META-ANALYSIS

e-ISSN 1643-3750
© Med Sci Monit, 2019; 25: 643-655
DOI: 10.12659/MSM.913026

Received: 2018.09.02
Accepted: 2018.10.09
Published: 2019.01.22

Identification of Recent Trends in Research on Vitamin D: A Quantitative and Co-Word Analysis

Authors' Contribution:
Study Design A
Data Collection B
Statistical Analysis C
Data Interpretation D
Manuscript Preparation E
Literature Search F
Funds Collection G

ABCDEF 1 Aolin Yang*
ABCDEF 1 Qingqing Lv*
B 2 Feng Chen
DG 2 Difei Wang
DG 3 Ying Liu
AD 1 Wanying Shi

1 Department of Nutrition, The First Hospital of China Medical University, Shenyang, Liaoning, P.R. China
2 Department of Geriatric Endocrinology, The First Hospital of China Medical University, Shenyang, Liaoning, P.R. China
3 Department of Biochemistry and Molecular Biology, China Medical University, Shenyang, Liaoning, P.R. China

* These authors contribute equally to the study
Corresponding Author: Wanying Shi, e-mail: swying555@sina.com
Source of support: This work was supported by the National Natural Science Foundation of China (grant 31570819); the Science and Technology Projects of Shenyang (grant Z18-5-104); and the Local Development Foundation of Science and Technology Guided by the Central Commission (grant 2016007024)

Background: In recent years, many studies on vitamin D have been published. We combed these data for hot spot analyses and predicted future research topic trends.

Material/Methods: Articles (4625) concerning vitamin D published in the past 3 years were selected as a study sample. Bibliographic Items Co-occurrence Matrix Builder (BICOMB) software was used to screen high-frequency Medical Subject Headings (MeSH) terms and construct a MeSH terms-source article matrix and MeSH terms co-occurrence matrix. Then, Graphical Clustering Toolkit (gCLUTO) software was employed to analyze the matrix by double-clustering and visual analysis to detect the trends on the subject.

Results: Ninety high-frequency major MeSH terms were obtained from 4625 articles and divided into 5 clusters, and we generated a visualized matrix and a mountain map. Strategic coordinates were established by the co-occurrence matrix of the MeSH terms based on the above classification, and the 5 clusters described above were further divided into 7 topics. We classified the vitamin D-related diseases into 12 categories and analyzed their distribution.

Conclusions: The analysis of strategic coordinates revealed that the epidemiological study of vitamin D deficiency and vitamin D-related diseases is a hot research topic. The use of vitamin D in the prevention and treatment of some diseases, especially diabetes, was found to have a significant potential future research value.

MeSH Keywords: **Bibliometrics • Diabetes Mellitus • Neoplasms • Vitamin D**

Full-text PDF: https://www.medscimonit.com/abstract/index/idArt/913026

3041 5 6 41

Background

In recent years, the research interest in studies of vitamin D functions has considerably increased. The functions of this vitamin are not limited only to its substantial skeletal effects, but are also closely related to non-skeletal effects. A growing body of literature has identified negative associations between the level of 25-hydroxyvitamin D and the occurrence of various diseases, such as obesity [1, 2], diabetes [3], thyroid disease [4], coronary heart disease [5], ischemic stroke [6], respiratory tract diseases [7,8], neoplasms [9,10], Alzheimer's disease [11,12], gestational diseases [13], digestive system diseases [14, 15], and skin diseases [16,17]. Vitamin D supplementation decrease the occurrence/development of some diseases, such as neoplasms [18–20], acute respiratory tract infections [21], and premature infant recurrent wheezing [22]. But the beneficial effects of vitamin D supplements are still controversial [23–26]. The widespread expression of vitamin D receptor in various tissues of the body suggests its extensive biological effects. Many studies have shown the extensive role of vitamin D in skeletal and non-skeletal tissues [27–30].

Appropriate literature sources can be used in the application of bibliometric methods and tools to judge the development status of a discipline and predict its development prospects. Co-word analysis is an important method of bibliometric analysis that can identify trends and hot topics in a subject [31]. In a document, if 2 subject words co-occur, they are considered as likely to have potential relationships. Moreover, the frequent co-occurrence of 2 subject words in an article indicates their close relationship. Through the "relationship" between co-occurring subject words, statistical methods, such as cluster and factor analyses, can be applied for further analysis. More specifically, cluster analysis can be used to find the semantic relations of certain research topics. Compared with the traditional clustering methods [32], double-clustering analysis simultaneously clusters rows and columns of a matrix, facilitating the clustering of global information [33,34]. Additionally, double-clustering simultaneously reduces the dimensions of the biaxial directions, requiring fewer parameters.

The number of publications in the PubMed database associated with "Vitamin D" has increased over the years. These articles contain a large amount of important experimental data, population epidemiological studies, randomized clinical trials, and other important data resources. In this study, we conducted a bibliometric analysis using a MeSH terms co-occurrence matrix, established the strategic coordinates of vitamin D, and explored the distribution of vitamin D-related diseases, trying to assess the current status of research on vitamin D and predict its future development trends.

Material and Methods

Data collection

The data used in this study were obtained from the PubMed database, a comprehensive medical database developed and operated by the National Biotechnology Information Center of the United States. We used MeSH Major Topic (major) and time-limited search strategies. Using the first search strategy, "vitamin D"[major] AND ("2015/05/14"[PDAT]: "2018/05/14"[PDAT]), on the search date 2018/05/14, we found 4625 related documents. The second search strategy, "vitamin D"[major] AND ("2002/01/01" [PDAT]: "2005/01/01"[PDAT]), contributed to the identification of 1938 related documents. The third search strategy was "vitamin D"[major], and the relevant documents were extracted year by year, from 1997 to 2017. The 3 groups were downloaded and saved in XML format.

Data extraction and analysis

Data extraction and matrix construction were performed using BICOMB software developed by Professor Cui of the China Medical University and available freely online [35]. To explore the research focused on vitamin D effects, after the data extraction, we found the most common major MeSH terms (Table 1).

Next, we used BICOMB to generate a MeSH terms-source article matrix – the source document set was the row, and the high-frequency MeSH terms was the column (Table 2). We then utilized gCLUTO 1.0 software – a Graphical Cluster Toolkit, developed by Rasmussen and Karypis of Minnesota University – to perform double-cluster analysis on the MeSH terms-source article matrix [36]. Referring to previous literature [37], we employed repeated bisection as the clustering method; cosine was selected as the similarity function, and I2 was selected as the clustering criterion function. To determine the optimum high-frequency MeSH terms classification, we repeated the search several times by selecting a different number of clusters. The external similarity (ESim) and the internal similarity (ISim) were used to optimize the results (Table 3). Further, we generated a visualization MeSH terms-source article matrix and a high-frequency MeSH terms mountain map. Based on the semantic relations, we analyzed clusters keyword meaning and representation of the article content to obtain the class of the hot spot topic associated with vitamin D in each cluster.

