles only, and employed relevancy assessments by the nursing students.[30] Okuma[25] used three expert judges to determine relevancy in a comparison of the suitability of MEDLINE and CINAHL for nursing students. Burnham and Shearer[31] compared

CINHAL, EMBASE and MEDLINE using three nursing topics, and used three nursing faculty members to conduct relevancy assessments.

There is no agreed methodology for assessing the performance of databases. Topic evaluation (single or multiple) often focus on inclusion, coverage and novelty, but the methods used for searching are not consistent (keyword, thesaurus term, title, title plus abstract, or combination of these possibilities). Statistical analysis tends to be over reliant on Precision and Recall which, whilst valuable measures, emphasise what a database can retrieve - not what a user might actually want. Relevancy judgements are sometimes made by independent judges and sometimes by users themselves. Uniqueness and access to information is touched upon, but these aspects are not analysed together with Precision and Recall. In addition, only a single study compared CINAHL, MEDLINE and EMBASE; databases often considered for use for nursing students, and no study analyzed BNI, a database for nursing students produced in the UK.

**OBJECTIVES**

This research examined which of four bibliographic databases (BNI, CINAHL, MEDLINE and EMBASE) could be considered most useful to nursing and midwifery students searching for information for an undergraduate dissertation. The traditional approach of testing information retrieval using recall and precision is combined with the additional factors of efficiency and accessibility. It produces quantifiable, measurable criteria that can be analyzed using statistical tests.

**SETTING AND ETHICAL APPROVAL**

The research was conducted at Homerton School of Health Studies (now part of Anglia Ruskin University) in 2005-2006. Ethical approval for this research has been granted by the following committees: Peterborough and Fenland Local Research Ethics Committee; Cambridge City PCT Ethics Committee; Peterborough and Stamford Hospitals NHS Trust Research Ethics Committee; Homerton School of Health Studies Research Ethics Committee.

All potential participants were given a letter of invitation that outlined the nature of the study and the extent of involvement. Those that chose to take part were given a Participant Information Sheet that stipulated the procedures involved and explained the informed consent. All participants were asked to sign a Consent Form and one copy was held by the researcher and one by the participant.

**METHODS**

OVID is the database provider for all four databases analyzed in this research. It was chosen as it was available at the study location. As the research investigated the content of the databases – not the search interface or indexing methods – the results ought to be identical if other database providers were used. The four databases used were: British Nursing Index, CINAHL, EMBASE and MEDLINE

Population and sample:

To obtain the raw data for analysis, students enrolled on the following dissertation modules at a Higher Education School of Nursing were invited to participate.

1. Entry to Register Nursing degree
2. Entry to Register Midwifery degree
3. Continuing Professional Development degree (either Primary and Community Care or Health Studies)

From these groups two of the three Entry to Register Nursing students took part, two out of the four Entry to Register Midwifery students took part; and five out of 16 Continuing Professional Development students took part (n=9 in total).

The students were asked to supply either a working title or a specific subject area that they would be using as a basis for their dissertation. These topics were then used by the researcher [PS] to formulate search strings and the resulting data obtained from the bibliographic databases were used in the analysis.

Students were given the option not to take part, and were not expected to agree to take part immediately. The researcher allowed up to a week of reflection if the potential participant so wished. They were approached at the time they were compiling their dissertation proposals.

The researcher conducted the searches as soon as possible after receiving the topics and forwarded the list of articles to the students. A stamped address return envelope was enclosed for those students who were based away from the college and who would have had difficulty returning the list in person.

**Information retrieval**

The searches included in the data collection took the form of a keyword search within record titles only. The rationale for this decision was

- To remove **bias of extra abstracts** on MEDLINE and EMBASE (not an aspect tested in this research)
- To negate differences in indexing practices – and an **identical search strategy** for each database analysis was a fairer test (and probably a fairer reflection of the way the students would search)
- A user search would not have been appropriate as each user would search differently and while this may have given an indication of how students search - it wouldn't have given a true reflection of the **content** of the databases

Searching within titles only has also been performed by Subirana et al[17] Brazier and Begley,[30] and Okuma.[25] From the perspective of assessing the value of the databases to users, there are limitations in this approach as a typical search would not be limited to titles only, but this approach provides the fairest test for all the databases.

**Precision and Recall**

Users may have different conceptions of a comprehensive search to librarians[32], but choosing the dissertation as the situation of need meant that students would be interested in maximum recall. Additional criteria were developed to determine the overall usefulness of the databases. The results from the four databases were pooled to determine all relevant documents.

**Relevancy judgments**

A simple dichotomous "yes/no" scale was used in order to negate order effects and minimize subjectivity. This type of scale is also easier to analyse. The students who provided topics conducted the relevancy judgements (as opposed to 'experts' or the researcher) as they were familiar with the topic, and to add authenticity to the search.

**Pilot study**

A small pilot study was first conducted to test the methodology that would be adopted for the research. Two students already enrolled on an undergraduate dissertation module agreed to take part in the study and supplied their working titles to the researcher. From these working titles simple search strings were formulated concerning the main topics using relevant truncation and Boolean operators and performed article title searches within the four selected databases. One search was then limited down to the last two years to restrict the number of hits to no more than 100 per database. The second search did not need limiting. The results were then 'deduplicated' and compiled into a list of article titles only. The two students then made relevancy judgments on this list. This pilot study worked in a manner deemed acceptable to the researcher and the same procedure then used for the main study. However, it became apparent that a title search on BNI included a brief summary of the reference

and as such a certain amount of filtering was necessary to only include those references in which the target word or words appeared in the actual title. This was noted for the main research project.

These two searches were not used in the final analysis as a database update had occurred which may have had implications for the research results.

**Evaluation methods**

The data obtained from the searches can be interrogated using a range of statistical tests. The Statistical Package for the Social Sciences (SPSS) was used to test the data. To answer the research question outlined earlier three tests were chosen and are detailed below within two groups.

***GROUP A***

For Group A the 'nature' of the data obtained from the searches were analysed within 3 distinct areas: effectiveness, efficiency and accessibility.

