Pre-exposure Prophylaxis in HIV Research: A Latent Dirichlet Allocation Analysis (GAP$_{RESEARCH}$)

Bach X. Tran[1,2]*, Hai T. Phan[3], Quang N. Nguyen[1,3,4], Olagoke A. Ewedairo[5], Giang T. Vu[6], Giang H. Ha[3], Thao P. Nguyen[6], Carl A. Latkin[2], Cyrus S.H.-Ho[7], and Roger C.M.-Ho[8,9,10]

[1]*Institute for Preventive Medicine and Public Health, Hanoi Medical University, Hanoi, Vietnam;* [2]*Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA;* [3]*Institute for Global Health Innovations, Duy Tan University, Da Nang, Vietnam;* [4]*UnivLyon, Université Claude Bernard Lyon 1, Villeurbanne, France;* [5]*GlaxoSmithKline Pharmaceuticals Limited, Philadelphia, PA, USA;* [6]*Center of Excellence in Evidence-based Medicine, Nguyen Tat Thanh University, Ho Chi Minh City, Vietnam;* [7]*Department of Psychological Medicine, National University Hospital, Singapore;* [8]*Center of Excellence in Behavioral Medicine, Nguyen Tat Thanh University, Ho Chi Minh City, Vietnam;* [9]*Department of Psychological Medicine, Yong Loo Lin School of Medicine;* [10]*Institute for Health Innovation and Technology (iHealthtech). National University of Singapore, Singapore*

Table S1. Summary of analytical techniques for each data types

Type of data	Unit of analysis	Analytical methods	Presentations of results
Keywords	Words	Frequency of cooccurrence	Map of keywords clusters
Countries of authors' affiliations	Words	Frequency of cooccurrence	Map of intercountry collaboration
Abstracts	Words	Exploratory factors analyses	Top 50 constructed research domains Clustering map of the landscapes constructed by these domains
Abstracts	Papers	LDA	Ten classifications of research topics
WoS classification of research areas	WoS research areas	Frequency of cooccurrence	Dendrogram of research disciplines (WoS classification)

