

SUBJECT CONTENT ANALYSIS OF HIV/AIDS RESEARCH IN EASTERN AND SOUTHERN AFRICA, 1980–2005¹

OMWOYO BOSIRE ONYANCHA

Department of Information Science
University of South Africa
onyanob@unisa.ac.za

DENNIS N. OCHOLLA

Department of Information Studies
University of Zululand
docholla@pan.uzulu.ac.za

ABSTRACT

This paper provides a content analysis of HIV/AIDS research in Eastern and Southern Africa between 1980 and 2005 as indexed in the MEDLINE database, with a view to assisting in the identification of HIV/AIDS indexing terms which can be used to access HIV/AIDS literature. Data were analyzed using Bibexcel software in order to establish the trends in HIV/AIDS research in the two regions. Results indicate that the number of keywords/terms used to index HIV/AIDS research outputs has grown exponentially, thus providing a number of options for accessing HIV/AIDS research findings. The ranking of main subject headings was initially unstable, especially in the 1980s, but had stabilized by the mid-1990s and thereafter. Concerning the sub-fields of HIV/AIDS, it was noted that most research is conducted on epidemiology, prevention and control, transmission, complications, and drug therapy. Drug therapy and antiretrovirals are rapidly emerging as the main areas of HIV/AIDS research, implying that research has shifted from the causal factors and diagnosis (which were the major areas of concern in the 1990s) to the care of those with HIV. Other findings, discussions, conclusions, and recommendations are provided. Areas for further research are also identified.

¹ This paper has been extracted from OB Onyancha's PhD thesis, completed in 2007 at the University of Zululand, South Africa.

KEYWORDS

AIDS, HIV, informetrics, content analysis, Eastern Africa, Southern Africa, sub-Saharan Africa

1 INTRODUCTION

Acquired immunodeficiency syndrome (AIDS), commonly known as HIV/AIDS (HIV being the abbreviation for human immunodeficiency virus), remains one of the most pressing concerns for developing countries, even more so in sub-Saharan Africa, which is home to about 25 million infected persons. Of these, 17.4 million (69.6%) live in Eastern and Southern Africa: Eastern Africa accounts for 5.8 million (23.2%) of sub-Saharan cases, while Southern Africa is home to 11.6 million (46.2%) of sub-Saharan cases (United Nations Programme on HIV/AIDS (UNAIDS) 2003). According to UNAIDS (2003), HIV prevalence has remained alarmingly high in the general population across Eastern and Southern Africa. This organization estimates that about 30% of people living with HIV/AIDS worldwide live in Southern Africa, yet this region is home to less than 2% of the world's population. Consequently, the crisis calls for novel alliances between social and biological sciences, especially with regard to designing effective interventions to prevent or treat the complications arising from HIV transmission (Farmer 1999). Apart from epidemiologists, physicians, scientists and policy makers, other professionals (e.g. social workers, information scientists, educationists, statisticians, economists and counselors) are increasingly becoming involved in solving the myriad problems associated with the pandemic (Onyancha 2006).

An area that in our view requires monitoring in order to encourage informed decision making regarding funding in various focus areas of HIV/AIDS research as well as subject access is the content of HIV/AIDS research. We concur with Cohen's (2000) observation that the acquisition of knowledge concerning the manifestations and features of HIV/AIDS in Africa would lead to the proper development of appropriate strategies for AIDS patients. These would include, for example, strategies to promote health and prolong lives by supplying AIDS patients with the appropriate anti-retroviral drugs. Knowledge of the shifts in HIV/AIDS research is essential when designing intervention programs. Emerging areas of research tend to mirror new problem areas in a subject domain, and therefore different approaches or an improvement in the methods used in dealing with new phenomena would be required. Concerning subject access, Macias-Chapula, Sotolongo-Aguilar, Magde & Solorio-Lagunas (1999:565) argue that subject content analysis of AIDS literature would mirror "not only the construction of this field by specific institutions and countries, but also what happens to subject access as the knowledge base and environment of a discipline grow and change". Bierbaum and Brooks (1995) also acknowledge that in order to make maximum usage of a database or have comprehensive access to AIDS literature, it is necessary to be knowledgeable about

the new terms and phrases used to index the literature. They observe that: “In order to be assured of making a comprehensive search of the medical literature for AIDS-related citations, one would have to have a rather thorough and up-to-date knowledge of the terminology of the field” (Bierbaum & Brooks 1995:536).

A content analysis can also make it possible to identify what Bierbaum and Brooks (1995:533) term the “rising and falling frequency” of the occurrence of subject headings. These authors observe that such data can be used to “infer the changing level of interest in a particular aspect of AIDS research” and to “track the introduction of new terms that reflect innovations and discoveries in the knowledge base” (Bierbaum & Brooks 1995:533). Kizito (2002) notes that content analysis can be used to “find out whether or not writers, producers, media organizations are biased against a specific subject of reporting such as HIV/AIDS. They use it to find out what prominence is given to a specific area of reporting”.

As HIV/AIDS in Africa is said to be distinct or unique (see Cohen 2000; Onyancha & Ocholla 2009), there is a need to identify the terms used in the literature disseminating HIV/AIDS research for the purposes of subject access in the region. This paper attempts to answer questions such as: What are the major keywords used to describe HIV/AIDS in Eastern and Southern Africa? What are the terms that can be used to access the HIV/AIDS literature in the region? What are the emerging terms that are used to index HIV/AIDS literature in the region? and Which sub-fields of HIV/AIDS research are the most researched in the region?

2 PURPOSE AND OBJECTIVES OF THE STUDY

This paper presents the preliminary findings of a content analysis of HIV/AIDS research in Eastern and Southern Africa between 1980 and 2005 as indexed in the MEDLINE database. Specifically, the objectives of the study were:

- to examine the growth in the number of HIV/AIDS subject indexing terms (main MESH terms) between 1980 and 2005
- to identify the MESH terms² most commonly used in indexing HIV/AIDS literature
- to offer a trend analysis of research into various aspects of HIV/AIDS
- to assess publishing activity in the sub-fields of HIV/AIDS research in order to present a clear picture of the endeavors in those sub-fields

2 In this paper, the MESH subject headings are referred to as MESH terms, descriptors, topics or simply terms.

3 LITERATURE REVIEW

A literature review reveals that studies conducted to analyze publishing trends associated with and patterns of HIV/AIDS research have been general in nature. For instance, although Onyanacha and Ocholla (2006) sought to examine HIV/AIDS literature specific to young persons, they nevertheless referred to the subject content of HIV/AIDS literature specific to the youths. In another study, Onyanacha and Ocholla (2004) aimed to compare HIV/AIDS research on/in Kenya and Uganda in general terms. However, one of the variables analyzed in the latter study was publishing activity in various sub-fields of HIV/AIDS. Similarly, Macias-Chapula & Mijangos-Nolasco (2002) conducted a bibliometric study on AIDS literature in Central Africa in which they also considered topical HIV/AIDS issues in the papers published. A review of Macias-Chapula's (2000) study on AIDS in Haiti reveals approaches similar to those adopted in the aforementioned studies.