Strategic coordinates

Co-occurrence analysis was used to understand and describe the relationship between the scientific topics identified. Additionally, we utilized Microsoft Excel to create a MeSH terms co-occurrence matrix (Table 4) consisting of high-frequency MeSH terms to calculate the intra-class link average

Table 1. High-frequency major MeSH terms from the included publications on vitamin D (n=76 125).

No.	MeSH terms	Frequency n (%*)		Cumulative percentage (%)
1	Humans	4094	(5.38)	5.38
2	Female	2848	(3.74)	9.12
3	Male	2459	(3.23)	12.35
4	Vitamin D/blood	2087	(2.74)	15.09
5	Vitamin D/analogs and derivatives	1613	(2.12)	17.21
6	Middle-aged	1495	(1.96)	19.17
7	Adult	1475	(1.94)	21.11
8	Aged	1067	(1.40)	22.51
9	Vitamin D deficiency/blood	903	(1.19)	23.70
10	Dietary supplements	857	(1.13)	24.83
11	Animals	834	(1.10)	25.92
12	Young adult	644	(0.85)	26.77
13	Vitamin D/administration & dosage	641	(0.84)	27.61
14	Adolescent	563	(0.74)	28.35
15	Risk factors	553	(0.73)	29.07
16	Vitamin D deficiency/epidemiology	513	(0.67)	29.75
17	Cross-sectional studies	499	(0.66)	30.40
18	Vitamin D deficiency/complications	490	(0.64)	31.05
19	Vitamin D/therapeutic use	481	(0.63)	31.68
20	Child	437	(0.57)	32.25
21	Vitamin D/metabolism	411	(0.54)	32.79
22	Vitamin D deficiency/drug therapy	366	(0.48)	33.27
23	Biomarkers/blood	360	(0.47)	33.75
24	Aged, 80 and over	359	(0.47)	34.22
25	Prospective studies	353	(0.46)	34.68
26	Vitamin D/pharmacology	339	(0.45)	35.13
27	Treatment outcome	323	(0.42)	35.55
28	Pregnancy	317	(0.42)	35.97
29	Case-control studies	315	(0.41)	36.38
30	Prevalence	305	(0.40)	36.78
31	Child, preschool	276	(0.36)	37.15
32	Cholecalciferol/administration and dosage	269	(0.35)	37.50
33	Double-blind method	268	(0.35)	37.85
34	Vitamin D Deficiency/diagnosis	263	(0.35)	38.20
35	Seasons	260	(0.34)	38.54

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System]
[ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica]
[Chemical Abstracts/CAS]

Table 1 continued. High-frequency major MeSH terms from the included publications on vitamin D (n=76 125).

No.	MeSH terms	Frequency n (%*)		Cumulative percentage (%)
36	Receptors, calcitriol/metabolism	259	(0.34)	38.88
37	Body mass index	253	(0.33)	39.21
38	Parathyroid hormone/blood	252	(0.33)	39.54
39	Calcifediol/blood	246	(0.32)	39.86
40	Vitamins/therapeutic use	244	(0.32)	40.19
41	Mice	241	(0.32)	40.50
42	Calcitriol/pharmacology	234	(0.31)	40.81
43	Infant	230	(0.30)	41.11
44	Follow-up studies	213	(0.28)	41.39
45	Cohort studies	212	(0.28)	41.67
46	Vitamins/administration and dosage	205	(0.27)	41.94
47	Receptors, calcitriol/genetics	196	(0.26)	42.20
48	Calcium/blood	195	(0.26)	42.45
49	Dose-response relationship, drug	189	(0.25)	42.70
50	Retrospective studies	180	(0.24)	42.94
51	Time factors	177	(0.23)	43.17
52	Cholecalciferol/therapeutic use	172	(0.23)	43.40
53	Infant, newborn	161	(0.21)	43.61
54	Sunlight	155	(0.20)	43.81
55	Cholecalciferol/pharmacology	152	(0.20)	44.01
56	Severity of Illness Index	152	(0.20)	44.21
57	Vitamin D deficiency/prevention and control	152	(0.20)	44.41
58	Nutritional status	146	(0.19)	44.60
59	Rats	142	(0.19)	44.79
60	Diet	138	(0.18)	44.97
61	Logistic models	132	(0.17)	45.14
62	Vitamin D	131	(0.17)	45.31
63	Vitamins/pharmacology	129	(0.17)	45.48
64	Surveys and Questionnaires	127	(0.17)	45.65
65	Cells, cultured	122	(0.16)	45.81
66	Calcitriol/analogs and derivatives	121	(0.16)	45.97
67	Age factors	119	(0.16)	46.13
68	Bone density	115	(0.15)	46.28
69	Disease models, animal	114	(0.15)	46.43
70	Cell line, tumor	114	(0.15)	46.58

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

Table 1 continued. High-frequency major MeSH terms from the included publications on vitamin D (n=76 125).

No.	MeSH terms	Frequency n (%*)		Cumulative percentage (%)
71	Vitamins/blood	114	(0.15)	46.73
72	Randomized Controlled Trials as Topic	114	(0.15)	46.88
73	Prognosis	114	(0.15)	47.03
74	Polymorphism, single-nucleotide	112	(0.15)	47.17
75	Incidence	110	(0.14)	47.32
76	Vitamin D Deficiency/metabolism	110	(0.14)	47.46
77	Odds ratio	104	(0.14)	47.60
78	Vitamin D Deficiency/etiology	102	(0.13)	47.73
79	Linear models	101	(0.13)	47.87
80	Multivariate analysis	101	(0.13)	48.00
81	Cell line	98	(0.13)	48.13
82	Calcitriol/administration and dosage	95	(0.12)	48.25
83	Calcitriol/therapeutic use	94	(0.12)	48.38
84	Ultraviolet rays	94	(0.12)	48.50
85	Sex factors	94	(0.12)	48.62
86	Vitamin D Deficiency/physiopathology	94	(0.12)	48.75
87	Insulin resistance	92	(0.12)	48.87
88	Diabetes Mellitus, Type 2/blood	92	(0.12)	48.99
89	Calcium/metabolism	91	(0.12)	49.11
90	Bone Density/drug effects	90	(0.12)	49.23

* Proportion of the frequency among 76 125 appearances.

Table 2. High-frequency major MeSH terms-source articles matrix (localized).

No.	Major MeSH terms	PMID of source article				
		21642832	23109511	23784946	…	29677309
1	Humans	1	1	0	…	1
2	Female	0	0	0	…	1
3	Male	0	1	0	…	1
4	Vitamin D/blood	0	0	0	…	0
…	…	…	…	…	…	…
89	Calcium/metabolism	0	0	0	…	0
90	Bone Density/drug effects	0	0	0	…	0

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

Table 3. Descriptive and discriminating features and representative articles.