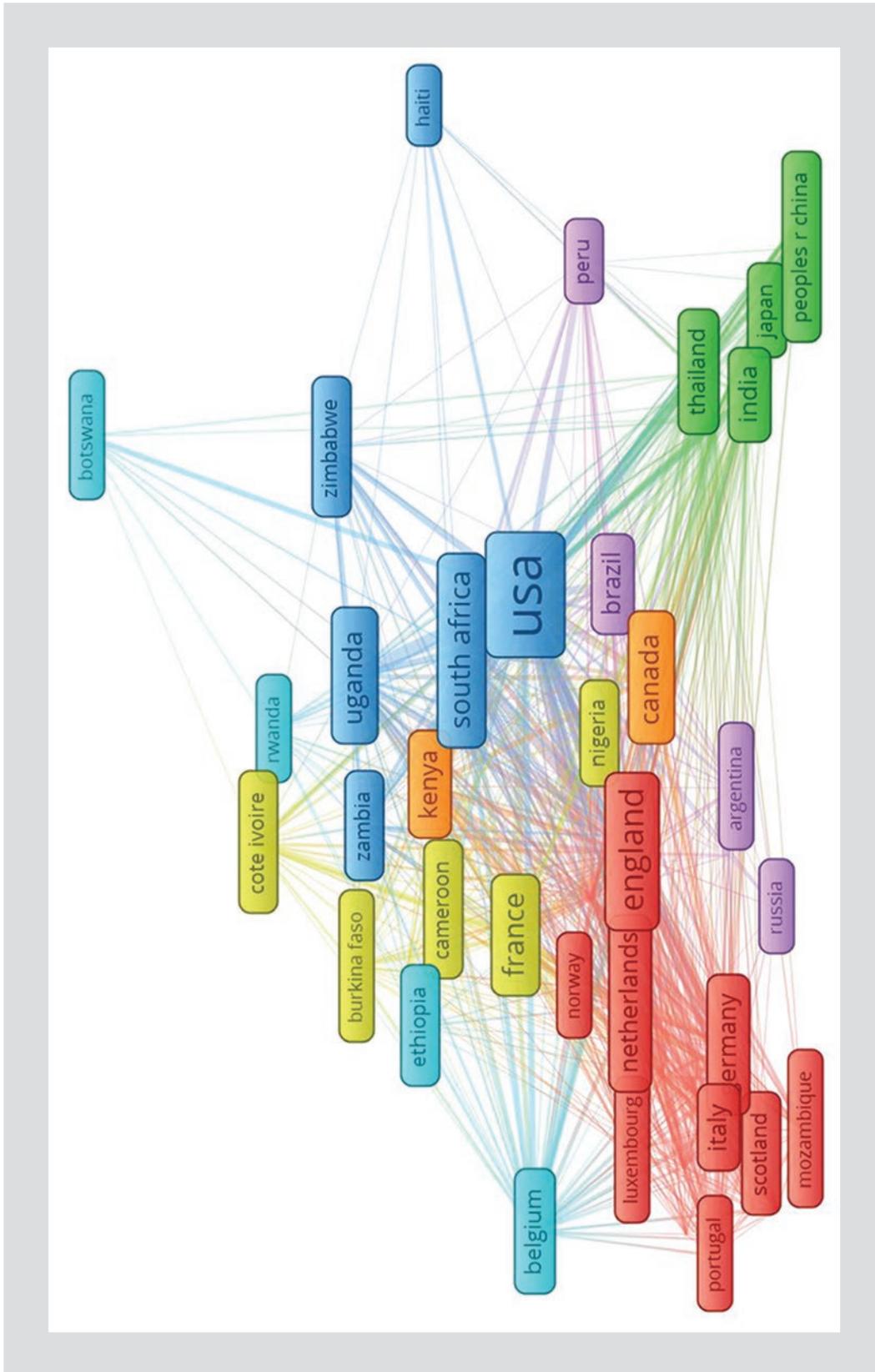WoS: Web of Science, LDA: Latent Dirichlet Allocation