That notwithstanding, these studies and others have identified focus areas in HIV/AIDS research, especially at an international level. Patterns of literature production indicate that when ranked, the main subject headings include HIV infections, acquired immunodeficiency syndrome, sex education and sex behavior (Bierbaum & Brooks 1995; Macias-Chapula 2000; Macias-Chapula & Mijangos-Nolasco 2002; Macias-Chapula et al 1999; Onyanacha & Ocholla 2004; Onyanacha & Ocholla 2006).

Bierbaum and Brooks (1995) offer an in-depth analysis of the subject content of literature on AIDS at an international level. Using the AIDSLINE database to collect data, the authors analyzed a total of 12,987 HIV/AIDS records published in 1992 and compared the 30 top-ranked subject headings in 1992 with those of 1989. The 1992 rankings showed some stability, particularly as regards the first 8 subjects, with descriptors that included human, HIV infections, acquired immunodeficiency syndrome, male, female, adult, HIV seropositivity and support, non-U.S. govt. There were, however, shifts in subsequent group ranks. Terms that did not feature at all among the 30 top-ranked terms in 1989 emerged in the 1992 list. These comprised HIV infections, HIV seropositivity, comparative study, sex behavior, molecular sequence data, opportunistic infections, substance abuse and intravenous. Others were leukocyte count, base sequence, T-4 lymphocytes, pregnancy, Zidovudine, amino acid sequence, cell line, polymerase chain reaction, and follow-up studies. Some of the indexing terms, derived from the medical subject headings (MESH) thesaurus of the National Library of Medicine, give an indication of the type of research, geographic region, type of study, and form of literature. In the analysis of the records by the sub-heading terms, /complications, /epidemiology, /prevention and control, and /transmission topped the list of the 10 top-ranked terms in 1992. The sub-headings /epidemiology, /prevention and control, and /treatment ranked highly for both AIDS and HIV infections. Slight shifts in rank position occurred in a few of the main subject headings and sub-headings during the two decades of study.

The rest of the studies (Macias-Chapula 2000; Macias-Chapula & Mijangos-Nolasco 2002; Onyancha & Ocholla 2004; Onyancha & Ocholla 2006) support the findings of Bierbaum and Brooks (1995) and Macias-Chapula et al (1999). Their surmised observation is that HIV infections and acquired immunodeficiency syndrome are dominant main subject headings, while epidemiology, prevention and control and transmission are highly regarded areas of study within HIV/AIDS research. Further research to explore the trends – through trend analysis – in research on the main subject headings and sub-headings of HIV/AIDS in Africa over a longer period of time may provide a clearer picture of how to support decision-making processes in the region.

4 METHODS AND PROCEDURES

Generally, the content analysis method was used to evaluate publishing activity associated with various topics of HIV/AIDS research. Content analysis is a research technique used to determine the presence of words or concepts in collections of textual documents. It is defined as a “research technique for the objective, systematic, and quantitative description of manifest content of communications” (Berelson in Palmquist [no date]). Palmquist [no date] observes that content analysis is used to “determine the presence of certain words, concepts, themes, phrases, characters, or sentences within texts or sets of texts and to quantify this presence in an objective manner”. He defines texts as books, book chapters, essays, interviews, discussions, newspaper headlines and articles, historical documents, speeches, conversations, advertising, theater, informal conversation, or effectively any occurrence of communicative language. The uses of content analysis are many and varied. The method may be used to reveal international differences in communication content; detect the existence of propaganda; identify the intentions, focus or communication trends of an individual, group or institution; describe attitudinal and behavioral responses to communications; or determine the psychological or emotional state of persons or groups.

The geographical region covered in this study consists of all countries in Eastern and Southern Africa (see Appendix A). Of the regions on the continent of Africa, these, as mentioned in the introduction, are the worst affected by HIV/AIDS. The study was limited to the first 25 years of the existence of HIV/AIDS. The paper therefore reports preliminary findings spanning these 25 years as efforts are made to expand the period of study up to 2010. Data were extracted from the MEDLINE database using an advanced search mode whereby 26 HIV/AIDS-specific terms were combined with the names of countries/geographic regions in Eastern and Southern Africa (see Appendices A and B). In order to extract relevant data, two types of records were downloaded:

1. Records published by authors whose institutional affiliation was African (i.e. the country of origin of the publication was in sub-Saharan Africa). The search query was in the format: AF Kenya.

2. Records about HIV/AIDS published in sub-Saharan African countries by both local and international authors. In this regard, the search was not limited to any institutional affiliation. The search was conducted within the Subjects field. For example, publications describing HIV/AIDS in Kenya were extracted from the database using the search query SU Kenya.

As suggested by Bierbaum, Brooks and Brooks (1992:10), the precaution was taken of including the terms by which HIV/AIDS was known at the beginning of the epidemic (during the early 1980s). Several other sources (e.g. literature indexed by MEDLINE on bibliometric studies of HIV/AIDS literature (see the literature review section)) were used to obtain the list of HIV/AIDS-specific terms. The MEDLINE database provides an elaborative classification of the topics indexed in it, a fact that largely dictated the choice of the database on which to conduct a content analysis of HIV/AIDS literature for this study. In addition, the database includes citations from the *Index medicus*, *International nursing index*, *Index to dental literature*, *PREMEDLINE*, *AIDSLINE*, *BIOETHICSLINE*, and *HealthSTAR*.

Two approaches were used to study the content of HIV/AIDS literature in Eastern and Southern Africa:

1. An examination of the MESH sub-headings relating to HIV and AIDS in order to distinctly clarify the efforts made in the various sub-fields of HIV/AIDS research in Eastern and Southern Africa.
2. A trend analysis of MESH terms in order to determine the most researched topics in HIV/AIDS research in Eastern and Southern Africa.