Descriptive and discriminating features				
Cluster 0	**Size* 26**	**ISim 0.246**	**ESim 0.084**	
Descriptive	26826045**	25300588	26845632	27488178
Discriminating	26868944	27488178	26826045	26291437
Cluster 1	**Size 18**	**ISim 0.166**	**ESim 0.025**	
Descriptive	26794222	27998003	27154546	26630444
Discriminating	26794222	27998003	27154546	26630444
Cluster 2	**Size 16**	**ISim 0.178**	**ESim 0.061**	
Descriptive	27413130	28323044	28615261	27717236
Discriminating	27413130	28323044	27717236	26173598
Cluster 3	**Size 14**	**ISim 0.165**	**ESim 0.058**	
Descriptive	28331054	26628439	26861385	28333101
Discriminating	28331054	26628439	27776564	26938997
Cluster 4	**Size 16**	**ISim 0.169**	**ESim 0.065**	
Descriptive	25901090	26184826	28882871	26498119
Discriminating	25901090	26184826	28882871	26009498

ISim – Internal Similarity; ESim – External Similarity. * Size: number of cluster objects; ** PubMed Unique Identifiers of literature.

Table 4. A co-word matrix of high-frequency major MeSH terms (localized).

No.	Major MeSH terms	Humans	Female	...	Bone density/drug effects
1	Humans	4094	2703	...	78
2	Female	2703	2848	...	78
3	Male	2261	2115	...	50
...	...	...	...	...	...
90	Bone Density/drug effects	78	78	...	90

Table 5. Centrality and density of the 5 clusters identified in this study.

Cluster	Intra-class link average	Density-Y	Inter-class link average	Centrality-X
0	260.04	174.07	44.62	18.44
1	38.31	–47.67	10.43	–15.75
2	57.36	–28.61	27.66	1.48
3	36.13	–49.84	21.92	–4.27
4	38.03	–47.95	26.30	0.11
Average	85.97		26.19	

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

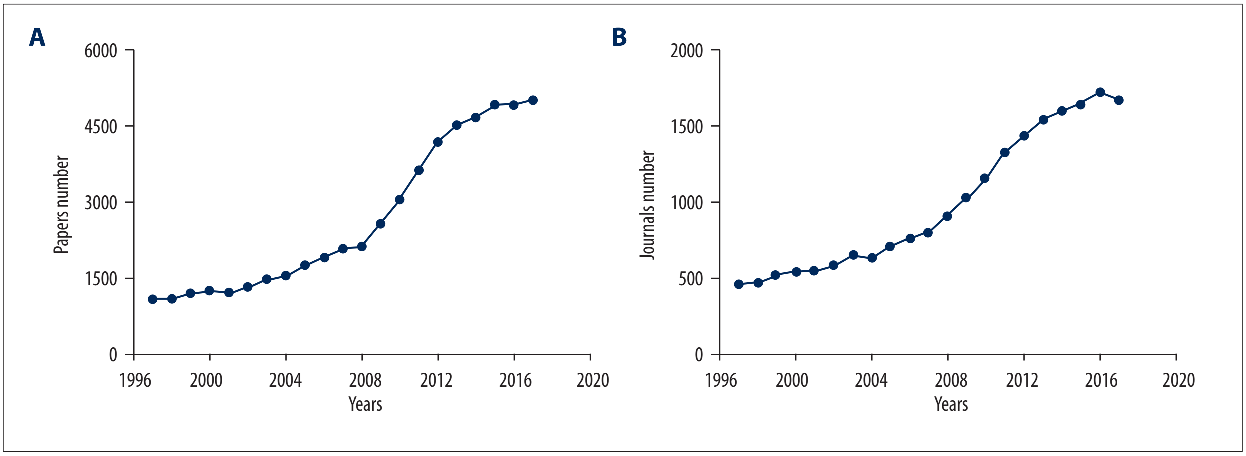


Figure 1. Changes in the number of vitamin D-related papers (**A**) and journals (**B**) from 1997 to 2017.

and the inter-class link average to calculate the centrality and density (Table 5). Two-dimensional coordinates with centrality and density were employed as parameters to construct a graph describing the internal integrity of certain topics and their interaction with other disciplines.

Results

Growth of the papers and journals

"Vitamin D" [major] was retrieved from the PubMed database and 16,678 articles were extracted from 1997/01/01 to 2017/12/31. We used BICOMB software to extract the number of papers and the number of journals year by year, and then statistical analyses were conducted. An explosive growth has been observed in the number of the research papers in the past 10 years (Figure 1A), and the number of journals has also grown rapidly (Figure 1B).

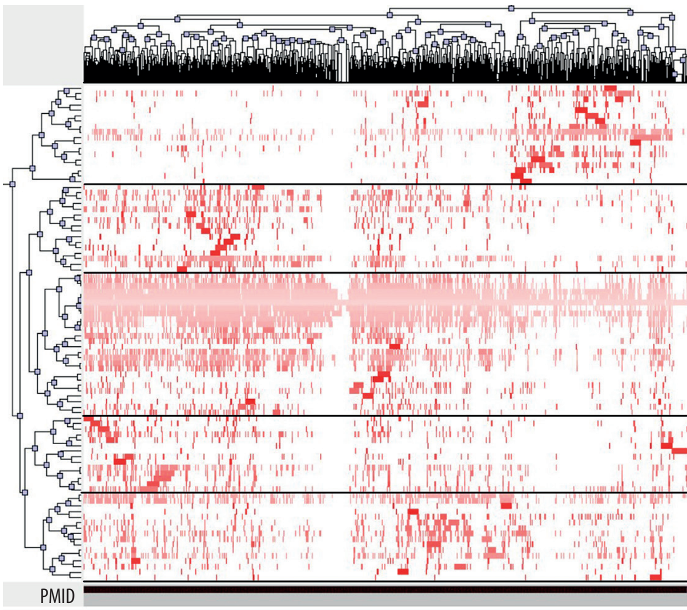
Trends in research on vitamin D

"Vitamin D" [major] was retrieved from the PubMed database, and 4625 documents were found to be released from 2015/05/14 to 2018/05/14. BICOMB software was employed to extract a total number of 11 961 MeSH terms with a cumulative frequency of 76 125 times from 4625 articles. Referring to the definition of the H index, proposed by the American physicist J. E. Hirsch in 2005, we considered the subject whose frequency was greater than its rank to be the high-frequency MeSH terms. In the topic word list, the top 90 were selected as high-frequency MeSH terms, and the cumulative percentage of the frequencies appearing in the literature was 49.23% (37 473/76 125) (Table 1). Based on the co-occurrence of these high-frequency words in an article, we established a MeSH terms-source article matrix. The number "1" in the cell indicates that the word appeared in the article, and "0" indicates that it did not appear. Then, we established a MeSH terms co-occurrence matrix in which numbers represented the number of co-occurrences between the 2 words (Table 4).