Table S2. General characteristics of publications

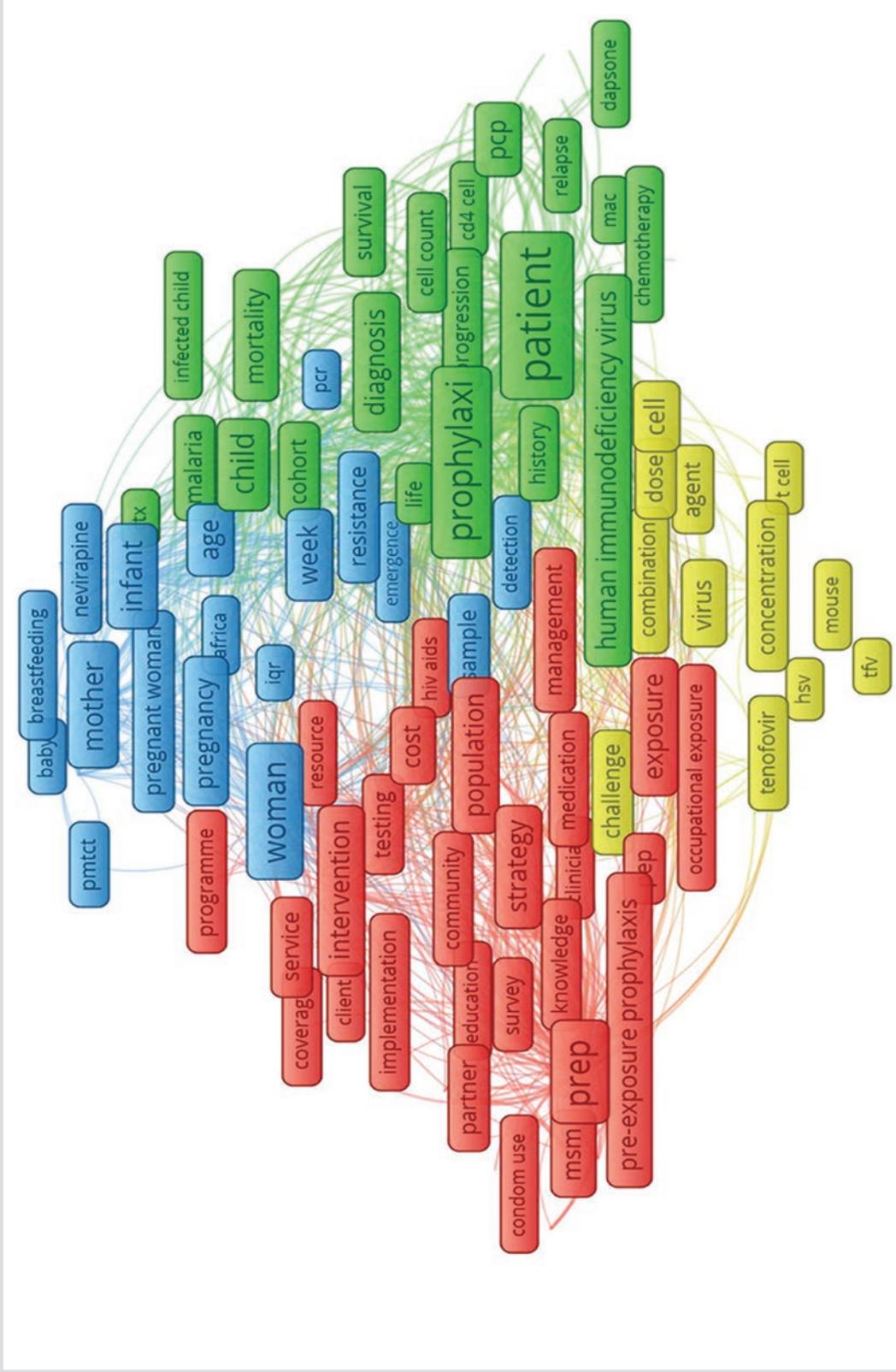| Year published	Total number of papers	Total citations	Mean cite rate per year[1]	Total usage[2] past 6 months	Total usage past 5 years	Mean use rate past 6 months[3]	Mean use rate past 5 years[4]
2017	473	983	2.1	744	1941	1.6	0.8
2016	441	3319	3.8	319	2879	0.7	1.3
2015	375	4409	3.9	175	2747	0.5	1.5
2014	348	6009	4.3	142	3180	0.4	1.8
2013	322	6201	3.9	115	2954	0.4	1.8
2012	268	8626	5.4	113	2566	0.4	1.9
2011	239	5262	3.1	42	1085	0.2	0.9
2010	201	8271	5.1	54	1125	0.3	1.1
2009	153	3984	2.9	32	615	0.2	0.8
2008	144	4248	3.0	21	427	0.1	0.6
2007	165	4853	2.7	28	490	0.2	0.6
2006	109	3483	2.7	22	287	0.2	0.5
2005	108	3480	2.5	18	224	0.2	0.4
2004	113	5819	3.7	17	294	0.2	0.5
2003	103	3318	2.1	15	183	0.1	0.4
2002	99	3951	2.5	13	166	0.1	0.3
2001	120	4785	2.3	19	253	0.2	0.4
2000	133	5940	2.5	13	263	0.1	0.4
1999	127	7820	3.2	18	299	0.1	0.5
1998	136	6251	2.3	16	158	0.1	0.2
1997	132	5788	2.1	18	233	0.1	0.4
1996	143	4220	1.3	13	188	0.1	0.3
1995	94	4601	2.1	6	111	0.1	0.2
1994	94	4091	1.8	10	118	0.1	0.3
1993	88	5134	2.3	5	98	0.1	0.2
1992	73	3589	1.9	5	54	0.1	0.1
1991	50	2867	2.1	4	58	0.1	0.2
1990	1	2	0.1	0	0	0.0	0.0