Bibexcel software was used to obtain frequencies of the main MESH terms and sub-topics of HIV/AIDS literature. The entire period of the study was divided into nine three-year periods, except for 2004–2005, and thirty top-ranking MESH terms were identified for each period: this made it easy to monitor any shifts in research. The approach that was used by Macias-Chapula et al (1999) to obtain the final MESH term frequencies was also adopted. Each subject heading was treated as distinct. That is, if a subject heading appeared twice or more in different formats, such as: *HIV*/Infections*, *HIV/Infections* or *HIV/*Infections*, they were treated as one descriptor, while being counted as three entries for HIV.

Data generated from this analysis were presented in tables. Terms in each period were compared with those in subsequent or previous periods in order to determine:

1. subject/topical terms that might have disappeared;
2. subject/topical terms that were ranked lower; and
3. emerging subject/topical terms.

5 RESULTS AND DISCUSSION

This section presents the findings under three broad sub-headings, namely:

- Growth in the number of MESH terms, 1980–2005
- Trends in main MESH terms relating to HIV/AIDS, 1980–2005
- Trends in MESH sub-topics relating to HIV/AIDS, 1980–2005

5.1 GROWTH IN THE NUMBER OF MESH TERMS, 1980–2005

Table 1 provides a trend analysis of the number of MESH main terms from 1980 to 2005. It was observed that the number of terms rose from just 127 at the beginning of the epidemic (1980–1982) to 25,524 in 2001–2003, which translates to a 19,998% increase.

Table 1: Growth in the number of MESH main terms, 1980–2005

Year	No. of terms	Change in no. of terms	Change in %	Records	Terms per record
2004–2005	17124	-8400	-33	1039	16.48
2001–2003	25524	5888	30	1603	15.92
1998–2000	19636	3497	22	1256	15.63
1995–1997	16139	6242	63	998	16.17
1992–1994	9897	3263	49	620	15.96
1989–1991	6634	4544	217	464	14.30
1986–1988	2090	1637	361	152	13.75
1983–1985	453	326	257	31	14.61
1980–1982	127	-	-	9	14.11

An inverse relationship between the number of MESH main terms and the number of terms per record was observed. Whereas the number of MESH main terms showed continued (exponential) growth, there was a consistent pattern in the average number of terms per record. The number of MESH terms per record ranged between 13.75 (in 1986–1988) and 16.48 (in 2004–2005). The average number of terms per year was therefore almost constant, implying a positive correlation between the number of terms and records for each period.

5.2 MAIN MESH TOPICS RELATING TO HIV/AIDS, 1980–2005

A trend analysis was performed between 1980 and 2005 to monitor the emergence and disappearance of certain topics and to determine the most researched topic throughout the period of study, as shown below.

5.2.1 1980–1982

Table 2 provides the top-ranking topics of research at the beginning of the HIV/AIDS epidemic, 1980 to 1982. At the top of the table is Burkitt lymphoma with 9 records, followed by antibodies, bacterial (7), antibodies, viral (5), and herpesvirus 4, human (5). Others are age factors (2), altitude (2), *Escherichia coli* (2), hepatitis, viral, human (2), immunoglobulin A, secretory (2), liver diseases (2), milk, human (2), parainfluenza virus 3, human (2), pregnancy (2) and antigens, viral (2). One record was registered for each of the other topics.

Table 2: Distribution of records by the main MESH terms, 1980–1982

Rank	MESH term	No. of hits	Rank	MESH term	No. of hits
1	Burkitt Lymphoma	9	6	Adenovirus Infections, Human	1
2	Antibodies, Bacterial	7	6	Adenoviruses, Human	1
3	Antibodies, Viral	5	6	Adolescence	1
4	Herpesvirus 4, Human	5	6	Antibody Specificity	1
5	Age Factors	2	6	Carcinoma, Hepatocellular	1
5	Altitude	2	6	Colostrum	1
5	<i>Escherichia coli</i>	2	6	DNA, Viral	1
5	Hepatitis, Viral, Human	2	6	Enterotoxins	1
5	Immunoglobulin A, Secretory	2	6	Hepatitis B Core Antigens	1
5	Liver Diseases	2	6	Hepatitis B Surface Antigens	1
5	Milk, Human	2	6	Hepatitis B virus	1
5	Parainfluenza Virus 3, Human	2	6	Hepatitis C	1
5	Pregnancy	2	6	Hepatovirus	1
5	Antigens, Viral	2	6	Immunoglobulin M	1
6	Adenoviridae Infections	1	6	Lactation	1

5.2.2 1983–1985

The period from 1983 to 1985 saw the emergence of a descriptor specific to the subject of HIV/AIDS, namely acquired immunodeficiency syndrome, as a subject indexing term. Acquired immunodeficiency syndrome tops the list of 30 top-ranking topics with a total of 29 records, while Kaposi sarcoma is in second position with 22 records, followed by retroviridae infections (15), adolescence (15), antibodies, viral (14), and deltaretrovirus (13). Burkitt lymphoma, which occupied first place in the previous table, is here rated sixth.

Likewise, herpesvirus 4, human has moved from position 4 in table 2 to position 7 in table 3. Besides acquired immunodeficiency syndrome, new entrants in the list of the 30 top-ranking topics include sarcoma, Kaposi; retroviridae infections; deltaretrovirus; homosexuality; and nasopharyngeal neoplasms. There were a total of 22, out of 30, new descriptors in the top-ranking terms in table 3. An equal number that featured in the previous period (1980–1982) did not appear in 1983–1985. Only 8 topics ranked highly in both periods (tables 2 and 3); these were adolescence, which moved up from sixth to third position; antibodies, viral; Burkitt lymphoma; herpesvirus 4, human; antigens, viral; age factors; immunoglobulin A; and hepatitis, viral, human.