We used the gCLUTO software to perform bi-cluster analysis on the MeSH terms-source article matrix derived from the BICOMB software. The 90 high-frequency MeSH terms were divided into 5 clusters, and then a visualization matrix and a mountain map were generated. In the visualization matrix, the colors represent the values in the original data matrix. The gCLUTO is represented in white when close to zero, whereas the progressively deeper red indicates a larger value. The rows of the matrix were rearranged so that the same type of MeSH terms is lined up together, and the black horizontal lines separate the clusters. A heat map of the double-cluster visualization is illustrated in Figure 2 shows, where the hierarchical tree on the left describes the relationship between these high-frequency MeSH terms, and the top-level hierarchical tree denotes the relationship between the articles. This image represents the semantic relationship between MeSH terms and articles used to identify and summarize topics for each cluster (Table 3). In Figure 3, the 90 high-frequency MeSH terms displayed by the visualization matrix are clustered in peaks that represent 5 clusters numbered from 0 to 4. The distance between peaks, the volume, height, and color of the peak are all used to describe the information of the associated cluster under the preset conditions. The distance between peaks indicates the relative similarity of the clusters. The height of each peak is proportional to the internal similarity of the cluster. Furthermore, the volume of the peak is proportional to the number of the major MeSH terms in each cluster. The color of the peak domain indicates the standard deviation within the cluster's objects, the red indicates low deviation, and the blue shows high deviation.

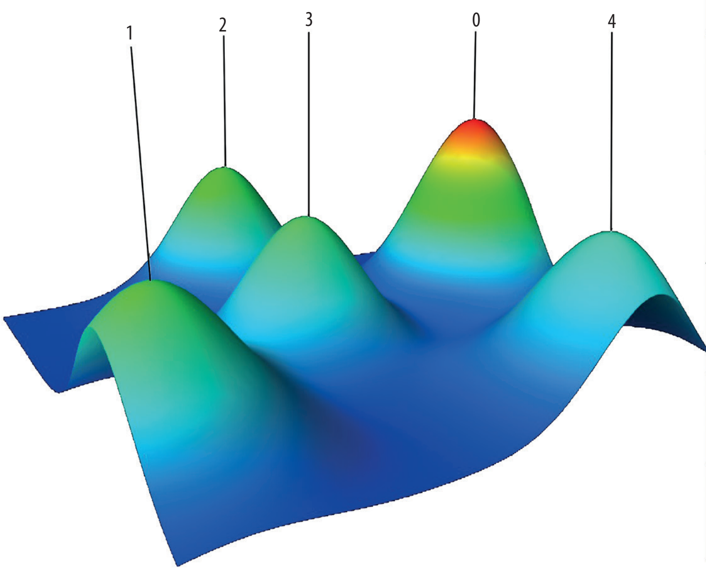
In addition, according to the connotation of the major MeSH terms and the semantic relationship between them, and the

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]



Cluster	MeSH terms
1	Animals, Vitamin D/metabolism, Vitamin D/pharmacology, Receptors Calcitriol/metabolism, Mice, Calcitriol/pharmacology, Receptors Calcitriol/genetics, Cholecalciferol/pharmacology, Rats, Vitamins/pharmacology, Cells Cultured, Calcitriol/analogs & derivatives, Disease Models, Animal, Cell Line Tumor, Cell Line, Calcitriol/administration & dosage, Calcitriol/therapeutic use, Calcium/metabolism
4	Risk Factors, Biomarkers/blood, Prospective Studies, Case-Control Studies, Vitamin D Deficiency/diagnosis, Follow-Up Studies, Retrospective Studies, Time Factors, Severity of Illness Index, Logistic Models, Prognosis, Polymorphism Single Nucleotide, Incidence, Odds Ratio, Linear Models, Multivariate Analysis
0	Humans, Female, Male, Vitamin D/blood, Vitamin D/analogs & derivatives, Middle Aged, Adult, Aged, Vitamin D Deficiency/blood, Young Adult, Adolescent, Vitamin D Deficiency/epidemiology, Cross-Sectional Studies, Vitamin D Deficiency/complications, Aged 80 and over, Prevalence, Body Mass Index, Parathyroid Hormone/blood, Calcifediol/blood, Cohort Studies, Calcium/blood, Age Factors, Bone Density, Sex Factors, Vitamin D Deficiency/physiopathology, Diabetes Mellitus Type 2/blood
3	Child, Pregnancy, Child Preschool, Seasons, Infant, Infant Newborn, Sunlight, Vitamin D Deficiency/prevention & control, Nutritional Status, Diet, Vitamin D, Surveys and Questionnaires, Vitamin D Deficiency/etiology, Ultraviolet Rays
2	Dietary Supplements, Vitamin D/administration & dosage, Vitamin D/therapeutic use, Vitamin D Deficiency/drug therapy, Treatment Outcome, Cholecalciferol/administration & dosage, Double-Blind Method, Vitamins/therapeutic use, Vitamins/administration & dosage, Dose-Response Relationship Drug, Cholecalciferol/therapeutic use, Vitamins/blood, Randomized, Controlled Trials as Topic, Vitamin D Deficiency/metabolism, Insulin Resistance, Bone Density/drug effects

Figure 2. Visualized matrix of bi-clustering of highly frequent major MeSH terms and PubMed Unique Identifiers (PMIDs) of articles on vitamin D. The rows represent the high-frequency major MeSH terms, listed on the right. The bottom of the matrix shows the PMID for each source article.



Cluster	MeSH terms
0	Humans, Female, Male, Vitamin D/blood, Vitamin D/analogs & derivatives, Middle Aged, Adult, Aged, Vitamin D Deficiency/blood, Young Adult, Adolescent, Vitamin D Deficiency/epidemiology, Cross-Sectional Studies, Vitamin D Deficiency/complications, Aged 80 and over, Prevalence, Body Mass Index, Parathyroid Hormone/blood, Calcifediol/blood, Cohort Studies, Calcium/blood, Age Factors, Bone Density, Sex Factors Vitamin D Deficiency/physiopathology, Diabetes Mellitus Type2/blood
1	Animals, Vitamin D/metabolism, Vitamin D/pharmacology, Receptors Calcitriol/metabolism, Mice, Calcitriol/pharmacology, Receptors Calcitriol/genetics, Cholecalciferol/pharmacology, Rats, Vitamins/pharmacology, Cells Cultured, Calcitriol/analogs & derivatives, Disease Models, Animal, Cell Line Tumor, Cell Line,
2	Calcitriol/administration & dosage, Calcitriol/therapeutic use, Calcium/metabolism Dietary Supplements, Vitamin D/administration & dosage, Vitamin D/therapeutic use, Vitamin D Deficiency/drug therapy, Treatment Outcome, Cholecalciferol/administration & dosage, Double-Blind Method, Vitamins/therapeutic use, Vitamins/administration & dosage, Dose-Response Relationship Drug, Cholecalciferol/therapeutic use, Vitamins/blood, Randomized, Controlled Trials as Topic, Vitamin D Deficiency/metabolism, Insulin Resistance, Bone Density/drug effects
3	Child, Pregnancy, Child Preschool, Seasons, Infant, Infant Newborn, Sunlight, Vitamin D Deficiency/prevention & control, Nutritional Status, Diet, Vitamin D, Surveys and Questionnaires, Vitamin D Deficiency/etiology, Ultraviolet Rays
4	Risk Factors, Biomarkers/blood, Prospective Studies, Case-Control Studies, Vitamin D Deficiency/diagnosis, Follow-Up Studies, Retrospective Studies, Time Factors, Severity of Illness Index, Logistic Models, Prognosis, Polymorphism Single Nucleotide, Incidence, Odds Ratio, Linear Models, Multivariate Analysis

Figure 3. Mountain visualization of bi-clustering of highly frequent major MeSH terms and articles on vitamin D. The 90 high-frequency terms, listed on the right, are clustered in peaks that represent 5 clusters numbered from 0 to 4.