**EFFECTIVENESS**

***This is a combination of Precision and Recall based on Relevancy (although the traditional view is that high recall=low precision and vice-versa [an inverse relationship]).***

1. Precision is tested in the traditional way (relevant articles in a database search divided by number of articles in search)

$$S_{1,2,3\ etc} \quad \frac{RD_1}{D_1} : \frac{RD_2}{D_2} : \frac{RD_3}{D_3} : \frac{RD_4}{D_4}$$

*Where S=each topic, RD=relevant hits for each respective database, and D= number of hits.*

2. Recall is tested in the traditional way (relevant articles in a database search divided by total number of relevant articles in all database searches)

$$S_{1,2,3\ etc} \quad \frac{RD_1}{\quad} : \frac{RD_2}{\quad} : \frac{RD_3}{\quad} : \frac{RD_4}{\quad}$$

$$\frac{}{RD_{1,2,3,4}} \quad \frac{}{RD_{1,2,3,4}} \quad \frac{}{RD_{1,2,3,4}} \quad \frac{}{RD_{1,2,3,4}}$$

*Where S=each topic, and RD=relevant hits for each respective database*

***NB: in order to obtain a score for 'all relevant articles' the total/pooled number of relevant articles for the four databases is used.***

**EFFICIENCY**

***This is a combination of Novelty and Originality based on Uniqueness.***

1. Novelty is calculated as the number of relevant records retrieved that are unique as a percentage of the number of relevant records retrieved in the search.

$$S_{1,2,3\ etc} \quad \frac{UD_1}{RD_1} : \frac{UD_2}{RD_2} : \frac{UD_3}{RD_3} : \frac{UD_4}{RD_4}$$

*Where S=each topic, UD=unique relevant hits for each database and RD= number of relevant hits in a database.*

2. Originality is calculated as the number of relevant records retrieved that are unique as a percentage of the total number of unique records.

$$S_{1,2,3\ etc} \quad \frac{UD_1}{UD_{1,2,3,4}} : \frac{UD_2}{UD_{1,2,3,4}} : \frac{UD_3}{UD_{1,2,3,4}} : \frac{UD_4}{UD_{1,2,3,4}}$$

*Where S=each topic, and UD=unique relevant hits for each respective database*

**ACCESSIBILITY**

***Accessibility is tested on a 'yes/no' scale of ease/cost of obtaining the relevant articles (yes=available within college or electronic, no=not available).***

***Accessibility is a combination of Availability and Retrievability and is based on Obtainability.***

1. Availability is calculated as the number of relevant records retrieved that are obtainable as a percentage of the number of relevant records retrieved in the search.

$$S_{1,2,3\ etc} \quad \frac{OD_1}{RD_1} : \frac{OD_2}{RD_2} : \frac{OD_3}{RD_3} : \frac{OD_4}{RD_4}$$

*Where S=each topic, OD=number of obtainable relevant hits for each respective database and RD= number of relevant hits in a database.*

2. Retrievability is calculated as the number of relevant articles retrieved that are obtainable divided by the total number of obtainable relevant articles.

$$S_{1,2,3\ etc} \quad \frac{OD_1}{OD_{1,2,3,4}} : \frac{OD_2}{OD_{1,2,3,4}} : \frac{OD_3}{OD_{1,2,3,4}} : \frac{OD_4}{OD_{1,2,3,4}}$$

*Where S=each topic, OD=number of obtainable relevant hits for each respective database.*

**Data analysis**

Tests of significance were calculated on the six criteria using Friedman's Test because:

1. It is a non-parametric test as data are not from a normal distribution.
2. It uses ranks which counteracts any particular skew or bias from any individual search.
3. It can test more than 2 variables.

There are no major assumptions for this test as it is distribution-free.

**Hypotheses for this test**

Set 1:

$H_0$: There is no difference in the precision of the four selected databases.

$H_1$: There is a difference in the precision of the four selected databases.

$H_0$: There is no difference in the recall of the four selected databases.

$H_1$: There is a difference in the recall of the four selected databases.

Set 2

$H_0$: There is no difference in the novelty of the four selected databases.

$H_1$: There is a difference in the novelty of the four selected databases.

$H_0$: There is no difference in the originality of the four selected databases.

$H_1$: There is a difference in the originality of the four selected databases.

Set 3

$H_0$: There is no difference in the availability of the four selected databases.

$H_1$: There is a difference in the availability of the four selected databases.

$H_0$: There is no difference in the retrievability of the four selected databases.

$H_1$: There is a difference in the retrievability of the four selected databases.

***GROUP B***

The second statistical analysis is an Odds ratio test which tests whether a database is more likely to retrieve a relevant (or unique; or obtainable) article rather than an irrelevant (or non-unique; or not obtainable) article. The data are compiled in a simple 2 X 2 contingency table. This process was then repeated for databases 2, 3, and 4. (Tables 1, 2, 3).

**Data analysis**

The odds ratio is used to test the likelihood of a particular database locating

1. relevant articles - effectiveness
2. unique articles - efficiency
3. obtainable articles - accessibility

This test is used to ascertain 'risk' in preference to the 'relative risk test' as the odds ratio will determine whether a database will select relevant as opposed to irrelevant articles rather than determine the chances of locating relevant articles from the search alone[33] .

**Hypotheses for this test**

$H_0$: Database X is no more likely to retrieve a relevant hit than an irrelevant hit.

$H_1$: Database X is more likely to retrieve a relevant hit than an irrelevant hit.

$H_0$: Database X is no more likely to retrieve a relevant hit that is unique than a relevant hit that is not unique.

$H_1$: Database X is more likely to retrieve a relevant hit that is unique than a relevant hit that is not unique.

$H_0$: Database X is no more likely to retrieve a relevant hit that is obtainable than a relevant hit that is not obtainable.

$H_1$: Database X is more likely to retrieve a relevant hit that is obtainable than a relevant hit that is not obtainable.

(Where X is either one of BNI, CINAHL, EMBASE, MEDLINE)

## RESULTS

### Raw data

From these results (table 4) it is clear that each database has its merits depending on the topic area that is being searched. The medical databases EMBASE and MEDLINE outperform the nursing databases BNI and CINAHL in searches 3 and 7, whilst EMBASE also does well for search 8. These three searches did not include a search string based on 'nurses'. EMBASE struggles in searches four and five which are focussed on nurses; MEDLINE however still performs comparatively well with the nursing databases for these two searches. For most of the searches the medical databases locate more relevant material, but also retrieve more irrelevant material. In three searches BNI locates only relevant material (100% precision), whilst CINAHL has 100% precision for two searches. On four occasions BNI has a 100% availability rating for relevant articles, but also fails to locate unique articles on five occasions.

The data can be more robustly tested by pooling these results.

### *GROUP A*

After pooling the data (combining the results for the nine separate searches) the mean ranks for each of the six criteria under investigation in Group A can be plotted as shown in Graph 1. *(Note the higher the ranking score the better the performance of the database.)*

### Precision

The two nursing databases: BNI and CINAHL have much higher overall ranks for precision than the medical databases EMBASE and MEDLINE.

### Recall

The inverse relationship between precision and recall is clear in the above table as BNI now has the lowest rank. CINAHL still performs well, but MEDLINE has the highest rank.

### Novelty

CINAHL is the top database for Novelty, BNI the lowest ranked. MEDLINE again outranks EMBASE.

### Originality

BNI again has the lowest rank, this time for Originality. MEDLINE has the highest rank, with CINAHL again

performing well. EMBASE registers its highest average rank thus far, but still only lies third.

**Availability**

The two nursing databases BNI and CINAHL have much higher ranks for Availability than the medical databases. BNI has the highest overall rank. Both medical databases register low average ranks.

**Retrievability**

Retrievability ranks appear to be the most consistent across the four databases. CINAHL has the highest rank, with EMBASE the lowest.

Across all six criteria CINAHL is not ranked lower than second, BNI and EMBASE are both lowest ranked for three criteria (although BNI is top ranked twice), and MEDLINE is top ranked twice.

In order to test the hypotheses given in the methodology section, these results can be quantified to allow the calculation of significance levels (Table 5).

