Time-Tracking-based Research of Drug Development in Pharmaceutical Fields ——A Case Study on Metformin

Shengnan Wu (vivian_sxmu@163.com)
Shanxi Medical University School of Management
Jizong Dong
Shanxi Medical University School of Management
Yidan Sun
Shanxi Medical University School of Management
Jian Gao
Shanxi Medical University School of Management
Ruonan Tian
Shanxi Medical University School of Humanities and Social Sciences
Qi Yu
Shanxi Medical University School of Management

Research Article

Keywords: metformin, correspondence analysis, research overview, drug reposition, AMPK

Posted Date: June 14th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1743720/v1

Abstract

[Purpose/Significance]: At present, the bibliometric analysis methods used in the field of medicine are difficult to dig deeply into the text content, and most of them still stay in the stage of macroscopic analysis of the text. This paper adopts an analysis method of time-tracking, taking the text of metformin drugs as an example, through further mining of the text content, which is helpful to deeply explore the life cycle of drugs, research overview in different periods, changes in drug therapeutic uses, research hotspots, and explore the reuse potential of future drugs in new fields.

[Method/Process]: In view of the shortcomings of the current research methods of text mining in the field of medicine, we extracted the inspection labels, subtitles and indexing terms of all metformin drugs texts included in the Pubmed database from 1959 to 2020. Using the correspondence analysis, cluster analysis and statistical methods to conduct dimensionality reduction of the data, and explore the information behind the text by a more microscopic way.

[Result]: The paper revealed the characteristics and reasons of metformin drugs in the process of entire historical development. It was found that metformin has drug reposition value in the treatment of obesity, cardio-cerebrovascular diseases, polycystic ovary syndrome, anticancer, anti-inflammatory, altering gut microbiome composition and anti-aging, etc. And it predicted that metformin may have the drug reuse potential in the treatment of Alzheimer's disease, neurodegenerative diseases and androdegenerative alopecia, etc. At the same time, the study of AMPK hypoglycemic mechanism will become the focus and hotspot of other hypoglycemic drug research and development institute.

[Conclusion]: The experimental results prove that the analysis method of time-tracking can effectively test the drug research overview.

1 Introduction

Because of the particularity of the pharmaceutical industry, the medical data resources present the characteristics of incompleteness, non-standardization, redundancy and high value. Therefore, it is of great significance to extract relevant data information from massive medical data for in-depth mining, and use time-tracking methods to explore the research profile of drugs that have been successfully listed, which is of great significance for studying the life cycle of drugs, research overview in different periods, research hotspots, and exploring potential indications, adverse reactions and drug repositioning. In the selection of data, we chose the medical database Pubmed, which ensures the comprehensive coverage of drug text, and its indexing terms are directly extracted from the thesaurus, avoiding the subjective problem of manual screening indexing terms. The main research trends of drugs can conduct to analysis through the indexing terms contained in the literature, which are specified by professional indexers and have undergone a certain degree of professionalism from the application of applications initially in experimental studies to publication in medical journals, and has a certain speciality. Indexing terms are an important basis for providing information on the evolution of drug research profiles in subsequent studies.

The chosen study subjects are metformin drugs. The discovery of metformin has a very long history. In the Middle Ages, it was found that Galega, which had the effect of relieving polyuria and reducing urine glucose in diabetic patients. Between 1920 and 1950, many guanidines derivatives were synthesized by degrees, and developed phenformin, buformin and metformin. In 1957, with the first use of metformin in clinical, the development process

of metformin finally turned a new page. Due to its long history, this drug has a chance to be chosen as a research subject, and its research trends and therapeutic uses have changed significantly in those years. Because Pubmed's first metformin text was published in 1959, we selected of all metformin drugs texts included in the Pubmed database from 1959 to 2020 as data sources.

The paper purposes that take the metformin drugs text as an example, through the further excavation of the text content, to explore the life cycle of drugs, research overview in different periods, research hotspots, and the prospects for future research trends, and to prove the effectiveness of time-tracking analysis methods for drug research overview exploration.

Through analyzing the relationship between adverb terms and chronology, this paper explores the research trend of metformin. Firstly, a correspondence analysis was used to detail the research trend plot of metformin over the past 62 years. Secondly, starting from the year predictions on the factor axis obtained from the corresponding analysis, the characteristics exhibited by the different periods of drug evolution are determined by cluster analysis. Finally, the research overview for different time periods is described according to their corresponding indexing terms.

Currently, due to most metformin texts are deep studies of molecular mechanisms and clinical applications, most review articles not only focus on specific aspects, but also on relatively limited aspects on metformin, such as pharmacological activity or pharmacokinetic profiles. Therefore, more and more people are struggling to conduct new research due to insufficient integration of data. Bibliometric analysis allows us to obtain an overview of the state of scientific research in a particular discipline, field or specialty. Long-term research is particularly important for understanding the dynamics of the discipline and determining which topics are emerging and which are tending to decline. In addition, it is of great significance for assessing the current research status and hotspots and the outlook for future research trends. At the same time, on the basis of bibliometric analysis, using the correspondence analysis, cluster analysis and statistical analysis, which can explore the life cycle of metformin from a more microscopic perspective, and excavate the main achievements and research hotspots and trends in different periods.

2 Materials And Methods

A total of 23,513 texts on metformin drugs from 1959 to 2020 were included in the pubmed database for data source selection.The inspection labels, subheadings, and index terms of metformin texts were extracted, and a total of 25,698 inspection labels, 33 subheadings, and 35,819 index terms were obtained.

We mainly used correspondence analysis, cluster analysis, and statistical analysis to study the data.