Table 3: Distribution of records by the main MESH terms, 1983–1985

Rank		No. of hits	Rank		No. of hits
1	Acquired Immunodeficiency Syndrome	29	10	Enzyme-Linked Immunosorbent Assay	3
2	Sarcoma, Kaposi	22	10	Disease Outbreaks	3
3	Retroviridae Infections	15	10	Diarrhea	3
3	Adolescence	15	10	Body Weight	3
4	Antibodies, Viral	14	10	Antigens, Viral	3
5	Deltaretrovirus	13	10	Age Factors	3
6	Burkitt Lymphoma	12	11	Syndrome	2
7	Herpesvirus 4, Human	8	11	Neoplasms, Multiple Primary	2
8	Homosexuality	5	11	Lymphopenia	2
9	Nasopharyngeal Neoplasms	4	11	Immunoglobulin A	2
9	Diagnosis, Differential	4	11	Hepatitis, Viral, Human	2
10	Sex Factors	3	11	Epidemiologic Methods	2
10	Risk	3	11	Entamoebiasis	2
10	Malaria	3	11	Cytomegalovirus Infections	2
10	Lymphoma	3	11	Chlamydia Infections	2

5.2.3 1986–1988

This period too saw a number of HIV/AIDS-specific subject descriptors, namely HIV, HIV antibodies, HIV infections, and HIV seropositivity, added to the MESH thesaurus, as shown in table 4. These descriptors are ranked sixth, fifth, seventh and second behind acquired immunodeficiency syndrome, which tops the list, antibodies, viral in position three and adolescence in position four. Only one-fifth of the 30 top-ranked descriptors in table 3 appeared in table 2. Most (24/30 or 80%) of the descriptors had been relegated to the periphery or had disappeared altogether. There were 24 new descriptors that were introduced into the MESH thesaurus or whose ranking improved.

These include the aforementioned HIV/AIDS-specific descriptors; pregnancy; health education; organization and administration; sexually transmitted diseases; disease; and communication. Others that thereafter maintained a continued presence were: risk factors; prostitution; and sexual behavior.

Table 4: Rank distribution of MESH terms by the total number of records, 1986–1988

Rank		No. of hits	Rank		No. of hits
1	Acquired Immunodeficiency Syndrome	172	14	Research	12
2	HIV Seropositivity	38	15	Mass Screening	11
3	Antibodies, Viral	36	15	Enzyme-Linked Immunosorbent Assay	11
4	Adolescence	34	16	Risk Factors	10
5	HIV Antibodies	32	16	Prostitution	10
6	HIV	30	16	Health Services	10
7	Virus Diseases	28	16	Health Planning	10
7	HIV Infections	28	16	Delivery of Health Care	10
8	Pregnancy	20	17	Malaria	9
9	Health Education	18	18	Sexual Behavior	8
10	Organization and Administration	16	18	Population Characteristics	8
11	Sexually Transmitted Diseases	15	18	Information Services	8
12	Sarcoma, Kaposi	14	18	Infection	8
12	Disease	14	18	AIDS-Related Complex	8
13	Communication	13	19	Education	7

5.2.4 1989–1991

Table 5 provides a list of the 30 top-ranking descriptors between 1989 and 1991. It was noted that there were 13 new descriptors, indicating that the number of descriptors that maintained high-ranking status in two consecutive periods (i.e. for 6 years) had improved by over 100% (i.e. from just 6 in 1986–1988 to 13 in 1989–1991). The new descriptors included, in descending order, HIV-1; prevalence; tuberculosis, pulmonary; tuberculosis; health knowledge, attitudes, practice; AIDS-related opportunistic infections; urban population; pregnancy complications, infectious; demography; counseling; contraceptive devices, male; and socioeconomic factors. 17 descriptors appeared in both tables 4 and 5, and were led by acquired immunodeficiency syndrome, followed by, among others, HIV infections; HIV seropositivity; adolescence; sexual behavior; risk factors; and sexually transmitted diseases. The number of hits for most of

the common descriptors demonstrated an upward trend. Also worth noting is the rapid rise of the subject heading HIV infections from position seven in table 4 to position two in table 5.

Table 5: Distribution of records by the main MESH terms, 1989–1991

Rank		No. of hits	Rank		No. of hits
1	Acquired Immunodeficiency Syndrome	407	15	Tuberculosis	35
2	HIV Infections	262	16	HIV Antibodies	35
3	HIV Seropositivity	127	16	Health Knowledge, Attitudes, Practice	32
4	Adolescence	114	17	AIDS-related Opportunistic Infections	30
5	HIV-1	103	18	Enzyme-Linked Immunosorbent Assay	29
6	Sexual Behavior	63	19	Urban Population	28
6	Risk Factors	63	20	Prostitution	28
7	Sexually Transmitted Diseases	57	21	Delivery of Health Care	26
8	Health Education	53	24	Pregnancy Complications, Infectious	24
9	Virus Diseases	50	24	Demography	24
10	Disease	46	24	Counseling	24
11	Prevalence	42	24	Contraceptive Devices, Male	24
12	Tuberculosis, Pulmonary	41	25	Socioeconomic Factors	23
13	Pregnancy	40	25	Organization and Administration	23
14	HIV Seroprevalence	36	26	Research	22

5.2.5 1992–2005

The previous years were characterized by turbulence, with most descriptors changing position. New terms were introduced into the MESH thesaurus, while others either disappeared completely or were relegated to below the threshold of thirtieth position. Nevertheless, as table 4 shows, descriptors had started stabilizing by 1989–1991, a period during which the rankings of 17 descriptors remained stable and/or improved when compared with the previous years. This is clearly shown in table 6, which provides a rank distribution of 57 MESH subject headings from 1992 to 2005 (14 years) quite unlike that of the 89 descriptors that were high-ranking topics of research during the 12 years between 1980 and 1991.

Until 1992, acquired immunodeficiency syndrome was the most researched topic. Subsequently, and as table 6 shows, this MESH term changed positions with HIV infections which, from 1992 to 2005, occupied first position. Other descriptors that ranked highly throughout the period were: HIV-1; adolescence; pregnancy complications, infectious; anti-HIV agents; pregnancy; risk factors; disease transmission, vertical; HIV seropositivity; prevalence; and sexual behavior. A few terms emerged to join the ranks of the top 10 in 2004–2005. These include anti-HIV agents and disease transmission, vertical, which appeared in 1998–2000, Nevirapine, viral load, CD8-positive T-lymphocytes, anti-retroviral therapy (highly active), treatment outcome, risk taking, breast feeding, sexual partners and anti-retroviral agents. Interestingly, most of the top-ranking terms in 1992–1994, 1995–1997 and 1998–2000 did not feature in the top MESH subject headings for 2004–2005. These terms include virus diseases, disease, HIV seroprevalence, age factors, health education, demography, population, HIV antibodies, economics, rural population, family planning services, education and health planning.