Results for Precision, Originality and Availability are significant to the $p>0.05$ level. Recall and Novelty - whilst not statistically significant - also show considerable differences. Retrievability cannot be considered significant.

Thus:

The null hypothesis that there is no difference in precision between the databases is *rejected.*

The null hypothesis that there is no difference in recall between the databases is *supported.*

The null hypothesis that there is no difference in novelty between the databases is *supported.*

The null hypothesis that there is no difference in originality between the databases is *rejected.*

The null hypothesis that there is no difference in availability between the databases is *rejected*.

The null hypothesis that there is no difference in retrievability between the databases is *supported*.

***GROUP B***

**Odds ratios**

The Odds Ratio test can be used to compare the four databases in the following manner **(ad)/(bc). For example: relevancy is determined by (relevant hits X not relevant misses)/(not relevant hits X relevant misses).** (Table 6).

**Effectiveness** (the likely odds that a database will retrieve a 'relevant' hit) Graph 2

The odds ratio test infers that the nursing databases are more likely to retrieve relevant hits than irrelevant hits. BNI has the highest effectiveness rating (1.717) as it retrieves very few irrelevant hits. CINAHL also retrieves a much higher proportion of relevant hits and has an odds ratio of 1.392. Although the medical databases EMBASE and MEDLINE retrieve more relevant hits overall, they lose 'effectiveness' by retrieving a higher proportion of irrelevant hits than the nursing databases.

**Efficiency** (the likely odds that a database will retrieve a 'unique + relevant' hit) Graph 3

CINAHL has the highest efficiency rating suggesting that is the database that is more likely to retrieve relevant hits that are not contained on other databases. MEDLINE and EMBASE contain the highest number of unique/relevant hits, but lose 'efficiency' due to their high retrieval rate of not unique/relevant articles. BNI has very few unique/relevant articles.

**Accessibility** (the likely odds that a database will retrieve an 'obtainable + relevant' hit) Graph 4

BNI clearly is the database that contains readily accessible materials for students. Despite having the lowest number of obtainable/relevant hits, the fact that so few of the relevant hits on BNI were not easily available gave this database a very high accessibility rating. CINAHL also had a higher proportion of obtainable/relevant hits to not obtainable/relevant articles and also had a high accessibility rating. The medical databases EMBASE and MEDLINE, whilst retrieving a comparable amount of obtainable/relevant hits retrieved many more hits that were not readily accessible and thus had low accessibility ratings.

## DISCUSSION

Although using relevancy criteria as the basis for quantitative analysis is a contentious issue, it is the only way to test the effectiveness (precision and recall) of databases. Subjectivity was reduced by using a 'yes/no' scale to negate order effects and relevancy enhanced by using the students to judge the results of the searches. Bias was also reduced by searching within title fields only, which also enabled identical search strategies to be performed across the four databases. The approach ignored interface issues, but these, as well as the skills of

the searcher and the subject indexing, may affect the final valuation of the usefulness of a database for a user.

The way the search is conducted may influence the relevancy of the search results. In some cases broader headings have been used than would normally have been in order to obtain analyzable results from all four databases. This may result in a lower overall relevancy ranking. Identical search strings, however, were used for all four databases to negate this effect.

Judgements had to be made by the participants on the relevancy of the record title only. No abstracts or additional information such as authors or journal titles were given. Again this was to enable consistency across all four databases to be maintained, but the student judge could have lost interest after screening a certain number of references, and the simple (yes/no) scale may be used inconsistently.

A large part of the analysis is novel to this research and cannot therefore be compared to previous studies. Many studies have evaluated MEDLINE and CINAHL[17 25 30] or included EMBASE as well.[31] No study has compared these three databases with the addition of BNI. This research also introduces novel testing criteria. However, the descriptive results that MEDLINE retrieves twice as much relevant material for nursing students than CINAHL is in line with research conducted by Brazier and Begley[17]; and the finding that much material on CINAHL is unique is supported by Okuma[25]. However, although Subirana et al.[17] found that MEDLINE had a higher rating for both Recall and Precision, this study found that MEDLINE does have a higher Recall rating but a lower Precision rating than CINAHL. Burnham and Shearer's assertion that searching CINAHL alone would not be sufficient[31] is in line with these findings, however their conclusion that MEDLINE finds enough relevant material to be searched alone is debatable.

**Group A**

This first group aimed to test six hypotheses within the three areas of: effectiveness, efficiency and accessibility.

The plotting of the mean ranks for the six criteria: precision, recall, novelty, originality, availability and retrievability enabled a visual comparison of the results for the four databases. These initial basic descriptive results showed the following:

- BNI had the best average rank for precision and availability, the lowest rank for recall, novelty and originality, and was ranked third for retrievability
- CINAHL had the best average rank for novelty and retrievability and was ranked second for all other criteria

- MEDLINE was top ranked for recall and originality, second ranked for novelty and retrievability, and third ranked for precision and availability
- EMBASE did not achieve a top or second highest average rank for any criterion. It was ranked third for recall, novelty and originality, and fourth for precision, availability and retrievability

From these results the inference is that CINAHL is a consistently 'good' performer within these criteria, and EMBASE the least good. BNI is sometimes very good, other times very poor, and MEDLINE consistently good although less so than CINAHL.

The use of the Friedman test yields levels of significance that can be used to test the six hypotheses. This test showed that whilst superficial differences can be seen from the descriptive statistics some significant differences do exist. In two of the three rejected null hypotheses ('precision' and 'availability' criteria) BNI has the highest rank and MEDLINE has the highest rank for the other rejected hypothesis concerning 'originality'. Does this make these two databases the most useful? Not necessarily. The significant difference in the 'precision' criterion is due to the high ranks for BNI and CINAHL coupled with the low rank for EMBASE. This reinforces the difference between the databases; it doesn't suggest that BNI is significantly better than the other three databases. This is also true for 'originality' where CINAHL and EMBASE have high ranks along with MEDLINE, but BNI has a very low rank. This may suggest that BNI is the least useful database for this criterion, but we can't say that any of the others are the most useful. For 'availability' BNI and CINAHL have much higher average ranks than both MEDLINE and EMBASE, so for this criterion we could surmise that the nursing databases are more useful than the medical databases. What is clear is that there are differences between the databases when they are tested within these parameters.

**Group B**

The second group of statistics - odds ratios - confirm that there are differences between databases.

It shows the likelihood of a particular database finding relevant, unique or accessible articles when a search is conducted and thus does not compare the performance of the four selected databases with each other. This test takes the results for Group A a step nearer to finding the most useful database. The odds ratio test shows whether a database is more likely to retrieve:

a. relevant hits rather than irrelevant hits
b. relevant hits that are unique rather than relevant hits that are not unique
c. relevant hits that are obtainable rather than relevant hits that are not obtainable

As such these three results can be used to test the effectiveness, efficiency, and accessibility of each database with single scores rather than using the two scores as shown in Group A.