Correspondence analysis is to use the idea of dimensionality reduction to achieve the purpose of simplifying the data structure, and at the same time, the rows and columns in the data table are processed to seek the relationship between rows and columns in the data table represented by low-dimensional graphs.Correspondence analysis is widely used to study the data of contingency table composed of attribute variables. Correspondence analysis can simultaneously draw the situation of different values of attribute variables on a two-dimensional diagram. Each column of each row of contingency table is expressed by a point on the two-dimensional diagram, and the relationship between various states of attribute variables and the relationship between different attribute variables is described in an intuitive and concise form.When determining the number of dimensions, usually the first 2–3 dimensions have been able to explain most of the original information, each dimension has the corresponding

singular value, the square of the singular value is the inertia, the inertia reflects the degree that the results of each dimension can explain the connection between the two variables, and the sum of inertia in all dimensions can still indicate the size of the total information.The proportion of inertia is the proportion of variance explanation, and the proportion of inertia in each dimension to the sum of inertia reflects the amount of information carried by each dimension.The dimension score in the line point overview gives the score of each category in each dimension (two dimensions are extracted by default).The numerical results show the contribution of this class to inertia.The position and distance of the scatter points in the correspondence analysis diagram in space (this distance is by default the chi-square distance, that is, the weighted Euclidean geometric distance) reflect the relationship between various types.The closer the distance between the scatter points, the correlation.In the same dimension, if different categories of the same variable are close in a certain direction, it indicates that these categories are not very different in this dimension.The classification points of different variables within approximately the same region in the same orientation (represented as the scatter points with different colors in Fig).The closer or farther the distance between scatter points, the more obvious the association.

The basic idea of cluster analysis is to cluster variables that are close in distance and variables that are distant in distance, and the process continues until each variable is clustered into the appropriate class.We choose year as the variable of clustering, and then analyze the subheadings contained in each cluster, so as to reflect the research hotspots and trends of different periods more microscopically, as well as excavate the achievements and breakthroughs made in each period.

Statistical analysis is mainly achieved by means of Tableau plotting tools for data visualization.Statistical analysis is mainly used for the statistical analysis of metformin document quantity statistics, test label statistics, index terms, keywords, and at the same time, through the further labeling and processing of graphics, it assists in mining the information contained behind the images.

3 Results

We plotted the annual publication volume of a total of 23513 literatures from 1959-2020 with Tableau plotting tool (as shown in Figure 1). From the figure the following findings have been drawn.

1.Before the 1990s, it showed a low volume of published literature, a slow growth rate, and a relatively flat trend of literature growth.

2.According to the dashed line of the literature publication trend and the axis of the mean value, it can be seen that 2000 was a turning point of literature growth. From 2000 to the present, there has been an explosive growth in the number of literature with an increasing growth rate. Especially, an exponential growth trend was shown after 2010.

3.1 Subheading analysis

We statistically counted all metformin-related subtitles from 1959-2020 and found that the five subtitles with the highest percentage were TU (therapeutic use), PD (pharmacology), AE (adverse effect), AD (administration and dose), and AA (analogues and derivatives). Their percentages at each time period are shown in Figure 2. We observed that TU (therapeutic use) and PD (pharmacology) both showed an overall trend of decreasing and then increasing, while AE (adverse effect) showed a trend of increasing, then decreasing and then increasing. For AD (administration and dose) and AA (analogues and derivatives), although there were also some variation trends in each period, the overall trend was relatively flat.

3.2 Statistical analysis for inspection tag

Based on a time period of 5 years, we calculated the percentage of the frequency of the top 9 test labels (The top 9 test labels were taken in consideration of the fact that the percentage of other labels in the total literature frequency was less than 1% for most time periods and did not show significant trends and characteristics) in the frequency of total literature test labels for each 5-year period with Tableau plotting tool (as shown in Figure 3). An analysis of the test labels assigned to the literatures provided information about the general aspects of the studies over 60 years. (Data for 1959-1960 were excluded because there were only three papers in total for these two years, which are not representative; if included, they might generate some noise and have some impact on the overall model)

1. The label “Human” has always been dominant but the percentage of it has been showing a decreasing trend from 1961 to 1978 and decreased from 20% to 12.65%, and gradually showed an overall increasing trend only since 1978.

2. The percentages of the labels “Male”, “Female”, “Middle Aged”, “Adult” and “Aged” all showed an increasing trend from 1961 to 1972, but all showed a decreasing trend during 1972-1978.

3. The percentage of “Male” label was always higher than that of “Female” label during 1972-1997, while the percentage of “Female” label was always higher than that of “Male” label during 1997-2020.

4. The label “Clinical” has shown an overall increasing trend after 2000 compared with that before 2000.

3.3 Correspondence analysis of year and subtitle

We used the statistical analysis software SPSS Statistics 26 to perform a corresponding analysis of the subtitles of metformin, aiming to explore the development pattern of metformin more further. In SPSS, the default data requires each row to be a single case, and since the two-dimensional table is a cross-summary of two categorical variables, after converting it to a one-dimensional table, each row of data would still correspond to the number of cases summarized from the categorical variables. Therefore, we have to perform case weighting for the frequency of the subtitles first. We selected the first two dimensions obtained from the analysis (explaining a total of 63.8% of the inertia, as shown in Figure 4). It reflects the relationship between years and subtitles. The first dimension explains 51.9% of the inertia and the second dimension explains 11.9% of the inertia. The red circles in the figure denote years, the blue circles denote subtitle names, and the horizontal and vertical coordinates of each point denote its score on each dimension (i.e., contribution of the dimension). The chi-square distance was used for the measurement of the distance between two years or two subtitles. The Euclidean distance between their outlines was used for the measurement of the distance between two years, and the inverse of the quality was used to weight each component of the index term (or relative frequency). In the same dimension, the closer the category distances of the same variable, the higher their correlation. From the figure, we can see that the nodes of years and subtitles are almost distributed at two sides of the function Y=X, and most of the nodes distributed above Y=X are year nodes, while most of the nodes distributed below Y=X are subtitle nodes. The closer the total distance (i.e., the minimum straight-line distance) of a year node to Y=X and a subtitle node to Y = X, the higher the correlation of the two nodes. Among them, the 6 subtitles of pharmacokinetics, surgery, therapy, pharmacology, physiology, and therapeutic use are the nodes closest to Y=X, while the year nodes of 1960, 1961, 2005, 2006, 2009-2013, and 2015-2020 are the nodes closest to Y=X. By comparing them collectively, it can be found that the closest year nodes to the subtitle pharmacokinetics are 1960, 1961, 2005, and 2006; the closest year nodes to the subtitle