Table 6: Rank distribution of the main MESH terms, 1992–2005

No.	Main MESH terms	1992–1994	1995–1997	1998–2000	2001–2003	2004–2005
1	HIV Infections	1	1	1	1	1
2	Acquired Immunodeficiency Syndrome	2	2	2	2	2
3	HIV-1	4	4	3	3	3
4	Adolescence	3	3	4	4	4
5	Pregnancy Complications, Infectious	21	24	8	9	5
6	Anti-HIV Agents	-	-	6	3	6
7	Pregnancy	14	15	5	6	7
8	Risk Factors	7	10	14	11	8
9	Disease Transmission, Vertical	-	-	11	13	9
10	HIV Seropositivity	6	7	7	7	10
11	Prevalence	13	12	10	10	11
12	Sexual Behavior	10	11	15	15	12
13	AIDS-Related Opportunistic Infections	8	8	12	8	13
14	Sexually Transmitted Diseases	11	9	9	12	13
15	Tuberculosis	15	13	13	16	14
16	Tuberculosis, Pulmonary	15	68	15	14	14
17	Condoms	20		22	21	15
18	Health Knowledge, Attitudes, Practice	19	14	17	20	16

19	CD4 Lymphocyte Count	-	-	22	18	17
20	Nevirapine	-	-	-	19	17
21	Viral Load	-	-	-	-	18
22	CD8-Positive T-Lymphocytes	-	-	-	-	19
23	Incidence	-	26	18	17	20
24	Anti-Retroviral Therapy, Highly Active	-	-	-	-	21
25	Treatment outcome	-	-	-	-	22
26	Social Economic Factors	-	25	-	28	23
27	Risk Taking	-	-	-	-	24
28	Breast Feeding	-	-	-	-	25
29	Sexual Partners	-	-	-	-	26
30	Anti-Retroviral Agents	-	-	-	27	27
31	Virus Diseases	5	5	14	-	-
32	Disease	6	6	14	-	-
33	HIV Seroprevalence	9	23	21	-	-
34	Age Factors	12	24	-	-	-
35	Health Education	16	16	20	-	-
36	Demography	17	18	-	-	-
37	Population	18	20	-	-	-
38	HIV Antibodies	18	25	-	-	-
39	Economics	22	19	-	-	-
40	Rural Population	23	-	23	23	-
41	Family Planning Services	23	-	-	-	-
42	Education	24	-	-	-	-
43	Health Planning	25	-	-	-	-
44	Population Characteristics	26	27	-	-	-
45	Research	-	17	-	-	-
46	Organization and Administration	-	21	-	-	-
47	Sarcoma, Kaposi	-	28	-	-	-
48	Disease outbreaks	-	-	16	22	-
49	Zidovudine	-	-	19	-	-
50	Health policy	-	-	23	-	-
51	HIV Envelope Protein gp120	-	-	23	-	-
52	Politics	-	-	23	-	-
53	AIDS vaccines	-	-	-	27	-
54	Drug industry	-	-	-	24	-
55	HIV seronegativity	-	-	-	26	-
56	Molecular Sequence Data	-	-	-	26	-
57	Rural Health	-	-	-	25	-

5.3 MESH SUB-TOPICS RELATING TO HIV/AIDS, 1981–2005

Macias-Chapula et al (1999) and Bierbaum et al (1992) opine that a study of the sub-headings would reveal the importance placed on specific areas of research in a broader subject or topic at a given time. Table 7 reveals that there were a number of sub-headings relating to both acquired immunodeficiency syndrome (AIDS) and HIV infections. In total, there were 32 sub-headings relating to HIV/AIDS. The heading acquired immunodeficiency syndrome accounted for 29 sub-headings, while all 32 sub-headings were focus areas of research under HIV infections. At the top of the list are, in descending order according to the total number of postings, epidemiology (1,986), prevention and control (1,565), transmission (1,036), complications (948), and drug therapy (561). Reading table 7 from left to right reveals that epidemiology and immunology ranked first in 1983–1985 with 6 postings each, followed by diagnosis (5), and transmission and complications, which yielded 3 hits each. Other sub-headings that featured in 1980–1982 are microbiology and pathology, which ranked fifth overall in that year, followed by blood and etiology, which ranked sixth. These sub-headings were overtaken by others in subsequent periods to rank eighteenth, nineteenth, thirteenth and fifteenth respectively. Furthermore, all the sub-headings appearing between 1983 and 1988 were associated with AIDS. The descriptor HIV infections was introduced in 1988 (Bierbaum et al 1992), as shown in table 3 above, but there were no subject sub-fields associated with it. The rapid rise in the use of the term drug therapy, particularly since 1998, perhaps reveals increased research focus on the use of drugs associated with HIV/AIDS. Other sub-headings that rose quickly to rank among the terms most frequently associated with HIV/AIDS include virology, mortality and economics.

Table 7: Number of publications dealing with research in the sub-fields of HIV and AIDS