Taking the graphs at face value, BNI appears to be the most effective and accessible database with CINAHL the most efficient. Looking more deeply we can see that the odds ratios for all databases in the effectiveness and efficiency ratings are very low and the confidence intervals quite large. As the confidence intervals cross for all four databases for the effectiveness ratings, the results must be treated with caution: we cannot predict that these results would occur in a further study. For efficiency we can be less cautious as the confidence intervals for CINAHL and MEDLINE do not cross, and CINAHL only marginally crosses with BNI and EMBASE. Nevertheless, we still cannot predict that these results would occur again. The accessibility results are much more definite. BNI has very high odds that it will retrieve accessible relevant material, far more than both MEDLINE and EMBASE. CINAHL also has a high odds ratio, again far more than MEDLINE and EMBASE, but does cross with the confidence interval of BNI.

The crossing of confidence intervals together with the low ratings across the range for uniqueness should be expected. The odds ratio is comparing each database individually against the pool of data; it does not compare the four databases with each other. Therefore the graphical representation is of greater importance as it provides a clear indication of which database is most useful; even though statistical testing cannot be conclusive.

Although pooling the data in this way ‘quashes’ any differences for individual searches it does enable more sophisticated testing to be used. The raw data shows that some databases perform well for particular searches, but the pooled data can be used to determine the likelihood that certain databases are generally better than others across a broad range of topics.

**CONCLUSION**

Whilst it was statistically challenging to come to a definitive conclusion as to which database could be considered the most useful, this approach confirms that searching a single database is likely to miss relevant articles, and that some databases may be general good performers (e.g. CINAHL). The methodology could be useful for other library services as the measure for accessibility differentiated between databases clearly and the odds ratios might help in decision making about database purchase.

Further research to investigate what users really want from a database is needed. It is clear that qualitative data on the needs and preferences of the student would provide additional data that could supplement this research. By examining whether students prefer a database that provides a few relevant hits that are easily accessible, would like to find unique articles, or want to locate as many relevant articles as possible, further analysis could be performed to ascertain which of these databases is in fact the most useful.

## REFERENCES


[1] Shorten, A, Wallace, M.C., & Crookes, P.A, Developing information literacy: a key to evidence-based nursing. International Nursing Review 2001, 48, 86-92.
[2] Wallace, M.C., Shorten, A., & Crookes, P.A. Teaching information literacy skills: an evaluation. *Nurse Education Today* 2000, **20**, 485-489.
[3] Kaplan-Jacobs, S., Rosenfeld, P., & Haber, J. Information literacy as the foundation for evidence-based practice in graduate nursing education: a curriculum-integrated approach. *Journal of Professional Nursing* 2003, **19**, 320-328.
[4] Verhey, M.P. Information literacy in an undergraduate nursing curriculum: development, implementation and evaluation. *Journal of Nursing Education* 1999, **38**, 252-259.
[5] van-Rijsbergen, C.J. *Information retrieval*. 2nd edn. London: Butterworths, 2004
[6] Korfhage, R.R. *Information storage and retrieval*., New York: John Wiley, 1997.
[7] Schamber, L., Eisenberg, M.A., & Nilan, M.S. A re-examination of relevance: toward a dynamic, situational definition. *Information Processing and Management* 1990, **26**:755-776.
[8] Barry, C.L., & Schamber, L. Users' criteria for relevance evaluation: a cross situational comparison. *Information Processing and Management* 1998, **34**, 219-236.
[9] Spink, A., & Greisdorf, H. Regions and levels: mapping and measuring users' relevance judgments. *Journal of the American Society for Information Science* 2001, **52**, 161-173.
[10] Spink, A., Greisdorf, H., & Bateman, J. From highly relevant to nonrelevant: examining different regions of relevance. *Information Processing and Management* 1998, **34**, 599-622.
[11] Greisdorf, H., & Spink, A. Recent relevance research: implications for information professionals. *Online Information Review* 2000, **24**, 389-395.
[12] Vakkari, P., & Hakala, N. Changes in relevance criteria and problem stages in task performance. *Journal of Documentation* 2000, **56**, 540-562.
[13] Rieh, S.Y. Judgment of the information quality and cognitive authority of the web. *Journal of the American Society for Information Science and Technology* 2002, **53**, 145-161.
[14] McDonald, S., Taylor, L., & Adams, C. Searching the right database: a comparison of four databases for psychiatry journals. *Health Libraries Review* 1999, **16**, 151-156.
[15] Brettle, A.J., & Long, A.F. Comparison of bibliographic databases for information on the rehabilitation of people with severe mental illness. *Bulletin of the Medical Library Association* 2001, **89**, 353-362.
[16] Subirana, M., Sola, I., Garcia, J.M., Gich, I., & Urrutia, G. A nursing systematic review required MEDLINE and CINAHL for study identification. *Journal of Clinical Epidemiology* 2005, **58**, 20-25.
[17] Subirana, M., Sola, I., Garcia, J.M., et al. Importance of the database in the literature search: the first step in a systematic review. *Enfermeria Clinica* 2002, **12**, 296-300.
[18] Watson, M.M., & Perrin, R. A comparison of CINAHL and MEDLINE CD-ROM in four allied health areas. *Bulletin of the Medical Library Association* 1994, **82**, 214-216.
[19] Yonker, V.A., Young, K.P., Beecham, S.K., Horwitz, S., & Cousin, K. Coverage and overlaps in bibliographic databases relevant to forensic medicine: a comparative analysis of MEDLINE. *Bulletin of the Medical Library Association* 1990, **78**, 49-56.
[20] Suarez-Almazor, M.E., Belseck, E., Homik, J., Dorgan, M., & Ramos-Remus, C. Identifying clinical trials in the medical literature with electronic databases: MEDLINE alone is not enough. *Controlled Clinical Trials* 2000, **21**, 476-487.
[21] Watson, R.J.D., & Richardson, P.H. Accessing the literature on outcome studies in group psychotherapy: the sensitivity and precision of MEDLINE and PsycINFO bibliographic database searching. *British Journal of Medical Psychology* 1999, **72**, 127-134.
[22] Marson, A.G., & Chadwick, D.W. How easy are randomized controlled trials in epilepsy to find on MEDLINE? The sensitivity and precision of two MEDLINE searches. *Epilepsia* 1996, **37**, 377-380.
[23] Jenuwine, E.S., & Floyd, J.A. Comparison of medical subject headings and textword searches in MEDLINE to retrieve studies on sleep in healthy individuals. *Journal of the Medical Library Association* 2004, **92**, 349-535.
[24] Gehanno, J.-F. Assessment of bibliographic databases performance in information retrieval for occupational and environmental toxicology. *Occupational and Environmental Medicine* 1998, **55**, 562-566.
[25] Okuma, E. Selecting CD-ROM databases for nursing students: a comparison of MEDLINE and the Cumulative Index to Nursing and Allied Health Literature. *Bulletin of the Medical Library Association* 1994, **82**, 25-29.
[26] Brown, C.M. Complementary use of the SciSearch database for improved biomedical information searching. *Bulletin of the Medical Library Association* 1998, **86**, 63-67.