therapy and surgery are 2009-2013; and the closest year nodes to the subtitle pharmacology, physiology, and therapeutic use are 2015-2020, indicating a strong correlation between these subtitles and years. For further presenting the correlation between years and the correlation between subtitles, we also performed corresponding analysis for years and subtitles respectively (as shown in Figure 5 and Figure 6), and cluster analysis for years (as shown in Figure 7). From Figure 4, it can be observed that the year nodes before the 21st century are generally distributed in the negative half-axis of the second dimensional factor axis (i.e., below the green solid line), and the nodes are distributed in a relatively scattered manner. Comparatively, the year nodes after the 21st century are distributed more concentrated and generally have a higher contribution to the model compared to the year nodes before the 21st century, indicating that the 21st century is an important turning point for the research on metformin. From Figure 5, it can be seen that the four subtitle nodes of rehabilitation, transmission, pathogenicity, and urine are distributed relatively marginally on the model, but except for the urine node, which has a low contribution in both dimensions, the transmission, pathogenicity, and rehabilitation nodes shows high contributions in both the first and second dimensions, indicating that these subtitle nodes are all high-quality nodes although they exhibit some differences from the other subtitle nodes. Considering keeping the amount of inertia carried by each historical period to be as relatively even as possible, we used the clustering tool gCLUTO to perform cluster analysis for all time nodes from 1959-2020 (as shown in Figure 6), and eventually limited the number of clusters to 5, i.e., 1959-1998, 1999-2005, 2006-2009, 2010-2013, and 2014-2020 (The larger span of the first cluster 1959-1998 is due to the relatively few publications before the 21st century, and further clustering would result in clusters carrying even less inertia). According to the figure of the cluster peaks, it can be noticed that the peak volume of the cluster 1959-1998 is rather small and the color variation within this cluster is large, suggesting that the information inertia carried in this period is rather small, which also indicates that the research on metformin in this period was still in the initial stage, and there were large variations among the research themes. In comparison with the cluster 1959-1998, the peak volume of the 1999-2005 cluster has increased, and the color variation within the cluster shows that there has been stronger linkage among research themes in this period. And the peak volumes of the three clusters of 2006-2009, 2010-2013 and 2014-2020 have significantly increased compared with the previous two clusters, and the colors within the clusters are more uniform, illustrating that the research on metformin has significantly increased in recent years with a continuous expansion in the depth and breadth of the research content.

Next we counted the top five subheadings in frequency ranking under each cluster (i.e., each historical period), and the percentage of these subheadings occupied internally and globally (as shown in Table 1). Through Table 1 we found that the studies of metformin in 1959-1998 mainly focused on several aspects: UR (urine), PO (toxicity), NU (nursing), PHY (physiology), TU (therapeutic use). The 1999-2005 study focused on several aspects: RE (rehabilitation), PHY (pathophysiology), VE (veterinary), PO (toxicity), and ST (standard). From 2006 to 2009, research focused on RE (rehabilitation), PO (toxicity), VE (veterinary), PHY (pathophysiology), and PHA (kinetics). The 2010-2013 study focused on several aspects: NU (care), VI (virology), ST (standard), MO (mortality), and OA (organization and management). The 2014-2020 study focused on several aspects: TR (transmission), PA (pathogenicity), PARA (parasitology), TRAN (transplantation), MI (microbiology).

Table 1 Cluster analysis for 1959-2020 years

Description of five research periods based on the study of Metformin subheadings				Prob.
Clusters/Subheadings	%Global	Internal freq.	Global freq	
Cluster 1. Years 1959-1998				
urine	10.61	19	179	0
poisoning	9.28	9	97	0.002
nursing	6.25	1	16	0.004
physiology	5.59	814	14571	0
therapeutic use	5.49	852	15523	0
Cluster 2. Years 1999-2005				
rehabilitation	20.51	8	39	0.006
physiopathology	15.95	351	2200	0
veterinary	14.81	4	27	0.007
poisoning	14.43	14	97	0.004
standards	12.67	28	221	0
Cluster 3. Years 2006-2009				
rehabilitation	30.77	12	39	0.002
poisoning	22.68	22	97	0.004
veterinary	22.22	6	27	0.01
physiopathology	16.27	358	2200	0
pharmacokinetics	15.06	155	1029	0
Cluster 4. Years 2010-2013				
nursing	43.75	7	16	0.008
virology	30.00	21	70	0.003
standards	27.15	60	221	0
mortality	23.01	180	782	0
organization and administration	22.71	186	819	0
Cluster 5. Years 2014-2020				
transmission	100	4	4	0.007
pathogenicity	96.67	29	30	0.01
parasitology	91.67	11	12	0.04
transplantation	83.33	10	12	0.006
microbiology	82.50	198	240	0

3.4 Correspondence analysis between indexing term joint subtitle and year

Analyzing indexing terms in association with their subtopics can reveal more concrete information about textual implication than using indexing terms alone. We conducted a year-by-year statistical analysis of all indexing terms for the 62 years from 1959 to 2020 and discovered that there were 219 commonly used indexing terms with 35,819 occurrences. The frequency distribution of indexing terms is highly skewed, with 61% of indexing terms occurring only once (incidental words). To avoid excessive noise, we did not take into account incidental words that appear only once in our analysis. To avoid excessive noise, we did not take into account incidental words that appear only once in our analysis. We analyzed the corresponding 219 indexing terms, and the first two dimensions explained a total of 82.1% of the inertia (as presented in Figure 8). Since the quantity of indexing terms is exceedingly large, we only show the top 30 indexing terms that contribute in the first two dimensions in order to avoid the overlapping display of indexing terms. Meanwhile, as the five indexing terms, namely Metformin TU, Hypoglycemic-Agents TU, Insulin TU, Diabetes-Mellitus-Type2 DT, and Metformin PD, accounted for a relatively high percentage of the global text (as presented in Table 2), it might result in the analysis of other indexing terms being affected by extreme values. And all five indexing terms characterized the role of metformin with respect to hypoglycemia. This application has been studied throughout the development of metformin and has been generally recognized. Our primary goal was to explore the application of metformin in fields other than glucose-lowering, so we excluded these 5 indexing terms from the corresponding analysis and ultimately manifested only 25 indexing terms. In Figure 7 we can observe the relationship between the joint subtitle of the indexing terms and the year.