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

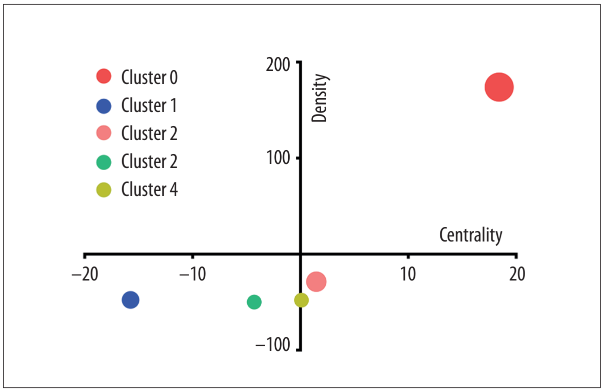


Figure 4. Strategic diagram of the clusters.

representative literature in each major cluster (Table 3), some major clusters were sub-divided into smaller topics. We divided the 90 high-frequency MeSH terms in the studies of vitamin D in the past 3 years into 5 clusters and further sub-divided them into the following 7 topics:

1. Epidemiological investigation of vitamin D deficiency in the population (cluster 0);
2. Epidemiological studies of vitamin D-related diseases (cluster 0);
3. Vitamin D metabolism, genetics, and pharmacological mechanisms (cluster 1);
4. Pharmacological and therapeutic effects of calcitriol and its analogs (cluster 1);
5. Vitamin D prevention- and treatment-related diseases, including vitamin D deficiency, diabetes, osteoporosis (cluster 2);
6. Reasons and prevention of vitamin D deficiency in the population (pregnant women, infants, children, etc.), including sunlight, diet (cluster 3);
7. Population blood vitamin D levels as biomarkers for diagnosing and predicting diseases (cluster 4).

We established topic strategic coordinates (Figure 4) based on the double-clustering results and the MeSH terms co-occurrence matrix. The horizontal axis of the strategic coordinates represents the centrality, and the vertical axis represents the density. Centrality measures the degree of closeness between each cluster of MeSH terms and other clusters of MeSH terms and expresses the degree of interaction between a subject area and other subject areas. Density is the degree of closeness of the MeSH terms within each cluster and indicates the ability of the class to maintain itself and develop itself. As show in Figure 4, cluster 0 is in the first quadrant; cluster 1 and cluster 3 are located in the third quadrant; cluster 2 and cluster 4 are in the fourth quadrant.

Distribution of vitamin D-related diseases

In addition to the above analysis, we retrieved 1,938 articles using "vitamin D" [major] AND ("2002/01/01" [PDAT]: "2005/01/01" [PDAT]), excluding literature reviews, comments, letters, guidelines, and conference discussions. Altogether, a total number of 707 articles of vitamin D-related diseases were screened. Similarly, we selected 2351 literature sources published from 2015 to 2018 reporting vitamin D-related diseases. We then classified the selected literature by disease types and attempted to explore the distribution of vitamin D-related diseases and the status of certain disease subgroups. Two researchers independently reviewed and evaluated the studies and reached consensus on the inclusion for analysis. The concordance rate between them was 0.90, indicating a strong agreement. With reference to the International Classification of Diseases (ICD-10), we classified the selected diseases into the following categories:

1. Endocrine and metabolic system diseases such as thyroid disease, diabetes mellitus, obesity, polycystic ovary syndrome;
2. Musculoskeletal diseases: osteoporosis, fractures, osteoarthritis, muscle strength, rickets, etc.;
3. Neoplasms such as breast neoplasms, colorectal neoplasms, skin neoplasms, prostate neoplasms, leukemia;
4. Neuropsychological diseases such as multiple sclerosis, Alzheimer's disease, cognitive decline, mood disorders;
5. Gestational diseases such as gestational diabetes, pregnancy-induced hypertension, pregnancy and offspring disease risk;
6. Infections such as non-specific infections, hepatitis B, hepatitis C, tuberculosis;
7. Urologic diseases such as uremia, acute kidney injury, benign prostatic hyperplasia;
8. Circulation system diseases such as myocardial infarction, coronary heart disease, hypertension, atherosclerosis, cerebral hemorrhage, cerebral infarction;
9. Digestive system diseases such as inflammatory bowel disease, pancreatitis, non-alcoholic fatty liver, liver cirrhosis;
10. Respiratory tract diseases such as asthma, bronchiectasis, sleep apnea-hypopnea syndrome, chronic obstructive pulmonary disease;
11. Skin diseases such as psoriasis, keratoses, atopic dermatitis, vitiligo, alopecia;
12. Other diseases.

As can be seen in Figure 5A, the cumulative proportion of the musculoskeletal diseases, neoplasms, skin diseases, endocrine and metabolic system diseases, and urologic diseases was 90.7%. The cumulative proportion of musculoskeletal diseases and neoplasms was 59.3%. As shown in Figure 5B, the cumulative number of the 11 listed diseases accounted for 90.6%. We found that in the category of endocrine and metabolic system