Rank	Overall rank	AIDS HIV	Sub-field of study	1983-1985		1986-1988		1989-1991		1992-1994		1995-1997		1998-2000		2001-2003		2004-2005		Sub-total		TOTAL	
				AIDS	HIV	AIDS	HIV	AIDS	HIV	AIDS	HIV	AIDS	HIV	AIDS	HIV	AIDS	HIV	AIDS	HIV	AIDS	HIV		
1	1	1	Epidemiology	6	0	47	0	101	66	88	114	61	159	106	284	132	450	67	305	608	1378	1986	
2	2	2	Prevention & Control	0	0	23	0	54	21	72	43	79	113	100	241	127	361	72	259	527	1038	1565	
3	3	3	Transmission	3	0	22	0	52	20	40	51	47	103	44	194	38	210	25	187	271	765	1036	
4	4	4	Complications	3	0	16	0	48	40	41	65	28	105	25	148	28	225	14	162	203	745	948	
5	5	5	Drug Therapy	0	0	1	0	5	2	2	5	10	8	43	59	71	181	47	127	179	382	561	
6	6	6	Immunology	6	0	11	0	13	14	7	21	9	34	16	65	22	92	6	74	90	300	390	
7	7	7	Diagnosis	5	0	10	0	17	10	13	16	7	32	14	71	10	81	12	79	88	289	377	
8	8	8	Virology	0	0	0	0	0	0	0	9	7	22	38	72	13	103	9	78	67	284	351	
9	9	9	Psychology	0	0	2	0	18	11	20	9	29	30	12	31	24	45	17	53	122	179	301	
10	10	10	Mortality	0	0	1	0	5	1	10	11	16	10	19	35	52	50	28	51	131	158	289	
11	11	11	Economics	0	0	0	0	6	2	5	2	3	9	18	19	44	51	13	29	89	112	201	
12	12	12	Therapy	0	0	1	0	4	5	20	8	11	10	22	22	22	34	10	28	90	107	197	
13	13	13	Blood	1	0	1	0	2	1	6	12	5	20	4	27	5	55	4	25	28	140	168	
14	14	14	Ethnology	0	0	0	0	3	6	3	6	11	17		14	9	25	8	12	34	80	114	
15	15	15	Etiology	1	0	3	0	3	1	1	4	2	11	21	16	7	17	1	19	39	68	107	
16	16	16	Physiopathology	0	0	0	0	2	3	1	3	4	2	7	18	10	22	1	16	25	64	89	
17	17	17	Nursing	0	0	2	0	5	1	6	3	3	2	7	4	14	15	7	14	44	39	83	
18	18	18	Microbiology	2	0	0	0	5	2	6	12	0	5	3	7	4	11	0	5	20	42	62	
19	19	19	Pathology	2	0	1	0	1	2	0	4	3	3	5	7	3	18	3	8	18	42	60	
20	20	20	Genetics	0	0	0	0	0	0	0	0	1	1	2	8	1	9	1	11	5	30	35	
21	21	21	Classification	0	0	0	0	2	0	1	2	1	3	0	5	2	4	0	4	6	18	24	
22	22	22	Metabolism	0	0	0	0	1	0	0	0	0	0	2	3	0	6	0	5	3	14	17	
23	23	23	Congenital	0	0	0	0	1	0	2	1	1	0	0	2	1	3	0	2	5	8	13	
24	24	24	Parasitology	0	0	0	0	2	0	1	0	0	1	0	0	2	2	0	2	5	5	10	
25	25	25	History	0	0	0	0	0	0	1	2	0	0	0	0	3	1	0	1	4	4	8	
26	26	26	Radiology	0	0	0	0	0	0	0	0	1	2	0	0	0	3	0	0	1	5	6	
27	27	27	Surgery	0	0	0	0	0	0	0	0	1	0	0	1	0	3	0	1	1	5	6	
28	28	28	Rehabilitation	0	0	0	0	0	0	1	1	0	0	0	0	1	1	1	2	2	4	6	
29	29	29	Urine	0	0	0	0	1	0	0	0	0	0	0	1	0	1	0	1	1	3	4	
30	30	30	Cerebrospinal Fluid	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	1	
31	31	31	Diet Therapy	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	
32	32	32	Ultrasonography	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	1

6 CONCLUSIONS AND RECOMMENDATIONS

As illustrated in table 1, the number of subject terms relating to HIV/AIDS has grown exponentially. The growth in the number of subject headings appears to be associated with the increase in the number of records, as indicated in table 1: as the number of records increased, so did the number of terms. However, the number of publications alone could not have caused an increase in the number of terms, especially if those publications dealt with the same issues or topics. This therefore means that the publications must have covered different topics over time. New publications may have introduced new ideas, thereby making the growth of subject headings possible. The growth in the number of terms, which in part stems from the growth of ideas, may imply the growth of knowledge in the subject domain.

Another factor that may have led to the continued expansion in the number of terms is the variety of professionals to have become involved in HIV/AIDS research. HIV/AIDS is becoming an increasingly multidisciplinary topic. In a study conducted by Onyancha (2006), professionals who were involved in HIV/AIDS research in South Africa between 1986 and 2005 were drawn from disciplines as varied as theology, psychology, educational psychology, health studies, medical sciences, social anthropology, sociology (or social sciences), and nursing. In addition, this author, while analyzing records by the subject areas of study, found that most HIV/AIDS research originated from medical sciences, which yielded 195 records (22.49%), followed by psychology (173 or 19.95%), education (76 or 8.77%) and social work (72 or 8.30%), while sociology, health sciences and religion ranked fifth, sixth and seventh, with 59 (6.81%), 56 (6.46%), and 49 (5.65%), respectively. Others – with 10 or more projects, and in descending order – included business administration (33), law (26), communication science (19), anthropology (15), and economics (12) (see Onyancha, 2006). In total, HIV/AIDS research originated in 41 disciplines or subjects. It is assumed that each professional would use different approaches and ideas associated with his/her field of research to conduct a study on HIV/AIDS, thereby contributing new ideas from his/her field of specialization to HIV/AIDS research. Perhaps this may explain the continued growth of subject headings used to describe HIV/AIDS literature that simultaneously reflect a variety of disciplines engaged in HIV/AIDS research.

As regards the number of terms per paper, the current study discovered a mixed pattern in which the average number of terms rose from 14.11 in 1980–1982 to 14.61 in 1983–1985, only to decrease to 13.75 in 1986–1988. This figure stood at 14.30 in 1989–1991 and rose steadily during the next two periods to stop at 16.17 in 1995–1997. The average number of terms then fell to 15.63, increasing to 15.92 in 2001–2003 and to 16.48 in 2004–2005. Similar observations were made by Bierbaum and Brooks (1995). In their explanation of the pattern, these authors opined that the variances in indexing intensity may have been due to “changes in literature (such as greater complexity of individual articles) or to a greater depth and thoroughness,”

although, as they observed, such arguments were not readily apparent from the data (Bierbaum & Brooks 1995:533).