We can see that the first dimension interpreted a total of 54.9% of the inertia while the second dimension interpreted a total of 28.2% of the inertia. In a chronological order from far to near we discovered that the Anticholesteremic-Agents PD node is closer to the year node of 1970-1980, indicating a higher concentration of pharmacological studies on acidosis with biguanides carried out during this period; the Phenformin AE node is relatively close to the year node of 1976-1980, illustrating that the research on the adverse effects of phenformin was more concentrated in that period; the Obesity DT node is closer to the year node of 1980-1990, indicating that the research on metformin in the treatment of obesity has made some advancements in this period; the nodes of Myocardial-Infarction DT, Arteriosclerosis DT, Thromboembolism PC, and Coronary-Disease DT related to cardio-cerebrovascular disease are closer to the year nodes of 1997-2000, suggesting new advances regarding metformin in the treatment of cardio-cerebrovascular disease in that period; Polycystic-Ovary-Syndrome DT node is closer to the year node of 2000-2005, indicating new research on metformin in the treatment of polycystic ovary syndrome in that period; Neoplasms DT, Gastrointestinal-Microbiome DE are closer to the year nodes of 2008-2014, demonstrating new advancements in research on metformin in anticancer and gut microbiome composition in that period; Thiazolidinediones TU, Sitagliptin-Phosphate AD, Linagliptin AD, Canagliflozin AD, Exenatide TU, Liraglutide TU, and Glipizide AD are closer to the year nodes after 2000, with this performance being more evident especially in 2015-2020. On the basis of observations, we found that these nodes have a remarkable commonality with metformin in that both of them are hypoglycemic agents. The research focused on therapeutic use and administration and dosage, which could be substitutes for metformin. In this regard, can we raise the question that will these new glucose-lowering drugs have an impact on the position of metformin in the field of glucose-lowering? In the conclusion part we will elaborate on this issue. Anti-Inflammatory Agents PD node is closer to the year node of 2000-2020, demonstrating the advancements in pharmacological studies on metformin in anti-inflammatory aspects in that period; the Longevity DE node is closer to the year node of 2015-2020, indicating advancements on the application of metformin in delaying aging; the AMP-Activated-Protein-Kinases ME nodes are

all close to the 2008-2020 year nodes, indicating that the hypoglycemic mechanism about AMPK from 2008 to the present is the highlight and hot spot of the current and even future time research.

Table 2 Frequency and percentage of the top 20 indexed term joint subtitles in all metformin literatures

Most frequent indexing terms in Metformin documents

Descriptors	Freq.	% Doc.
Metformin TU	2617	7.31%
Hypoglycemic-Agents TU	2194	6.13%
Insulin TU	1936	5.40%
Diabetes-Mellitus-Type2 DT	1717	4.79%
Metformin PD	1494	4.17%
AMP-Activated-Protein-Kinases ME	693	1.93%
Neoplasms DT	576	1.61%
Polycystic-Ovary-Syndrome DT	504	1.41%
Gastrointestinal-Microbiome DE	467	1.30%
Acidosis PD	428	1.19%
Obesity DT	404	1.13%
Coronary-Disease DT	393	1.10%
Myocardial-Infarction DT	361	1.01%
Thromboembolism PC	348	0.97%
Arteriosclerosis DT	316	0.88%
Anticholesteremic-Agents BL	240	0.67%
Anticarcinogenic-Agents TU	236	0.66%
Diabetic-Retinopathy DT	228	0.64%
Intestinal-Absorption DE	212	0.59%
Longevity DE	207	0.58%
Thiazolidinediones TU	192	0.54%
Sitagliptin-Phosphate AD	183	0.51%
Canagliflozin AD	176	0.49%
Linagliptin AD	172	0.48%
Liraglutide TU	163	0.46%
Exenatide TU	154	0.43%
Acarbose AD	153	0.43%
Phenformin AE	142	0.40%
Glipizide AD	140	0.39%
Anti-Inflammatory Agents PD	136	0.38%

Only 25 indexing terms are presented due to the problem of considering the clarity of the information presented in the corresponding analysis figures. If clarity is neglected to show more indexing terms, it is expected that more applications of metformin in other fields can be further explored.

4 Discussion

In this study, a variety of bibliometric tools were used to perform bibliometric analysis on 23513 metformin-related texts from the PubMed database from 1959 to 2020. The conclusions will be elaborated from the following aspects:

I. This study explored the major events and relative enlightenment of metformin drugs in different historical periods.

In this study, combined with index terms and the corresponding analysis results of subheadings, it is speculated that there may be six major changes in metformin in therapeutic or drug use (except hypoglycemic effects) from 1959 to 2020: In the 1980s, metformin was found to have a certain effect on the treatment of obesity; From 1997 to 2000, metformin was found to play a certain role in the treatment of cardio-cerebrovascular diseases; In 2000-2005, metformin was found to be beneficial in the treatment of polycystic ovary syndrome; In 2008-2014, metformin was found to play a role in anticancer and altering gut microbiome composition; Since 2000, metformin has been found to be anti-inflammatory; In 2015-2020, metformin was found to be effective in delaying aging to some extent. By reviewing the latest research progress on metformin drugs, it is found that metformin does have the value of drug reuse in the treatment of obesity, cardio-cerebrovascular diseases, polycystic ovary syndrome, anti-cancer, anti-inflammation, altering gut microbiome composition and anti-aging. Moreover, the occurrence time of these six events coincides with the six time nodes obtained by correspondence analysis, which further proves that correspondence analysis can accurately explore the important events of a drug in different historical periods.