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

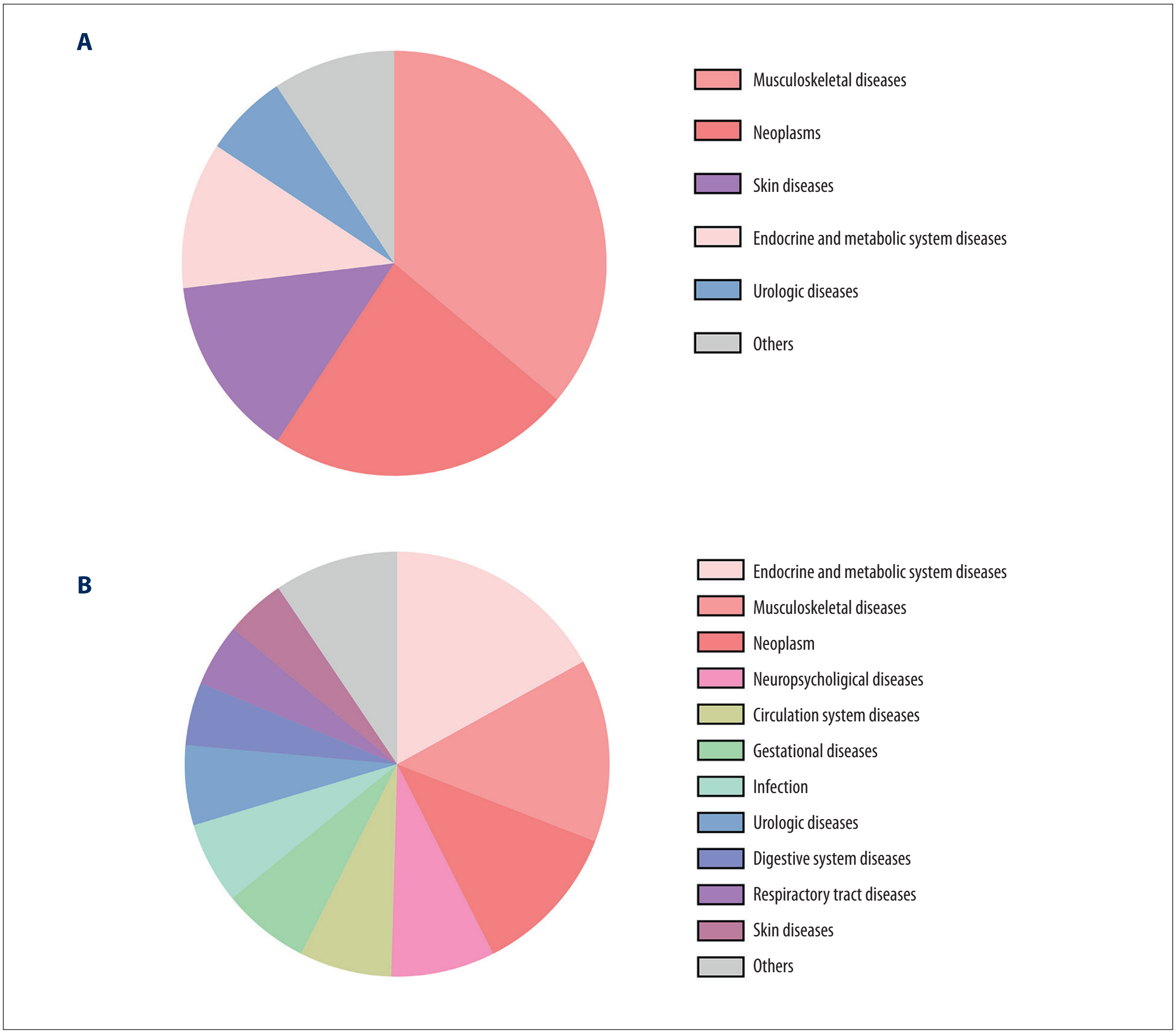


Figure 5. Vitamin D-related disease distribution from 2002 to 2005 (**A**) and from 2015 to 2018 (**B**). The numbers of publications are 707 (**A**) and 2351 (**B**), respectively.

diseases the following diseases were prevalent: diabetes mellitus, obesity, thyroid disease, and polycystic ovary syndrome. Their cumulative proportion was 92.1%, of which diabetes mellitus accounted for 50.5% (Figure 6A). The cumulative number of tumor subtypes listed in the neoplasm category accounted for 90.1% of the tumor-related literature (Figure 6B).

Discussion

In the past 10 years, an explosive growth has been observed in the volume of the research literature on vitamin D, and the number of journals has also grown rapidly (Figure 1), indicating that increasingly more researchers have begun to pay attention to and study vitamin D effects on humans in recent years.

To analyze the literature related to vitamin D that had been published in the past 3 years, we used BICOMB software to extract the MeSH terms-source article matrix and MeSH terms co-occurrence matrix. Additionally, we applied gCLUTO software to perform double-cluster analysis to obtain 5 major clusters, which were further sub-divided into 7 topics. Then, based on the double-clustering results and the MeSH terms co-occurrence matrix, topic strategic coordinates were established. As can be seen from Figure 4, cluster 0 is in the first quadrant and has high centrality and density, indicating the topics 1 and 2 are in the mature and core areas, and they are the most popular trends in vitamin D research. Cluster 1 and cluster 3 are located in the third quadrant, indicating the topics 3, 4, and 6 are located in the relatively cold marginal and immature regions.

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

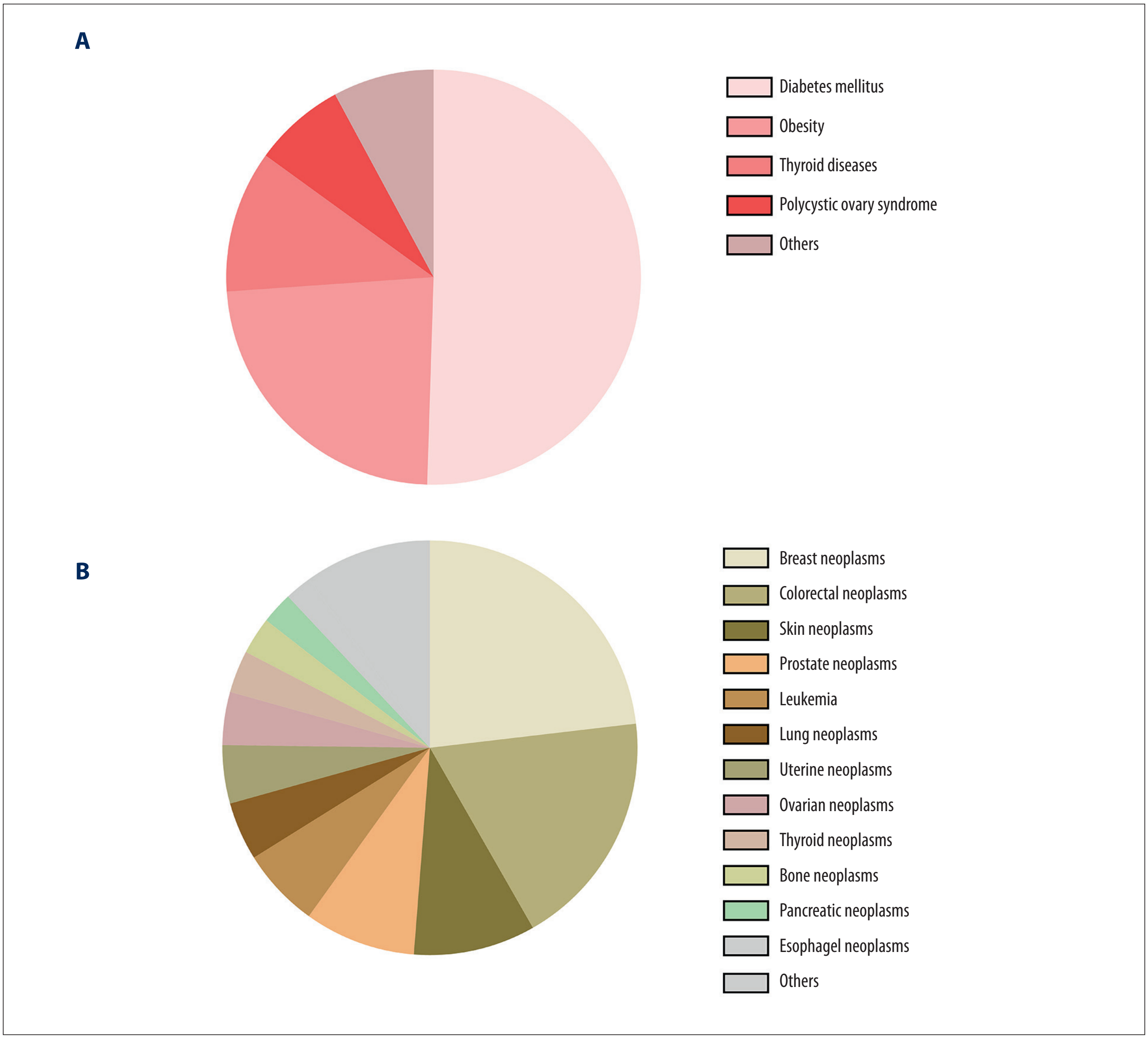


Figure 6. Vitamin D-related endocrine and metabolic system diseases (**A**) and neoplasms (**B**) distribution (2015–2018). The number of publications is 406 (**A**) and 242 (**B**), respectively.