[1]Mean cite rate per year=Total citations/[total citations*(2018-that year)], [2]Total usage: total downloads, [3]Mean use rate past 6 months=Total usage past 6 months/total number of papers, [4]Mean use rate past [5]years=total usage past 5 years/(total number of papers*5)

Table S3. Top research domains emerged from exploratory factor analysis of all abstracts' contents

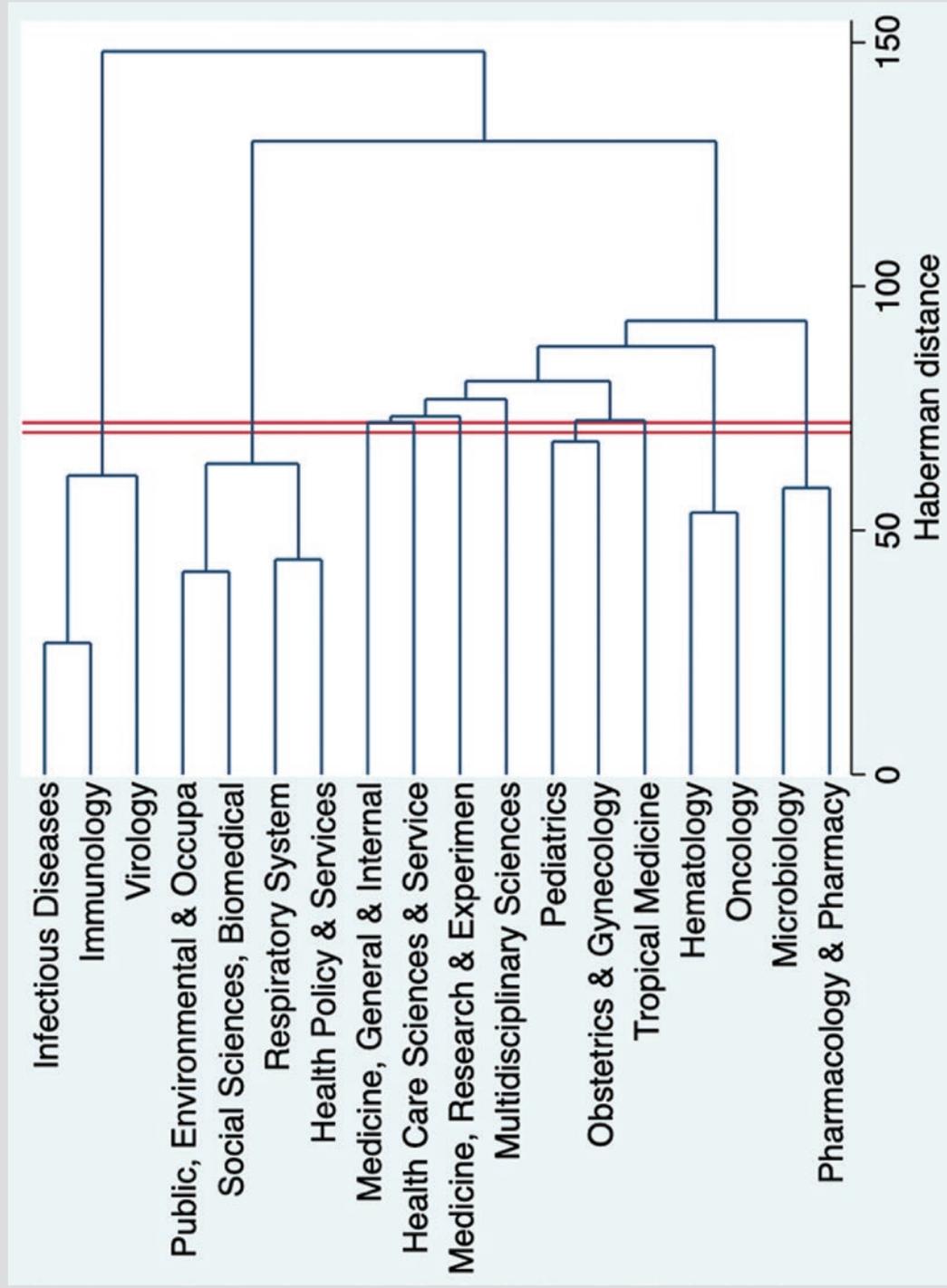| No.	Topic	Terms	Terms' frequency of appearance	Cases (%)
1.	MEN (MSM; SEX)	MSM; MEN; SEX; PREP; PRE; EXPOSURE; AWARENESS; WILLINGNESS; SEXUAL; RISK; PARTNERS; PREVENTION; BARRIERS; UPTAKE	10,976	71.45
2.	FUTURE; CLINICAL DEVELOPMENT	FUTURE; DEVELOPMENT; TRIALS; STRATEGIES; VACCINE; INCLUDE; RECENT; CLINICAL	4734	54.90
3.	CARE SERVICES; UPTAKE	SERVICES; CARE; UPTAKE; BARRIERS; ACCESS; SUPPORT; COMMUNITY; PMTCT; SOCIAL; COUNSELING; PROVIDED; SERVICE; HEALTH	5568	49.87
4.	HEALTHCARE WORKERS; OCCUPATIONAL	WORKERS; OCCUPATIONAL; HCWS; HEALTH CARE; EXPOSURES; PEP; POST-EXPOSURE; BLOOD; EXPOSURE; POST	4625	46.75
5.	INFANT; TRANSMISSION FROM MOTHER	INFANTS; MOTHER; INFANT; MOTHERS; MATERNAL; TRANSMISSION; BIRTH; BORN; PREGNANT; WOMEN; BREASTFEEDING; PMTCT; EXPOSED; PREGNANCY; MTCT	6750	41.81
6.	MORTALITY AND MORBIDITY; OPPORTUNISTIC	MORBIDITY; MORTALITY; OPPORTUNISTIC; MAJOR; INFECTIONS; COMMON	3510	41.24
7.	COST-EFFECTIVENESS; LIFETIME COSTS	COST; COSTS; LIFE; QUALITY; EFFECTIVENESS; EFFECTIVE	2796	34.92
8.	RESISTANCE MUTATIONS; DRUG RESISTANT	MUTATIONS; RESISTANCE; RESISTANT; DETECTED; SAMPLES; DRUG	2572	33.14
9.	DISEASE PROGRESSION	PROGRESSION; SURVIVAL; STAGE; DEATH; DISEASE; FREE	2466	32.63