At the same time, an important enlightenment is also obtained in the corresponding analysis using index terms combined with subheadings. It is found that in the previous years of these six events, the combination of index terms and subheadings corresponding to these events usually exists in a low frequency, which is especially obvious in recent years. For example, the retrieval found that the term “Polycystic-Ovary-Syndrome DT” first appeared in 1994, and that it had always appeared at a low frequency from 1994 to 1997. However, until 1998, the frequency of the term began to increase rapidly. Similarly, “Neoplasms DT”, “Gastrointestinal-Microbiome DE” and other terms also appeared at a low frequency in 2000-2005 and then suddenly increased rapidly after 2005. Almost all of these turning points were in the years before the six major events. Therefore, it is necessary to pay attention to the low-frequency index terms related to therapeutic use, pharmacology, drug action and other subheadings in recent years, as well as the terms with abnormally increased frequency, because they are likely to be applied to other fields of metformin in the next few years. According to the statistics of index terms combined with subheadings, it is found that in recent years, several terms such as Alzheimer’s disease, neurodegenerative diseases, and androgenetic alopecia have low-frequency combinations with subheadings such as therapeutic use, pharmacology, and drug action. Therefore, it can be speculated that metformin is likely to have the potential of drug reposition in the treatment of these diseases in the future.

II. The reason why metformin has a long life cycle and has not been replaced by other hypoglycemic drugs.

The reason why metformin can have a long life cycle, in addition to its continuous implementation of cross-domain applications, another important reason is that it has always occupied a dominant position in the field of lowering blood sugar. Exploring the causes of this problem still needs to be combined with the corresponding analysis of index terms and subheadings. As mentioned in the analysis above, Thiazolidinediones TU, Sitagliptin-Phosphate AD, Linagliptin AD, Canagliflozin AD, Exenatide TU, Liraglutide TU and Glipizide AD are all closer to the post-2000 nodes, especially in 2015-2020. However, these nodes represent hypoglycemic drugs, and they are alternatives to metformin, but why do they not shake the position of metformin in the field of hypoglycemic? The problems can be found by induction and comparative analysis. Thiazolidinedione belongs to thiazolidinedione hypoglycemic drugs; sitagliptin-Phosphate and Linagliptin belong to DPP-4 inhibitors; canagliflozin belongs to SGLT-2 inhibitor drugs; exenatide and Liraglutide belong to GLP-1 receptor agonist; glipizide belongs to sulfonylurea drugs; acarbose belongs to α-glucosidase inhibitors. By comparing the advantages and disadvantages of these six drugs with metformin (as shown in Table 3), it can be found that metformin has obvious therapeutic advantages, but it shows quite few side effects. However, other types of drugs do not have this feature. At the same time, the correspondence analysis found that the node of “AMP-Activated-Protein-Kinases ME” was close to the nodes between 2008 and 2020, and this time period coincided with the time period of the above-mentioned hypoglycemic drugs. Is there any necessary connection here? AMPK is known to be AMP-dependent protein kinase, which is the core of the study of diabetes-related diseases. It can effectively inhibit liver gluconeogenesis and reduce liver glucose output, which is unique to metformin drugs. At the same time, the AMPK mechanism is also the core of research on cancer, metabolism and other related diseases. This is also the key that metformin can always play a role in the field of hypoglycemic and can continuously achieve cross-domain applications. In the statistics of index terms, it is also found that in recent years, the frequency of AMPK has shown a rising trend in the related literature of metformin, indicating that the study of AMPK mechanism is a hot topic at present. In addition, relevant research teams have also realized the importance of AMPK mechanism. Therefore, this mechanism will be continuously concerned in the development of new hypoglycemic drugs in the future, and it is not ruled out that new drugs will shake the status of metformin.

Table 3 Comparison of advantages and disadvantages of six types of hypoglycemic drugs

Drug name	Advantage	Disadvantage
Metformin	1. Reduce the production of glucose in the liver; 2. Reduce the intestinal absorption of glucose; 3. It will not cause weight gain; 4. It is not easy to cause blood glucose lower than normal levels due to hypoglycemic	1. May cause nausea, diarrhea, abdominal discomfort, anorexia;
DPP-4 inhibitor	1. Increase the utilization rate of GLP-1 hormone; 2. Increase the availability of other molecules involved in maintaining blood glucose balance level	1. May cause colds and allergic reactions; 2. Pancreatitis is a rare side effect
SLGT-2 Inhibitor	1. Reducing blood glucose will slightly decrease blood pressure (which may benefit some people); 2. It is not easy to cause blood glucose lower than normal level due to lowering blood glucose	1. May cause mycotic genital infection and urinary tract infection; 2. May lead to decreased bone mineral density; 3. May cause high levels of ketones in the blood (diabetic ketoacidosis); 4. May lead to increased potassium levels in the blood (hyperkalemia)
GLP-1 receptor agonist	1. Reduce blood glucose level by increasing insulin secretion; 2. Prevent the rise of blood glucose level by reducing glucagon secretion	1. May cause nausea, vomiting, diarrhea, injection reactions; 2. Acute pancreatitis is a rare side effect
Sulfonylureas	Stimulation of insulin release	1. It is easy to cause weight gain; 2. It is easy to cause long-term hypoglycemia
A-glucosidase inhibitor	Does not cause weight gain	May cause flatulence, diarrhea, abdominal pain, cramping, nausea
Thiazolidinedione	1. Increased glucose uptake; 2. Enhanced sensitivity of adipocytes to insulin; 3. Decreased glucose release from the liver	1. Weight gain; 2. Fluid retention; 3. Worsening heart failure; 4. Increased risk of fracture; 5. May cause macular degeneration; 6. May increase the risk of bladder cancer

. The characteristics and causes of metformin in the whole development process.