[27] Brown, C.M. The benefits of searching EMBASE versus MEDLINE for pharmaceutical information. *Online & CDROM Review* 1998, **22**, 3-8
[28] Brand-de-Heer, D.L. A comparison of the coverage of clinical medicine provided by PASCAL BIOMED and MEDLINE. *Health Information and Libraries Journal* 2000, **17**, 110-116.
[29] McCain, K.W., White, H.D., & Griffith, B.C. Comparing retrieval performance in online databases. *Information Processing and Management* 1987, **23**, 539-553.
[30] Brazier, H., & Begley, C.M. Selecting a database for literature searches in nursing: MEDLINE or CINAHL? *Journal of Advanced Nursing* 1996, **24**, 868-875.
[31] Burnham, J., & Shearer, B. Comparison of CINAHL, EMBASE and MEDLINE databases for the nurse researcher. *Medical Reference Services Quarterly* 1993, **12**, 45-57.
[32] Saracevic, T., Mokros, H., & Su, L. Nature of interaction between users and intermediaries in online searching: a qualitative analysis. In: Henderson, D., (ed.). *ASIS '90 Information in the year 2000: from research to applications Proceedings of the 53rd annual meeting of the American Society for Information Science*. Learned Information, Medford, New Jersey, 1990: pp47-54.
[33] Riegelman, R.K. *Studying a study and testing a test*. Philadelphia: Lippincott, Williams and Wilkins, 2005

| Odds ratio | Relevancy | |
|---|---|---|
| | Relevant | Not relevant |
| Database 1 hit | A | B |
| Database 1 miss | C | D |

Table 1: showing the contingency table to calculate whether a database is likely to retrieve relevant or irrelevant hits.

| Odds ratio | Uniqueness | |
|---|---|---|
| | Unique | Not unique |
| Database 1 hit | A | B |
| Database 1 miss | C | D |

Table 2: showing the contingency table to calculate whether a database is likely to retrieve relevant hits that are unique or relevant hits that are not unique.

| Odds ratio | Obtainability | |
|---|---|---|
| | Obtainable | Not obtainable |
| Database 1 hit | A | B |
| Database 1 miss | C | D |

Table 3: showing the contingency table to calculate whether a database is likely to retrieve relevant hits that are obtainable or relevant hits that are not obtainable.

| **Search1: Use of hypnosis in labour and pregnancy** | | | | | |
|---|---|---|---|---|---|
| Effectiveness | | | | | |
| Database | Hits | Relevant | Not relevant | Precision | Recall |
| BNI | 4 | 4 | 0 | 1.00 | 0.20 |
| CINAHL | 4 | 4 | 0 | 1.00 | 0.20 |
| EMBASE | 19 | 11 | 8 | 0.58 | 0.55 |
| MEDLINE | 17 | 11 | 6 | 0.65 | 0.55 |
| *TOTAL* | *30* | *20* | *10* | | |
| | | | | | |
| Efficiency | | | | | |
| Database | Relevant | Unique/R | Not unique | Novelty | Originality |
| BNI | 4 | 2 | 2 | 0.50 | 0.18 |
| CINAHL | 4 | 3 | 1 | 0.75 | 0.27 |
| EMBASE | 11 | 3 | 8 | 0.27 | 0.27 |
| MEDLINE | 11 | 3 | 8 | 0.27 | 0.27 |
| *TOTAL* | *20* | *11* | *9* | | |
| | | | | | |
| Accessibility | | | | | |
| Database | Relevant | Obtainable | Not obtainable | Availability | Retrievability |
| BNI | 4 | 3 | 1 | 0.75 | 0.60 |
| CINAHL | 4 | 1 | 3 | 0.25 | 0.20 |
| EMBASE | 11 | 3 | 8 | 0.27 | 0.60 |
| MEDLINE | 11 | 2 | 9 | 0.18 | 0.40 |
| *TOTAL* | *20* | *5* | *15* | | |
| | | | | | |
| **Search 2: Continence in pregnancy and the post-natal period** | | | | | |

| Effectiveness | | | | | |
|---|---|---|---|---|---|
| Database | Hits | Relevant | Not relevant | Precision | Recall |
| BNI | 9 | 9 | 0 | 1.00 | 0.26 |
| CINAHL | 13 | 13 | 0 | 1.00 | 0.38 |
| EMBASE | 37 | 19 | 18 | 0.51 | 0.56 |
| MEDLINE | 43 | 25 | 18 | 0.58 | 0.74 |
| *TOTAL* | *59* | *35* | *24* | | |
| | | | | | |
| Efficiency | | | | | |
| Database | Relevant | Unique/R | Not unique | Novelty | Originality |
| BNI | 9 | 0 | 9 | 0.00 | 0.00 |
| CINAHL | 13 | 3 | 10 | 0.23 | 0.23 |
| EMBASE | 19 | 3 | 16 | 0.16 | 0.23 |
| MEDLINE | 25 | 7 | 18 | 0.28 | 0.54 |
| *TOTAL* | *35* | *13* | *22* | | |
| | | | | | |
| Accessibility | | | | | |
| Database | Relevant | Obtainable | Not obtainable | Availability | Retrievability |
| BNI | 9 | 9 | 0 | 1.00 | 0.60 |
| CINAHL | 13 | 9 | 4 | 0.69 | 0.60 |
| EMBASE | 19 | 9 | 10 | 0.47 | 0.60 |
| MEDLINE | 25 | 9 | 16 | 0.36 | 0.60 |
| *TOTAL* | *35* | *15* | *20* | | |
| | | | | | |
| **Search 3: Thrombolytic therapies in the myocardial infarction patient** | | | | | |
| Effectiveness | | | | | |
| Database | Hits | Relevant | Not relevant | Precision | Recall |
| BNI | 2 | 1 | 1 | 0.50 | 0.01 |
| CINAHL | 22 | 19 | 3 | 0.86 | 0.23 |
| EMBASE | 78 | 56 | 22 | 0.72 | 0.69 |
| MEDLINE | 86 | 59 | 17 | 0.69 | 0.73 |
| *TOTAL* | *113* | *81* | *32* | | |
| | | | | | |
| Efficiency | | | | | |
| Database | Relevant | Unique/R | Not unique | Novelty | Originality |
| BNI | 1 | 0 | 1 | 0.00 | 0.00 |
| CINAHL | 19 | 6 | 13 | 0.32 | 0.17 |
| EMBASE | 56 | 16 | 40 | 0.29 | 0.44 |
| MEDLINE | 59 | 15 | 44 | 0.25 | 0.42 |
| *TOTAL* | *81* | *36* | *45* | | |
| | | | | | |
| Accessibility | | | | | |
| Database | Relevant | Obtainable | Not obtainable | Availability | Retrievability |
| BNI | 1 | 1 | 0 | 1.00 | 0.09 |
| CINAHL | 19 | 7 | 12 | 0.37 | 0.64 |
| EMBASE | 56 | 9 | 47 | 0.16 | 0.82 |
| MEDLINE | 59 | 9 | 50 | 0.15 | 0.82 |