10.	TEST; TESTING	TESTS; TEST; TESTING; RAPID; LABORATORY; PERFORMED	2342	31.32
11.	SEXUALLY TRANSMITTED	SEXUALLY; TRANSMITTED; INFECTIONS	1806	26.62
12.	SEXUAL ASSAULT; NPEP	ASSAULT; NPEP; SEXUAL; PEP; POST; EXPOSURES; POST-EXPOSURE	2265	26.46
13.	PREGNANCY; PREGNANT WOMEN	PREGNANCY; PREGNANT; WOMEN; ANTENATAL; DELIVERY	2533	26.14
14.	PLASMA CONCENTRATIONS; DOSING	CONCENTRATIONS; PLASMA; DOSING; SAMPLES; MU; BLOOD	1867	25.16
15.	CASE	CASE; REPORT; CASES	1535	24.53
16.	RANDOMIZED; PLACEBO CONTROLLED	RANDOMIZED; PLACEBO; CONTROLLED; TRIAL; ARM; TRIALS	2501	24.40
17.	FTC/TDF; TENOFOVIR AND EMTRICITABINE	FTC; TDF; EMTRICITABINE; TENOFOVIR; TFV; PRE-EXPOSURE; PREP	2330	21.83
18.	COUPLES; PARTNER	COUPLES; PARTNER; PARTNERS; MALE; UNINFECTED	1417	18.24
19.	PNEUMOCYSTIS	PNEUMOCYSTIS; PNEUMONIA; CARINII; PCP; PENTAMIDINE; DAPSONE; TRIMETHOPRIM; SULFAMETHOXAZOLE	3065	17.43
20.	VIRAL LOAD	LOAD; VIRAL; VL; PLASMA	1287	16.70
21.	DOSE NEVIRAPINE	NEVIRAPINE; NVP; SINGLE; DOSE	1241	16.07
22.	ADVERSE EVENTS	ADVERSE; EVENTS; SAFETY	1117	15.27
23.	REGIMEN	REGIMENS; REGIMEN	885	14.27
24.	RESPIRATORY; ACUTE	RESPIRATORY; ACUTE; BACTERIAL; FAILURE	893	14.03
25.	IMMUNE RESPONSE	RESPONSE; IMMUNE	770	12.95
26.	DEVELOPING; COUNTRIES	DEVELOPING; COUNTRIES	780	12.25
27.	PRIOR TO INITIATION	PRIOR; INITIATION	682	12.23
28.	TUBERCULOSIS (TB)	TB; TUBERCULOSIS; ISONIAZID; PREVENTIVE	990	12.02
29.	PROPORTION	PROPORTION; DIAGNOSED	646	11.65
30.	PRE-EXPOSURE	PRE-EXPOSURE; POPULATIONS	669	11.62
31.	LONG TERM	TERM; LONG	859	10.62
32.	SOUTH AFRICA	SOUTH; AFRICA	745	10.56
33.	ANTIBODIES; ANTIBODY	ANTIBODIES; ANTIBODY; MICE; VACCINE	735	10.54
34.	INFECTIOUS	INFECTIOUS; DISEASES	650	9.85