Through the statistics of the published papers in the field of metformin and the corresponding analysis and cluster analysis of their years and subheadings, it was found that around 2000 was a time node for the significant transition of metformin. Before 2000, the number of metformin-related literature was always small. In the cluster analysis, it can also be seen that the inertia carried by this cluster from 1959 to 1998 is particularly small, and the color difference within the cluster is large, which fully shows that the study of metformin before 2000 is still in the initial stage, and there are large differences in research content, and there is no mature research system. What is the reason? In the correspondence analysis of the term “metformin” combined with the subheading, it can be found that the Anticholesteremic-Agents PD node was close to the node between 1970 and 1980, and the Phenformin AE

node was close to the node between 1976 and 1980. The two nodes coincide in recent years, and the term “Anticholesteremic-Agents” represents a side effect, which coincides with the joint subtitle “AE” by Phenformin. At the same time, according to the statistics of the five highest frequency subheadings, it can be seen that the subheading “AE” has always dominated from 1959 to 1980. It can also be seen in the frequency statistics that “Humans”, “Adult”, “Male”, “Female” and other human-related labels were significantly reduced in 1973-1978, or even the lowest, indicating that the development of metformin during this period may be in crisis. Therefore, it can be speculated that Phenformin may be found to have acidosis reaction in the 1980 s. In fact, phenformin and buformin were indeed found to have a high risk of lactic acidosis after 1971, and phenformin also increased the mortality of cardio-cerebrovascular diseases. Therefore, in the 1970s, both phenformin and buformin announced delisting, while metformin, which is also a biguanide drug, was inevitably affected and even on the verge of drug withdrawal. Until 1997, Professor Turner and Professor Holman of Oxford University found that metformin did not significantly increase the risk of lactic acidosis in addition to its hypoglycemic effect. At the same time, they also found that metformin had a certain protective effect on the cardiovascular system, so metformin began to be used again. It explains why metformin developed slowly before 2000.

However, after 2000, it was found that the amount of metformin-related literature increased significantly, and the subtitles also changed from dominated by adverse reactions to dominated by pharmacology and therapeutic uses, which reflects the great change in the research direction of metformin. In the peak diagram of cluster analysis, it can also be seen that the color of the cluster is generally single and the volume is increasing after 2000, which fully shows that the depth and breadth of metformin research are increasing, and gradually formed a more mature research system. At the same time, in the correspondence analysis of metformin index terminology combined with subheadings, it was also found that drug reposition was mostly concentrated after 2000. In statistics, the term “clinical” suddenly increased after 2000. These evidence shows that the research heat of metformin drugs continues to increase, and the research cycle of drug reposition is also shortening.

In addition, it is found that the advantage of extracting index terms and subheadings for corresponding analysis is that they have clear directivity and can focus on the specific aspects of the research object very accurately. However, the bibliometric methods currently exploring research hotspots still largely adopt keyword cluster analysis or burst items analysis, which can certainly explore the current or future research hotspots, but the results are lack of directivity, for keywords or burst items are only isolated words, which can only reflect the importance of these words, and cannot reflect that these words are acting on specific aspects of the research. The corresponding analysis method of index terms combined with subheadings effectively solves this problem. Of course, the corresponding analysis method also has shortcomings. For example, when exploring the six major events of metformin, it is found that year and index terms are continuous variables, so there may be a case where the node of a term is close to the node of multiple years. Based on this situation, the study can only determine that the event corresponding to this index term is in a certain period of time, but not specifically accurate to a certain year. This is also a question that will continue to be considered and explored in the future.

Through correspondence analysis, this paper mainly explores the milestone events of metformin drugs, its development characteristics and causes, as well as the research trends in various periods and prospects for future research hotspots. This study will provide some reference for the relevant researchers in the field of metformin to the current research progress and its future trend.

Declarations

Fund project

Supported by the National Natural Science Foundation of China under the project “Research on drug relationship based on co-occurrence network analysis of literature entity units in a multidimensional perspective” under Grant No. 71804102.

Author Contribution Statement:

Shengnan Wu: Provide research ideas and solutions;

Jizong Dong: Conduct experiments, write papers;

Jin Gao,Yidan Sun: data collection;

Ruonan Tian: Technical Support;

Qi Yu: final revision.

Declaration of conflict of interest:

All authors declare that there is no conflict of interest.

References

1. Ali, S., & Fonseca, V. (2012). Overview of metformin: special focus on metformin extended release. Expert opinion on pharmacotherapy, 13(12), 1797–1805. https://www.tandfonline.com/doi/full/10.1517/14656566.2012.705829
2. Andújar-Plata, P., Pi-Sunyer, X., & Laferrere, B. (2012). Metformin effects revisited. Diabetes research and clinical practice, 95(1), 1–9. https://www.sciencedirect.com/science/article/pii/S0168822711005249?via%3Dihub
3. Bailey, C. J. (2017). Metformin: historical overview. Diabetologia, 60(9), 1566–1576. https://linkspringer.53yu.com/article/10.1007/s00125-017-4318-z
4. Bharath, L. P., & Nikolajczyk, B. S. (2021). The intersection of metformin and inflammation. American Journal of Physiology-Cell Physiology, 320(5), C873-C879. https://journals.physiology.org/doi/abs/10.1152/ajpcell.00604.2020
5. Bordons, M., Bravo, C., & Barrigón, S. (2004). Time-tracking of the research profile of a drug using bibliometric tools. Journal of the American Society for Information Science and Technology, 55(5), 445–461. https://asistdl.onlinelibrary.wiley.com/doi/abs/10.1002/asi.10397
6. Bost, F., Sahra, I. B., Le Marchand-Brustel, Y., & Tanti, J. F. (2012). Metformin and cancer therapy. Current opinion in oncology, 24(1), 103–108. https://journals.lww.com/co-oncology/fulltext/2012/01000/Metformin_and_cancer_therapy.17.aspx
7. Cariou, V., & Qannari, E. M. (2018). Statistical treatment of free sorting data by means of correspondence and cluster analyses. Food quality and preference, 68, 1–11. https://sciencedirect.53yu.com/science/article/abs/pii/S0950329318300600
8. DeFronzo, R., Fleming, G. A., Chen, K., & Bicsak, T. A. (2016). Metformin-associated lactic acidosis: Current perspectives on causes and risk. Metabolism, 65(2), 20–