Clusters 2 and 4 are in the fourth quadrant, and, although the 2 topics have low densities, they have high centrality, suggesting that the internal trends cannot be self-contained, in combination with other types of research topics. However, the subject of this study is still in its infancy and has huge potential for development and substantial importance in research on vitamin D. Among the 2 topics mentioned above, cluster 2 has greater centrality than cluster 4, indicating that the former has a closer interaction with other research topics and has a greater potential for research.

We analyzed the distribution of vitamin D-related diseases from 2002 to 2005 and from 2015 to 2018. As shown in Figure 5, a relative reduction from 36.1% to 14.0% was observed for this period in the number of studies on musculoskeletal diseases. The subject of vitamin D-related diseases has undergone drastic changes and has changed from effects on skeletal to non-skeletal tissues over the last decade. Over the past 3 years, the cumulative proportion of the endocrine and metabolic system diseases, musculoskeletal diseases, neoplasms, and neuropsychological diseases is 50.4% and has become a subject of increased research interest for investigations of vitamin D-related diseases, suggesting that these 4 types of diseases are more closely associated with vitamin D. In the category of endocrine and metabolic system diseases, the proportion of diabetes mellitus was 50.5% (Figure 6), and in all vitamin D-related diseases, the frequency of diabetes mellitus is second only to bone diseases, indicating that diabetes mellitus has a great potential correlation with vitamin D. In the neoplasm category, the cumulative proportion of breast, colorectal, and

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]

skin neoplasms was 51.2%, meaning it had become a popular subject of studies in the field of vitamin D-related neoplasms. As show in Figure 5, the range of vitamin D-related diseases significantly increased between 2015 and 2018, and neuropsychological, circulation system, and gestational diseases, as well as infections and respiratory tract and digestive system diseases, have become new research trends.

The hot and the unpopular topics are constantly evolving. As the epidemiological studies of "vitamin D deficiency" and " association of vitamin D with various diseases" have increased dramatically and the research in this area has matured, use of vitamin D for prevention or treatment might have increasingly started to attract the attention of researchers. According to the distribution of the vitamin D-related diseases and strategic coordinates established in our examination, we speculate that vitamin D has important potential research value in the prevention and treatment of some diseases, especially diabetes mellitus. In addition, with the increasing body of research on the correlation between vitamin D and diseases, and the prevention or treatment of diseases by vitamin D administration, research on the molecular mechanisms and signaling pathways of vitamin D may gradually increase and deepen; related reviews have already appeared [27–29]. In a recent study, scientists found a new mechanism of interaction between vitamin D receptors and chromatin-associated proteins [38], which may lead to more explorations of the molecular mechanisms of vitamin D action.

Due to limitations in research methods, this study might have overlooked some of the emerging fields of research with high potential, such as molecular pathological epidemiology (MPE). MPE is a discipline combining epidemiology and pathology, whose approach associates potential risk factors with molecular disease pathology [39]. A major value of MPE resides in the better definition of phenotype that it provides, which can improve our understanding of disease etiology from host susceptibility and exposures [40]. MPE will not only enhance our understanding of the pathophysiological roles of the interactions among the environment, host, and tumor, but also promote the development of preventive and therapeutic strategies for precision medicine [41]. The approach of MPE will possibly play an important role in vitamin D-related studies.

The co-word clustering analysis of high-frequency MeSH terms is a new method of analysis, and there still may be some degree of prejudice that prevents researchers from choosing vocabulary when writing. Due to the smart limitations in the PubMed database retrieval system and the absence of other database retrieval results, the data set in the present study may be incomplete. In addition, the quality of the articles in the PubMed database is not uniform and some deviations might exist in the results of the research.

Conclusions

In this study, we summarized 5 categories and 7 popular trends for vitamin D research. We found that the population epidemiology study of vitamin D deficiency and vitamin D-related diseases has attracted extensive and active research interest. The use of vitamin D in the prevention and treatment of some diseases, especially diabetes, was found to have a significant potential future research value.

Acknowledgements

The authors thank Professors Lei Cui, Han Zhang, Lei Yan, Yuefang Hou, and Xiaoning Wang for providing the instructions for use of BICOMB and gCLUTO software and the method for analysis.

Conflict of interest

None.

References:

1. Kim J, Lee Y, Kye S et al: Association of serum vitamin D with osteosarcopenic obesity: Korea National Health and Nutrition Examination Survey 2008–2010. J Cachexia Sarcopenia Muscle, 2017; 8: 259–66
2. Pereira-Santos M, Costa PR, Assis AM et al: Obesity and vitamin D deficiency: A systematic review and meta-analysis. Obes Rev, 2015; 16: 341–49
3. Afzal S, Brondum-Jacobsen P, Bojesen SE, Nordestgaard BG: Vitamin D concentration, obesity, and risk of diabetes: A mendelian randomisation study. Lancet Diabetes Endocrinol, 2014; 2: 298–306
4. Wang J, Lv S, Chen G et al: Meta-analysis of the association between vitamin D and autoimmune thyroid disease. Nutrients, 2015; 7: 2485–98
5. Robinson-Cohen C, Hoofnagle AN, Ix JH et al: Racial differences in the association of serum 25-hydroxyvitamin D concentration with coronary heart disease events. JAMA, 2013; 310: 179–88
6. Brondum-Jacobsen P, Nordestgaard BG, Schnohr P, Benn M: 25-hydroxyvitamin D and symptomatic ischemic stroke: an original study and meta-analysis. Ann Neurol, 2013; 73: 38–47
7. Martineau AR, James WY, Hooper RL et al: Vitamin D3 supplementation in patients with chronic obstructive pulmonary disease (ViDiCO): A multicentre, double-blind, randomised controlled trial. Lancet Respir Med, 2015; 3: 120–30
8. Hollams EM, Teo SM, Kusel M et al: Vitamin D over the first decade and susceptibility to childhood allergy and asthma. J Allergy Clin Immunol, 2017; 139: 472–81.e9
9. Budhathoki S, Hidaka A, Yamaji T et al: Plasma 25-hydroxyvitamin D concentration and subsequent risk of total and site specific cancers in Japanese population: Large case-cohort study within Japan Public Health Center-based Prospective Study cohort. BMJ, 2018; 360: k671
10. Fuchs MA, Yuan C, Sato K et al: Predicted vitamin D status and colon cancer recurrence and mortality in CALGB 89803 (Alliance). Ann Oncol, 2017; 28: 1359–67