| | | | | | |
|---|---|---|---|---|---|
| *TOTAL* | *81* | *11* | *70* | | |
| | | | | | |
| **Search 4: Management of venous leg ulcers by district nurses** | | | | | |
| Effectiveness | | | | | |
| Database | Hits | Relevant | Not relevant | Precision | Recall |
| BNI | 6 | 5 | 1 | 0.83 | 0.71 |
| CINAHL | 7 | 5 | 2 | 0.71 | 0.71 |
| EMBASE | 3 | 1 | 2 | 0.33 | 0.14 |
| MEDLINE | 7 | 5 | 2 | 0.71 | 0.71 |
| *TOTAL* | *12* | *7* | *5* | | |
| | | | | | |
| Efficiency | | | | | |
| Database | Relevant | Unique/R | Not unique | Novelty | Originality |
| BNI | 5 | 1 | 4 | 0.20 | 1.00 |
| CINAHL | 5 | 0 | 5 | 0.00 | 0.00 |
| EMBASE | 1 | 0 | 1 | 0.00 | 0.00 |
| MEDLINE | 5 | 0 | 5 | 0.00 | 0.00 |
| *TOTAL* | *7* | *1* | *6* | | |
| | | | | | |
| Accessibility | | | | | |
| Database | Relevant | Obtainable | Not obtainable | Availability | Retrievability |
| BNI | 5 | 4 | 1 | 0.80 | 1.00 |
| CINAHL | 5 | 4 | 1 | 0.80 | 1.00 |
| EMBASE | 1 | 0 | 1 | 0.00 | 0.00 |
| MEDLINE | 5 | 3 | 2 | 0.60 | 0.75 |
| *TOTAL* | *7* | *4* | *3* | | |
| | | | | | |
| **Search 5: The role of the nurse practitioner in the orthopaedic setting** | | | | | |
| Effectiveness | | | | | |
| Database | Hits | Relevant | Not relevant | Precision | Recall |
| BNI | 9 | 1 | 8 | 0.11 | 0.17 |
| CINAHL | 20 | 4 | 16 | 0.20 | 0.67 |
| EMBASE | 6 | 1 | 5 | 0.17 | 0.17 |
| MEDLINE | 16 | 4 | 12 | 0.25 | 0.67 |
| *TOTAL* | *25* | *6* | *19* | | |
| | | | | | |
| Efficiency | | | | | |
| Database | Relevant | Unique/R | Not unique | Novelty | Originality |
| BNI | 1 | 0 | 1 | 0.00 | 0.00 |
| CINAHL | 4 | 1 | 3 | 0.25 | 0.33 |
| EMBASE | 1 | 0 | 1 | 0.00 | 0.00 |
| MEDLINE | 4 | 2 | 2 | 0.50 | 0.67 |
| *TOTAL* | *6* | *3* | *3* | | |
| | | | | | |
| Accessibility | | | | | |
| Database | Relevant | Obtainable | Not obtainable | Availability | Retrievability |
| BNI | 1 | 1 | 0 | 1.00 | 0.20 |

| | | | | | |
|---|---|---|---|---|---|
| CINAHL | 4 | 3 | 1 | 0.75 | 0.60 |
| EMBASE | 1 | 0 | 1 | 0.00 | 0.00 |
| MEDLINE | 4 | 3 | 1 | 0.75 | 0.60 |
| *TOTAL* | *6* | *5* | *1* | | |
| | | | | | |
| **Search 6: The image of breastfeeding** | | | | | |
| Effectiveness | | | | | |
| Database | Hits | Relevant | Not relevant | Precision | Recall |
| BNI | 2 | 2 | 0 | 1.00 | 0.29 |
| CINAHL | 7 | 5 | 2 | 0.71 | 0.71 |
| EMBASE | 10 | 2 | 8 | 0.20 | 0.29 |
| MEDLINE | 15 | 4 | 11 | 0.27 | 0.57 |
| *TOTAL* | *22* | *7* | *15* | | |
| | | | | | |
| Efficiency | | | | | |
| Database | Relevant | Unique/R | Not unique | Novelty | Originality |
| BNI | 2 | 0 | 2 | 0.00 | 0.00 |
| CINAHL | 5 | 3 | 2 | 0.60 | 0.75 |
| EMBASE | 2 | 0 | 2 | 0.00 | 0.00 |
| MEDLINE | 4 | 1 | 3 | 0.25 | 0.25 |
| *TOTAL* | *7* | *4* | *3* | | |
| | | | | | |
| Accessibility | | | | | |
| Database | Relevant | Obtainable | Not obtainable | Availability | Retrievability |
| BNI | 2 | 2 | 0 | 1.00 | 0.67 |
| CINAHL | 5 | 3 | 2 | 0.60 | 1.00 |
| EMBASE | 2 | 1 | 1 | 0.50 | 0.33 |
| MEDLINE | 4 | 2 | 2 | 0.50 | 0.67 |
| *TOTAL* | *7* | *3* | *4* | | |
| | | | | | |
| **Search 7: Psychosocial support for the laryngectomy patient** | | | | | |
| Effectiveness | | | | | |
| Database | Hits | Relevant | Not relevant | Precision | Recall |
| BNI | 2 | 1 | 1 | 0.50 | 0.05 |
| CINAHL | 5 | 2 | 3 | 0.40 | 0.11 |
| EMBASE | 25 | 10 | 15 | 0.40 | 0.53 |
| MEDLINE | 26 | 13 | 13 | 0.50 | 0.68 |
| *TOTAL* | *44* | *19* | *25* | | |
| | | | | | |
| Efficiency | | | | | |
| Database | Relevant | Unique/R | Not unique | Novelty | Originality |
| BNI | 1 | 1 | 0 | 1.00 | 0.08 |
| CINAHL | 2 | 1 | 1 | 0.50 | 0.08 |
| EMBASE | 10 | 3 | 7 | 0.30 | 0.25 |
| MEDLINE | 13 | 7 | 6 | 0.54 | 0.58 |
| *TOTAL* | *19* | *12* | *7* | | |