29. https://sciencedirect.53yu.com/science/article/pii/S0026049515003066

9. Del Prato, S., Rosenstock, J., Garcia-Sanchez, R., Iqbal, N., Hansen, L., Johnsson, E., … Mathieu, C. (2018). Safety and tolerability of dapagliflozin, saxagliptin and metformin in combination: Post-hoc analysis of concomitant add-on versus sequential add-on to metformin and of triple versus dual therapy with metformin. Diabetes, Obesity and Metabolism, 20(6), 1542–1546. https://dom-pubs.onlinelibrary.wiley.com/doi/full/10.1111/dom.13258
10. Foretz, M., Guigas, B., Bertrand, L., Pollak, M., & Viollet, B. (2014). Metformin: from mechanisms of action to therapies. Cell metabolism, 20(6), 953–966. https://sciencedirect.53yu.com/science/article/pii/S1550413114004410
11. Forst, T., & Bramlage, P. (2014). Vildagliptin, a DPP-4 inhibitor for the twice-daily treatment of type 2 diabetes mellitus with or without metformin. Expert opinion on pharmacotherapy, 15(9), 1299–1313. https://wwwtandfonline.53yu.com/doi/abs/10.1517/14656566.2014.920009
12. Ghazeeri, G. S., Nassar, A. H., Younes, Z., & Awwad, J. T. (2012). Pregnancy outcomes and the effect of metformin treatment in women with polycystic ovary syndrome: an overview. Acta obstetricia et gynecologica Scandinavica, 91(6), 658–678. https://obgyn.onlinelibrary.wiley.com/doi/full/10.1111/j.1600-0412.2012.01385.x
13. Glossmann, H. H., & Lutz, O. M. (2019). Metformin and aging: a review. Gerontology, 65(6), 581–590. https://www.karger.com/Article/Abstract/502257
14. Greenacre, M. (2010). Correspondence analysis of raw data. Ecology, 91(4), 958–963. https://esajournals.onlinelibrary.wiley.com/doi/full/10.1890/09-0239.1
15. Griffin, S. J., Leaver, J. K., & Irving, G. J. (2017). Impact of metformin on cardiovascular disease: a meta-analysis of randomised trials among people with type 2 diabetes. Diabetologia, 60(9), 1620–1629. https://linkspringer.53yu.com/article/10.1007/s00125-017-4337-9
16. Hajjar, J., Habra, M. A., & Naing, A. (2013). Metformin: an old drug with new potential. Expert opinion on investigational drugs, 22(12), 1511–1517. https://wwwtandfonline.53yu.com/doi/abs/10.1517/13543784.2013.833604
17. Hardie, D. G. (2014). AMP-activated protein kinase: maintaining energy homeostasis at the cellular and whole-body levels. Annual review of nutrition, 34, 31–55. https://
18. Lentferink, Y. E., Knibbe, C. A. J., & Van Der Vorst, M. M. J. (2018). Efficacy of metformin treatment with respect to weight reduction in children and adults with obesity: a systematic review. Drugs, 78(18), 1887–1901. https://linkspringer.53yu.com/article/10.1007/s40265-018-1025-0
19. Mallik, R., & Chowdhury, T. A. (2018). Metformin in cancer. Diabetes research and clinical practice, 143, 409–419. https://sciencedirect.53yu.com/science/article/abs/pii/S0168822717314109
20. Markowicz-Piasecka, M., M Huttunen, K., Mateusiak, L., Mikiciuk-Olasik, E., & Sikora, J. (2017). Is metformin a perfect drug? Updates in pharmacokinetics and pharmacodynamics. Current Pharmaceutical Design, 23(17), 2532–2550. https://chinesesites.library.ingentaconnect.com/content/ben/cpd/2017/00000023/00000017/art00012
21. Markowicz-Piasecka, M., Sikora, J., Szydłowska, A., Skupień, A., Mikiciuk-Olasik, E., & Huttunen, K. M. (2017). Metformin–a future therapy for neurodegenerative diseases. Pharmaceutical research, 34(12), 2614–2627. https://linkspringer.53yu.com/article/10.1007/s11095-017-2199-y

22. McCreight, L. J., Bailey, C. J., & Pearson, E. R. (2016). Metformin and the gastrointestinal tract. Diabetologia, 59(3), 426–435. https://linkspringer.53yu.com/article/10.1007/s00125-015-3844-9
23. Nasri, H., & Rafieian-Kopaei, M. (2014). Metformin: current knowledge. Journal of research in medical sciences: the official journal of Isfahan University of Medical Sciences, 19(7), 658. https://europepmc.org/article/pmc/4214027
24. Papanas, N., Maltezos, E., & Mikhailidis, D. P. (2009). Metformin: diamonds are forever. Expert opinion on pharmacotherapy, 10(15), 2395–2397. https://wwwtandfonline.53yu.com/doi/abs/10.1517/14656560903176453
25. Pasquali, R. (2014). Metformin in women with PCOS, pros. Endocrine, 48(2), 422–426. https://europepmc.org/article/med/24913417
26. Rena, G., Hardie, D. G., & Pearson, E. R. (2017). The mechanisms of action of metformin. Diabetologia, 60(9), 1577–1585. https://linkspringer.53yu.com/article/10.1007/s00125-017-4342-z
27. Rizos, C. V., & Elisaf, M. S. (2013). Metformin and cancer. European Journal of Pharmacology, 705(1–3), 96–108. https://sciencedirect.53yu.com/science/article/abs/pii/S0014299913001477
28. Romero, R., Erez, O., Hüttemann, M., Maymon, E., Panaitescu, B., Conde-Agudelo, A., … Grossman, L. I. (2017). Metformin, the aspirin of the 21st century: its role in gestational diabetes mellitus, prevention of preeclampsia and cancer, and the promotion of longevity. American journal of obstetrics and gynecology, 217(3), 282–302. https://sciencedirect.53yu.com/science/article/abs/pii/S0002937817307391
29. Rotermund, C., Machetanz, G., & Fitzgerald, J. C. (2018). The therapeutic potential of metformin in neurodegenerative diseases. Frontiers in endocrinology, 9, 400. https://
30. Salani, B., Del Rio, A., Marini, C., Sambuceti, G., Cordera, R., & Maggi, D. (2014). Metformin, cancer and glucose metabolism. Endocrine-related cancer, 21(6), R461-R471. https://erc.bioscientifica.com/view/journals/erc/21/6/R461.xml
31. Saraei, P., Asadi, I., Kakar, M. A., & Moradi-Kor, N. (2019). The beneficial effects of metformin on cancer prevention and therapy: a comprehensive review of recent advances. Cancer management and research, 11, 3295. https://wwwncbi.53yu.com/pmc/articles/PMC6497052/
32. Soukas, A. A., Hao, H., & Wu, L. (2019). Metformin as anti-aging therapy: is it for everyone?. Trends in Endocrinology & Metabolism, 30(10), 745–755. https://sciencedirect.53yu.com/science/article/abs/pii/S104327601930147X
33. Sourial, N., Wolfson, C., Zhu, B., Quail, J., Fletcher, J., Karunananthan, S., … Bergman, H. (2010). Correspondence analysis is a useful tool to uncover the relationships among categorical variables. Journal of clinical epidemiology, 63(6), 638–646. https://sciencedirect.53yu.com/science/article/abs/pii/S0895435609002376
34. Steinberg, G. R., & Carling, D. (2019). AMP-activated protein kinase: the current landscape for drug development. Nature reviews Drug discovery, 18(7), 527–551. https://wwwnature.53yu.com/articles/s41573-019-0019-2
35. Tuch, B. E. (2016). Clinical use of GLP-1 agonists and DPP4 inhibitors. Pancreatology, 16(1), 8–9. https://sciencedirect.53yu.com/science/article/abs/pii/S1424390315005669
36. van de Velden, M., D'Enza, A. I., & Palumbo, F. (2017). Cluster correspondence analysis. Psychometrika, 82(1), 158–185. https://linkspringer.53yu.com/article/10.1007/s11336-016-9514-0