A trend analysis of the periodic subject content analysis provides an insight into the important or main focus areas of research as reflected in HIV/AIDS literature during specific time periods. Through an analysis of AIDS literature, a number of writers have observed the explosive growth and astonishing development of the field (Bierbaum et al 1992; Macias-Chapula et al 1999). The findings of this study reveal that whereas the epidemic was first clinically diagnosed and given the name acquired immunodeficiency syndrome in 1982 (Begley, Check, Wingert & Conway 2001; Konforti 2001; National Institute of Allergy and Infectious Diseases [NIAID] 2003; Self, Filardo & Lancaster 1989), it was not until 1983 that this was used as a MESH indexing term (Macias-Chapula et al 1999). During 1980–1982, the high-ranking subject headings were Burkitt lymphoma, antibodies (bacterial and viral) and human herpesvirus 4. Hepatitis was also a subject of research during 1980–1982. According to a study conducted by Small and Greenlee (1990:171) on the co-citation of AIDS research, the only cluster that was formed in 1981, “although not concerned with AIDS, dealt with the problems that arise when the immune system is suppressed by cancer therapy”. The 1981 cluster, according to these authors, consisted of papers on opportunistic infections such as pneumocystis carinii pneumonia experienced by cancer patients, a fact borne out by table 2. Again, the absence of HIV/AIDS-related descriptors may be attributed to the fact that the disease was diagnosed only in October 1982, and the name was therefore still new; generally some time elapses between the coinage of a term and its usage (Chan, Jin, Rousseau, Vaughan & Yu 2002). Two aspects were the focus of research in 1981 and 1982, according to Small and Greenlee (1990); these were infection in cancer patients, and the discovery of AIDS.

The period 1983–1985 ushered in a new era of research dedicated to understanding the AIDS epidemic and its relationship with other viral infections. Naturally, upon the discovery of an epidemic such as AIDS, the immediate response of the scientific community is to seek to identify its cause and properties before embarking on eradication. Perhaps the high-ranking status of terms such as Kaposi sarcoma, retroviridae infections, adolescence, homosexuality, viral antibodies and viral antigens is a result of the aforementioned activities. Other notable entrants to the HIV/AIDS literature were other sexually transmitted diseases such as chlamydia and opportunistic diseases (cytomegalovirus infections), implying a close link perceived between these infections and AIDS during the early years of the disease.

The cause of AIDS, the human immunodeficiency virus, was the subject of considerable research in 1986–1988, as illustrated in table 4. MESH terms associated with HIV such as HIV seropositivity, HIV antibodies, HIV, and HIV infections emerged from nowhere (so to speak) to join the top ten most used descriptors. The subject headings ranked second, fifth, sixth and seventh respectively. The focus had apparently shifted from the definition of AIDS to its cause. However, acquired immunodeficiency syndrome still

remained at the top of the table. Other areas that attracted researchers' attention were the risk factors associated with HIV/AIDS, represented in table 4 by the emergence of terms such as risk factors, which ranked sixteenth, prostitution (ranking sixteenth), and sexual behavior (ranking eighteenth).

The exponential rise of the term HIV infections from position 7 in 1986–1988 to position 2 in 1989–1991 is significant. Acquired immunodeficiency syndrome still maintained its position at the head of the list. Table 5 also revealed that risk factors remained a subject of research. The rank status of the MESH terms sexual behavior and risk factors improved (although prostitution dropped in rank), while pregnancy complications, infectious was introduced into the list of the 30 top-ranking MESH terms. Particular attention was given to AIDS-related opportunistic infections. The emergence of the terms tuberculosis; tuberculosis, pulmonary and AIDS-related opportunistic infections attests to the shift in HIV/AIDS research to the study of opportunistic infections in HIV infected persons. This may indicate that research emphasis had shifted to the means of controlling the causal factors of deaths associated with HIV infected persons.

The rapid increase in the prevalence of the term HIV infections resulted in this MESH term occupying first place in the list of the 30 top-ranking MESH descriptors in 1992–1994 and thereafter. Acquired immunodeficiency syndrome switched positions with HIV infections and settled in second position between 1992 and 2005, as shown in table 6. HIV-1, a descriptor that emerged in 1989–1991, during which it ranked fifth, ranked fourth in 1992–1994 and 1995–1997, and third in 2001–2003 and 2004–2005. The absence of other types of HIV (i.e. HIV-2) was noted. Adolescence (or adolescent) is a term that has maintained a high rank throughout the period of study, perhaps because young people are the most vulnerable, or because prevention and control measures were, and still are, targeted at this age group: studies have shown that more than half of those newly infected with HIV are between 15 and 24 years old (United Nations Development Programme and Centre for African Family Studies as cited in Onyanacha & Ocholla 2006), and it is estimated that 11.8 million young people are living with HIV/AIDS. It was not possible to determine from the data whether the pattern of research on adolescents or adolescence is influenced by the high prevalence of the disease among the youth. A number of terms related to HIV/AIDS drugs have also emerged in the recent past, which is probably indicative of a shift in research. These terms include anti-HIV agents; Nevirapine; anti-retroviral therapy, highly active; anti-retroviral agents; Zidovudine; HIV envelope protein gp 120; AIDS vaccines; drug industry; and molecular sequence data. Most of these terms emerged in 1998–2000.

Concerning the sub-fields of research, table 7 provides a list of 32 sub-headings relating to HIV/AIDS that received attention from researchers between 1983 and 2005. The sub-topics fall into three categories, namely (1) the most researched (2) moderately researched and (3) the least researched areas. The most researched sub-topics (sub-fields) of HIV/AIDS include epidemiology, prevention and control, transmission, complications, and drug therapy. Each of these terms yielded over 500 postings. The

second category contained immunology, diagnosis, virology, psychology, mortality, economics, therapy, blood, ethnology, and etiology. Finally, the least researched sub-fields of HIV/AIDS included physiopathology, nursing, microbiology, pathology, genetics, classification, metabolism, congenital, parasitology, history, radiology, surgery, rehabilitation, urine, cerebrospinal fluid, diet therapy, and ultrasonography. A reading of the table from left to right reveals that epidemiology, prevention and control, transmission and complications arising from HIV/AIDS have dominated since the beginning of the epidemic. Researchers can therefore be said to be focusing on the epidemiological aspects of the disease. Similar findings were reported in Latin America and the Caribbean (LAC) (Macias-Chapula et al 1999). However, a few sub-headings that appeared among the highest ranked in this study (e.g. complications, diagnosis, therapy, drug therapy, and psychology) did not emerge as important areas of research in the study of the LAC region. The pattern of research in various sub-fields of HIV/AIDS observed by Small and Greenlee (1990) was, in many respects, similar to that observed in the current study. For instance, the above authors noted that in 1981 researchers were concerned primarily with finding the possible cause of immune suppression. Then, two factors were suspected, notably, cytomegalovirus (CMV) and herpesvirus. These authors observed that by 1983, research was focused on the clinical description of AIDS in addition to the possible origin and cause of the epidemic. One of the fields investigating HIV/AIDS at the time was immunology. Finally, the observation by Small and Greenlee that most papers published in 1984 were dedicated to research that focused on the homosexual population and/or various opportunistic infections was also confirmed in the current study.