11. Feart C, Helmer C, Merle B et al: Associations of lower vitamin D concentrations with cognitive decline and long-term risk of dementia and Alzheimer's disease in older adults. Alzheimers Dement, 2017; 13: 1207–16
12. Miller JW, Harvey DJ, Beckett LA et al: Vitamin D status and rates of cognitive decline in a multiethnic cohort of older adults. JAMA Neurol, 2015; 72: 1295–303
13. Aghajafari F, Nagulesapillai T, Ronksley PE et al: Association between maternal serum 25-hydroxyvitamin D level and pregnancy and neonatal outcomes: Systematic review and meta-analysis of observational studies. BMJ, 2013; 346: f1169
14. Trepo E, Ouziel R, Pradat P et al: Marked 25-hydroxyvitamin D deficiency is associated with poor prognosis in patients with alcoholic liver disease. J Hepatol, 2013; 59: 344–50
15. Ananthakrishnan AN, Khalili H, Higuchi LM et al: Higher predicted vitamin D status is associated with reduced risk of Crohn's disease. Gastroenterology, 2012; 142: 482–89
16. Aksu Cerman A, Sarikaya Solak S, Kivanc Altunay I: Vitamin D deficiency in alopecia areata. Br J Dermatol, 2014; 170: 1299–304
17. Upala S, Sanguankeo A: Low 25-hydroxyvitamin D levels are associated with vitiligo: A systematic review and meta-analysis. Photodermatol Photoimmunol Photomed, 2016; 32: 181–90
18. McDonnell SL, Baggerly C, French CB et al: Serum 25-hydroxyvitamin D concentrations >/=40 ng/ml are associated with >65% lower cancer risk: Pooled analysis of randomized trial and prospective cohort study. PLoS One, 2016; 11: e0152441
19. McDonnell SL, Baggerly CA, French CB et al: Breast cancer risk markedly lower with serum 25-hydroxyvitamin D concentrations ≥60 *vs.* <20 ng/ml (150 *vs.* 50 nmol/L): Pooled analysis of two randomized trials and a prospective cohort. PLoS One, 2018; 13(6): e0199265
20. Barry EL, Peacock JL, Rees JR et al: Vitamin D receptor genotype, vitamin D3 supplementation, and risk of colorectal adenomas: A randomized clinical trial. JAMA Oncol, 2017; 3: 628–35
21. Martineau AR, Jolliffe DA, Hooper RL et al: Vitamin D supplementation to prevent acute respiratory tract infections: Systematic review and meta-analysis of individual participant data. BMJ, 2017; 356: i6583
22. Hibbs AM, Ross K, Kerns LA et al: Effect of vitamin D supplementation on recurrent wheezing in black infants who were born preterm: The D-wheeze randomized clinical trial. JAMA, 2018; 319: 2086–94
23. Gulseth HL, Wium C, Angel K et al: Effects of vitamin D supplementation on insulin sensitivity and insulin secretion in subjects with type 2 diabetes and vitamin D deficiency: A randomized controlled trial. Diabetes Care, 2017; 40: 872–78
24. Sadiya A, Ahmed SM, Carlsson M et al: Vitamin D3 supplementation and body composition in persons with obesity and type 2 diabetes in the UAE: A randomized controlled double-blinded clinical trial. Clin Nutr, 2016; 35: 77–82
25. Lappe J, Watson P, Travers-Gustafson D et al: Effect of vitamin D and calcium supplementation on cancer incidence in older women: A randomized clinical trial. JAMA, 2017; 317: 1234–43
26. Mason C, Xiao L, Imayama I et al: Vitamin D(3) supplementation during weight loss: A double-blind randomized controlled trial. Am J Clin Nutr, 2014; 99: 1015–25
27. Christakos S, Dhawan P, Verstuyf A et al: Vitamin D: Metabolism, molecular mechanism of action, and pleiotropic effects. Physiol Rev, 2016; 96: 365–408
28. Girgis CM, Clifton-Bligh RJ, Hamrick MW et al: The roles of vitamin D in skeletal muscle: Form, function, and metabolism. Endocr Rev, 2013; 34: 33–83
29. Bouillon R, Carmeliet G, Lieben L et al: Vitamin D and energy homeostasis: Of mice and men. Nat Rev Endocrinol, 2014; 10: 79–87
30. Reid IR, Bolland MJ, Grey A: Effects of vitamin D supplements on bone mineral density: A systematic review and meta-analysis. Lancet, 2014; 383: 146–55
31. Law J, Bauin S, Courtial J-P, Whittaker J: Policy and the mapping of scientific change: A co-word analysis of research into environmental acidification. Scientometrics, 1988; 14: 251–64
32. Hartigan JA: Direct clustering of a data matrix. J Am Stat Assoc, 1972; 67: 123–29
33. Cheng Y, Church GM: Biclustering of expression data. Proc Int Conf Intell Syst Mol Biol, 2000; 8: 93–103
34. Pontes B, Giraldez R, Aguilar-Ruiz JS: Biclustering on expression data: A review. J Biomed Inform, 2015; 57: 163–80
35. Cui L: Development of a text mining system based on the co-occurrence of bibliographic items in literature databases. New Technol Libr Inf Serv, 2008; 70–75
36. Karypis Lab. gCLUTO-Graphical Clustering Toolkit. 2003 Nov 27 [cited 2018 May 30]. Available from: URL: *http://glaros.dtc.umn.edu/gkhome/cluto/gcluto/download*
37. Li F, Li M, Guan P et al: Mapping publication trends and identifying hot spots of research on internet health information seeking behavior: A quantitative and co-word biclustering analysis. J Med Internet Res, 2015; 17(3): e81
38. Wei Z, Yoshihara E, He N et al: Vitamin D switches BAF complexes to protect beta cells. Cell, 2018; 173: 1135–49.e15
39. Ogino S, Nowak JA, Hamada T et al: Insights into pathogenic interactions among environment, host, and tumor at the crossroads of molecular pathology and epidemiology. Annu Rev Pathol, 2018 [Epub ahead of print]
40. Ogino S, Nowak JA, Hamada T et al: Integrative analysis of exogenous, endogenous, tumour and immune factors for precision medicine. Gut, 2018; 67: 1168–80
41. Ogino S, Nishihara R, VanderWeele TJ et al: Review article: The role of molecular pathological epidemiology in the study of neoplastic and non-neoplastic diseases in the era of precision medicine. Epidemiology, 2016; 27: 602–11

Indexed in: [Current Contents/Clinical Medicine] [SCI Expanded] [ISI Alerting System] [ISI Journals Master List] [Index Medicus/MEDLINE] [EMBASE/Excerpta Medica] [Chemical Abstracts/CAS]