35.	FEMALE	FEMALE; MALE	605	9.70
36.	ANTIVIRAL	ANTIVIRAL; ACTIVITY; HSV	503	7.96
37.	ZIDOVUDINE (ZDV)	ZDV; ZIDOVUDINE	473	7.67
38.	COTRIMOXAZOLE (CTX)	CTX; COTRIMOXAZOLE; MALARIA	482	7.60
39.	REDUCTION	REDUCTION	368	7.04
40.	VAGINAL GEL	VAGINAL; GEL; TFV	427	6.33
41.	DIFFERENCES	DIFFERENCES	322	6.16
42.	PREVENTING	PREVENTING	320	6.12
43.	HEPATITIS B VIRUS (HBV	HBV; HEPATITIS	394	5.95
44.	UNITED STATES	UNITED; STATES	538	5.74
45.	TMP-SMX; SULFAMETHOXAZOLE-TRIMETHOPRIM	TMP; SMX; TRIMETHOPRIM; SULFAMETHOXAZOLE	717	5.61
46.	ASSOCIATION	ASSOCIATION	290	5.55
47.	BREASTFEEDING	BREAST; FEEDING; BREASTFEEDING	416	5.49
48.	DNA PCR	PCR; DNA	347	5.32
49.	FLUCONAZOLE	FLUCONAZOLE	94	1.80

3

Supplementary Figure 1. *The global networking countries having co-authorships in PrEP research in HIV/AIDS. The range of nodes represents the contribution of each country to the total number of publications and the thickness of lines illustrates the proportion of the volume of collaborations. These countries were classified into clusters depending on their level of international collaborations.*

Supplementary Figure 2. *Co-occurrence of most frequent author's keywords. The colors of the nodes refer to the principle components of data structure; the nodes size was scaled to the oc keywords' currences; the thickness of the lines was drawn based on the strength of the association between two keywords.*

5

Supplementary Figure 3. *Co-occurrence of most frequent topics emerged from exploratory factor analysis of abstracts contents.*

Supplementary Figure 4. *Dendrogram of coincidence of research areas using the WoS classifications.*