The following areas merit further research:

- A co-word analysis of the various subject terms so as to identify the inter-relatedness of HIV/AIDS-specific terms and other terms as indexed in the MEDLINE database
- For the purposes of triangulation, a co-citation analysis of HIV/AIDS literature needs to be conducted to study the emerging focus areas of HIV/AIDS research using other bibliographic databases, such as the Science Citation Index and Social Sciences Citation Index

REFERENCES

- Begley, S, Check, E, Wingert, P & Conway, F. 2001. AIDS at 20. *Newsweek* 137(24):34.
- Bierbaum, EG & Brooks, TA. 1995. The literature of acquired immunodeficiency syndrome (AIDS): continuing changes in publication patterns and subject access. *Journal of the American Society for Information Science* 46(7):530–536.
- Bierbaum, EG, Brooks, TA, & Brooks, RM. 1992. Subject control of the literature of acquired immunodeficiency syndrome (AIDS). *Information Processing and Management* 28(1): 89–98.

- Chan, LCY, Jin, B, Rousseau, R, Vaughan, L & Yu, Y. 2002. Newspaper coverage of SARS: a comparison among Canada, Hong Kong, Mainland China and Western Europe. *Cybermetrics* 6/7.
- <http://www.cindoc.csic.es/cybermetrics/articles/v6i1p1.html> (Accessed 16 July 2007).
- Cohen, J. 2000. Is AIDS in Africa a distinct disease? *Science* 288(5474):2153–2155.
- Farmer, P. 1999. AIDS and social scientists: critical reflections, in C. Becker et al (eds). *Experiencing and understanding AIDS in Africa*. Dakar: CODESRIA.
- Kizito, M. 2002. The importance of content analysis in HIV/AIDS reporting. Paper presented at a UNESCO sub-regional training workshop on legal and ethical issues in HIV/AIDS reporting for the media women, 5–8 March 2002, Nairobi, Kenya. <http://aids-africa.unon.org/events/Dr%20Kizitos.doc> (Accessed 12 January 2009).
- Konforti, B. 2001. Twenty years ago. *Nature Structural Biology* 8(7):582.
- Macias-Chapula, CA. 2000. AIDS in Haiti: a bibliometric analysis. *Medical Library Association Bulletin* 88(1):56–61.
- Macias-Chapula, CA & Mijangos-Nolasco, A. 2002. Bibliometric analysis of AIDS literature in Central Africa. *Scientometrics* 54(2):309–317.
- Macias-Chapula, CA, Sotolongo-Aguilar, GR, Magde, B, & Solorio-Lagunas, J. 1999. Subject analysis of AIDS literature, as produced in Latin America and the Caribbean. *Scientometrics* 46(3):563–574.
- National Institute of Allergy and Infectious Diseases (NIAID). (2003). General information about HIV/AIDS treatment. <http://www.niad.nih.gov/daids/therapeutics/geninfo/geninfo.htm#1> (Accessed 1 April 2006).
- Onyancha, OB. 2006. Empowering the South African community's AIDS intervention workforce: an informetric study of HIV/AIDS research projects, with special reference to master's and doctoral dissertations and theses. *South African Journal of Libraries and Information Science* 72(1):56–71.
- Onyancha, OB & Ocholla, DN 2004. A comparative study of the literature on HIV/AIDS in Kenya and Uganda: a bibliometric study. *Library & Information Science Research* 26:434–447.
- Onyancha, OB & Ocholla, DN. 2006. HIV/AIDS research and the youths: an informetric analysis of the literature. *South African Journal of Libraries & Information Science* 72(2):85–97.
- Onyancha, OB & Ocholla, DN. 2009. Is HIV/AIDS in Africa distinct? What can we learn from an analysis of the literature? *Scientometrics* 79(2):277–296.
- Palmquist, M. (no date). Content analysis. <http://www.colostate.edu/Depts/WritingCenter/references/research/content/page2.htm> (Accessed 21 June 2006).
- Self, PHC, Filardo, THW & Lancaster, FW. 1989. Acquired immunodeficiency syndrome (AIDS) and the epidemic growth of the literature. *Scientometrics* 17(1–2):49–60.
- Small, H & Greenlee, E. 1990. A co-citation study of AIDS research, in CL Borgman (ed), *Scholarly communication and bibliometrics*: 166–193. Newbury Park, CA: Sage.
- UNAIDS. 2003. AIDS epidemic update: December 2003. http://www.unaids.org/wad/2003/Epiupdate2003_en/Epi03_03_en.htm#P28_3962 (Accessed 21 June 2006).

APPENDIX

Appendix A: List of the names of countries and regions used in downloading papers from MEDLINE

Angola	Botswana	Djibouti	Eritrea	Ethiopia
Kenya	Lesotho	Malawi	Mozambique	Namibia
Somalia	South Africa	Sudan	Swaziland	Tanzania
Uganda	Zambia	Zimbabwe	Eastern Africa	Africa, East*
Southern Africa	Africa, South*			

Appendix B: List of terms used to identify HIV/AIDS papers from MEDLINE

Acquired Immuno-deficiency Syndrome	Immuno-deficiency Syndrome, Acquired	Immunologic Deficiency Syndrome, Acquired	Acquired Immune Deficiency Syndrome	Pneumonia, Pneumocystis Carinii
AIDS Arteritis, Central Nervous System	AIDS Dementia Complex	AIDS Seropositivity	HIV Seroprevalence	Immunologic Deficiency Syndromes
HIV*	HTLV-III	LAV-HTLV-III	Receptors, HIV	Immunoblastic Lymphadenopathy
Human T-Cell Lymphotropic Virus Type III	Sarcoma, Kaposi	Human Immuno-deficiency Virus	AIDS related complex	Human T Lymphotropic Virus Type III
Cytomegalic Inclusion Disease	Immuno-deficiency Virus, Human	Virus, Human Immuno-deficiency	Viruses, Human Immuno-deficiency	Reverse Transcriptase Inhibitors
Human T-Cell Leukemia Virus				