| Accessibility | | | | | |
|---|---|---|---|---|---|
| Database | Relevant | Obtainable | Not obtainable | Availability | Retrievability |
| BNI | 1 | 0 | 1 | 0.00 | 0.00 |
| CINAHL | 2 | 0 | 2 | 0.00 | 0.00 |
| EMBASE | 10 | 1 | 9 | 0.10 | 1.00 |
| MEDLINE | 13 | 1 | 12 | 0.08 | 1.00 |
| *TOTAL* | *19* | *1* | *18* | | |
| | | | | | |
| **Search 8: Psychosocial aspects of postnatal depression** | | | | | |
| Effectiveness | | | | | |
| Database | Hits | Relevant | Not relevant | Precision | Recall |
| BNI | 6 | 5 | 1 | 0.83 | 0.26 |
| CINAHL | 6 | 5 | 1 | 0.83 | 0.26 |
| EMBASE | 23 | 13 | 10 | 0.57 | 0.68 |
| MEDLINE | 10 | 4 | 6 | 0.40 | 0.21 |
| *TOTAL* | *30* | *19* | *11* | | |
| | | | | | |
| Efficiency | | | | | |
| Database | Relevant | Unique/R | Not unique | Novelty | Originality |
| BNI | 5 | 0 | 5 | 0.00 | 0.00 |
| CINAHL | 5 | 3 | 2 | 0.60 | 0.27 |
| EMBASE | 13 | 7 | 6 | 0.54 | 0.64 |
| MEDLINE | 4 | 1 | 3 | 0.25 | 0.09 |
| *TOTAL* | *19* | *11* | *8* | | |
| | | | | | |
| Accessibility | | | | | |
| Database | Relevant | Obtainable | Not obtainable | Availability | Retrievability |
| BNI | 5 | 4 | 1 | 0.80 | 0.40 |
| CINAHL | 5 | 5 | 0 | 1.00 | 0.50 |
| EMBASE | 13 | 4 | 9 | 0.31 | 0.40 |
| MEDLINE | 4 | 2 | 2 | 0.50 | 0.20 |
| *TOTAL* | *19* | *10* | *9* | | |
| | | | | | |
| **Search 9: The expert patient** | | | | | |
| Effectiveness | | | | | |
| Database | Hits | Relevant | Not relevant | Precision | Recall |
| BNI | 16 | 10 | 6 | 0.63 | 0.38 |
| CINAHL | 31 | 15 | 16 | 0.48 | 0.58 |
| EMBASE | 22 | 8 | 14 | 0.36 | 0.31 |
| MEDLINE | 19 | 12 | 7 | 0.63 | 0.46 |
| *TOTAL* | *51* | *26* | *25* | | |
| | | | | | |
| Efficiency | | | | | |
| Database | Relevant | Unique/R | Not unique | Novelty | Originality |
| BNI | 10 | 1 | 9 | 0.10 | 0.07 |
| CINAHL | 15 | 6 | 9 | 0.40 | 0.43 |
| EMBASE | 8 | 4 | 4 | 0.50 | 0.29 |

| | | | | | |
|---|---|---|---|---|---|
| MEDLINE | 12 | 3 | 9 | 0.25 | 0.21 |
| *TOTAL* | *26* | *14* | *12* | | |
| | | | | | |
| Accessibility | | | | | |
| Database | Relevant | Obtainable | Not obtainable | Availability | Retrievability |
| BNI | 10 | 8 | 2 | 0.80 | 0.40 |
| CINAHL | 15 | 13 | 2 | 0.87 | 0.65 |
| EMBASE | 8 | 6 | 2 | 0.75 | 0.30 |
| MEDLINE | 12 | 10 | 2 | 0.83 | 0.50 |
| *TOTAL* | *26* | *20* | *6* | | |

Table 4: showing the results of each search

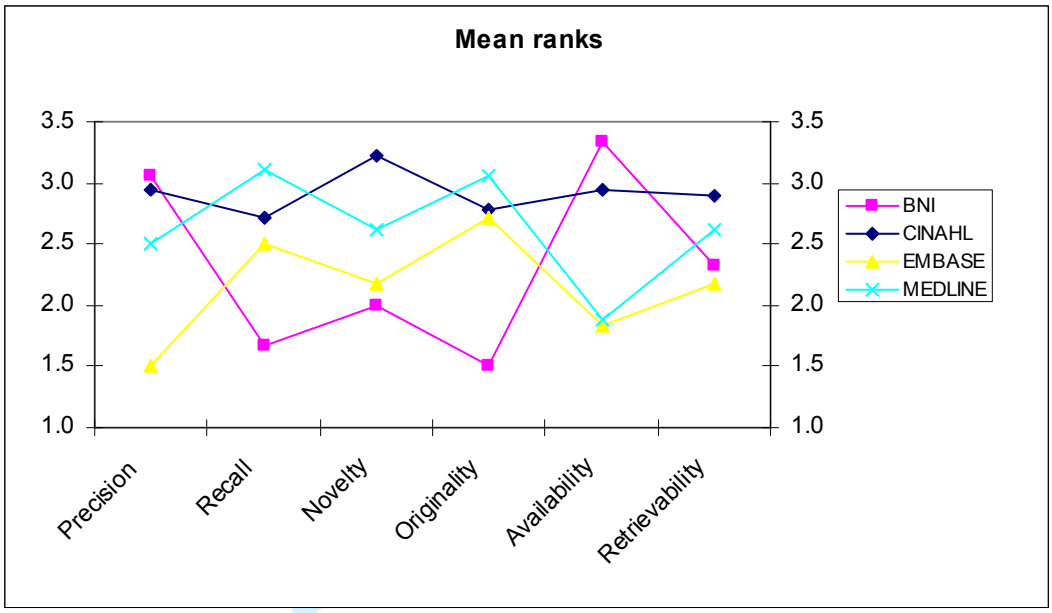
| **Criteria** | **BNI** | **CINAHL** | **EMBASE** | **MEDLINE** | $X^2$ | df | ***p>0.05*** |
|---|---|---|---|---|---|---|---|
| Precision | 3.06 | 2.94 | 1.50 | 2.50 | 8.819 | 3 | ***0.032*** |
| Recall | 1.67 | 2.72 | 2.50 | 3.11 | 6.788 | 3 | 0.079 |
| Novelty | 2.00 | 3.22 | 2.17 | 2.61 | 5.241 | 3 | 0.155 |
| Originality | 1.50 | 2.78 | 2.67 | 3.06 | 8.922 | 3 | ***0.030*** |
| Availability | 3.33 | 2.94 | 1.83 | 1.89 | 9.663 | 3 | ***0.022*** |
| Retrievability | 2.33 | 2.89 | 2.17 | 2.61 | 2.042 | 3 | 0.564 |

Table 5: showing the average ranks for each database, chi-square ($X^2$), Degrees of Freedom (df), and the level of significance (*p*).

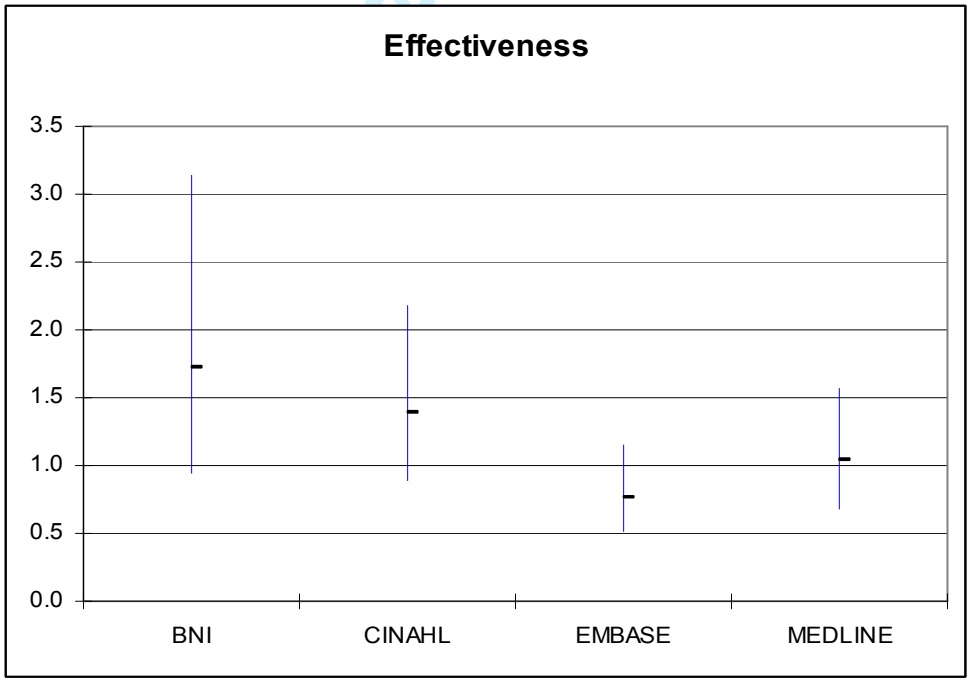
| **Database Hits** | **Relevant (a)** | **Not Relevant (b)** | **Unique/R (a)** | **Not Unique/R (b)** | **Obtain/R (a)** | **Not Obtain/R (b)** |
|---|---|---|---|---|---|---|
| BNI | 38 | 18 | 5 | 33 | 32 | 6 |
| CINAHL | 72 | 43 | 26 | 46 | 45 | 27 |
| EMBASE | 121 | 102 | 36 | 85 | 33 | 88 |
| MEDLINE | 137 | 102 | 39 | 98 | 41 | 96 |
| *Total* | *220* | *166* | *106* | *114* | *74* | *146* |
| | | | | | | |
| **Database Misses** | **Relevant (c)** | **Not Relevant (d)** | **Unique/R (c)** | **Not Unique/R (d)** | **Obtain/R (c)** | **Not Obtain/R (d)** |
| BNI | 182 | 148 | 101 | 81 | 42 | 140 |
| CINAHL | 148 | 123 | 80 | 68 | 29 | 119 |
| EMBASE | 99 | 64 | 70 | 29 | 41 | 58 |
| MEDLINE | 83 | 64 | 67 | 16 | 33 | 50 |

*Note: Totals (other than 'Unique/R') do not equal the number of hits due to more than a single database retrieving the same hit.*

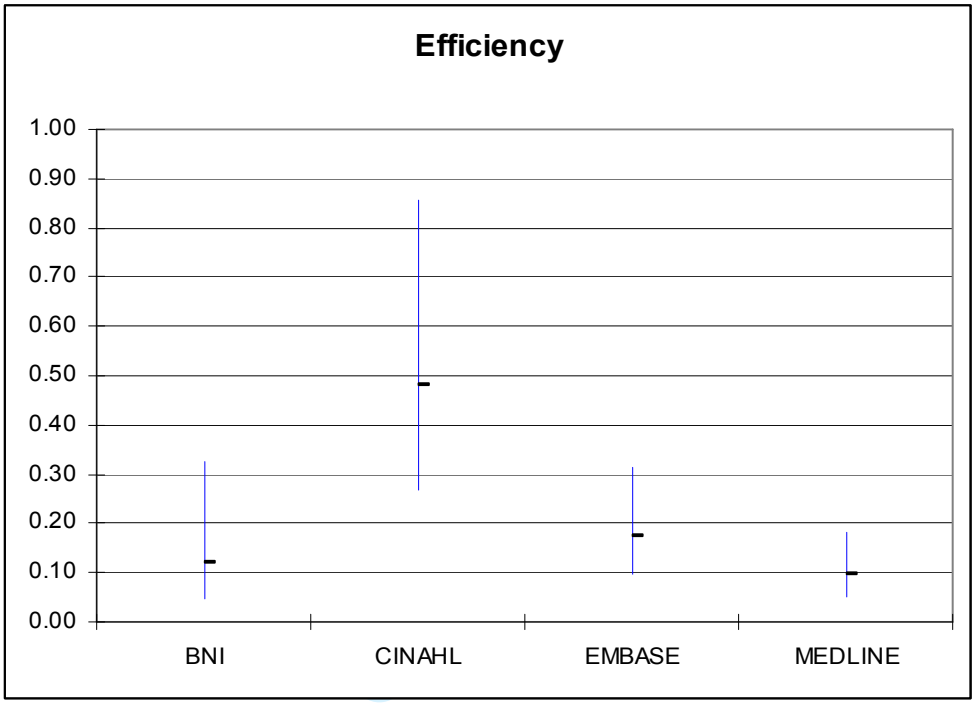
Table 6: showing the number of hits and misses for each database. (Unique/R = unique and relevant; Not Unique/R = not unique, but still a relevant hit), (Obtain/R = obtainable and relevant; Not Obtain/R = not obtainable, but still a relevant hit).



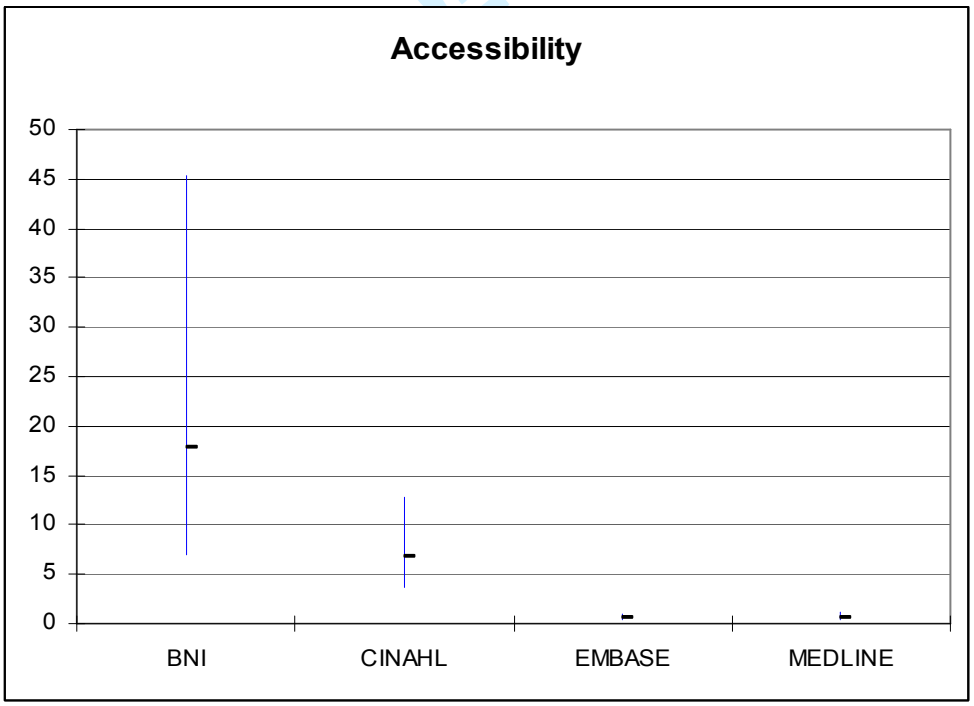

Graph 1: showing the mean ranks for six separate criteria for each database




Graph 2: showing the 'effectiveness' odds for each database and confidence intervals




Graph 3: showing the 'efficiency' odds for each database and confidence intervals




Graph 4: showing the 'efficiency' odds for each database and confidence intervals