37. Wróbel, M. P., Marek, B., Kajdaniuk, D., Rokicka, D., Szymborska-Kajanek, A., & Strojek, K. (2017). Metformin—a new old drug. Endokrynologia Polska, 68(4), 482–496. https://journals.viamedica.pl/endokrynologia_polska/article/view/48058

Figures

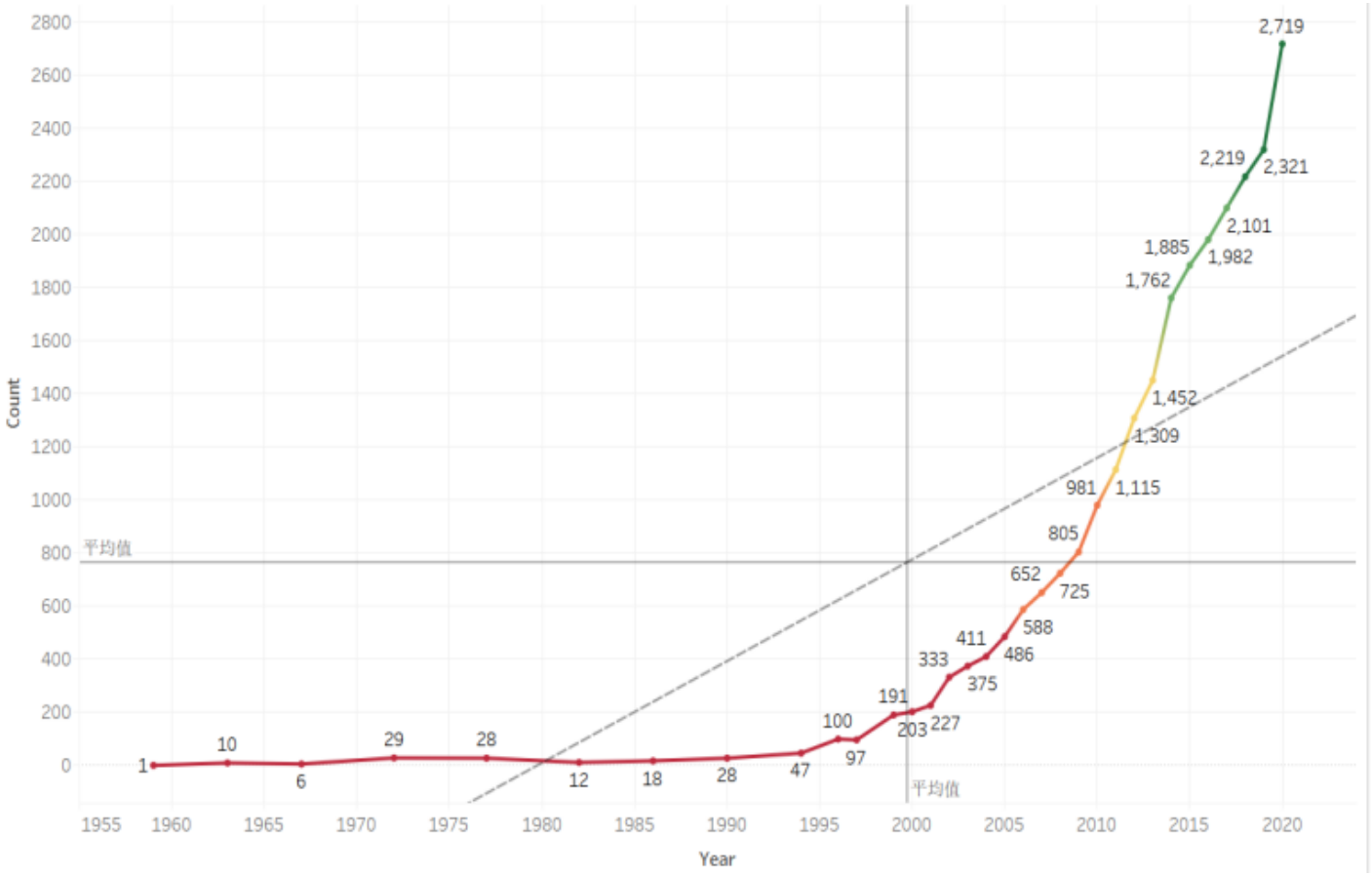

Figure 1

Statistics of metformin annual publications from 1959 to 2020

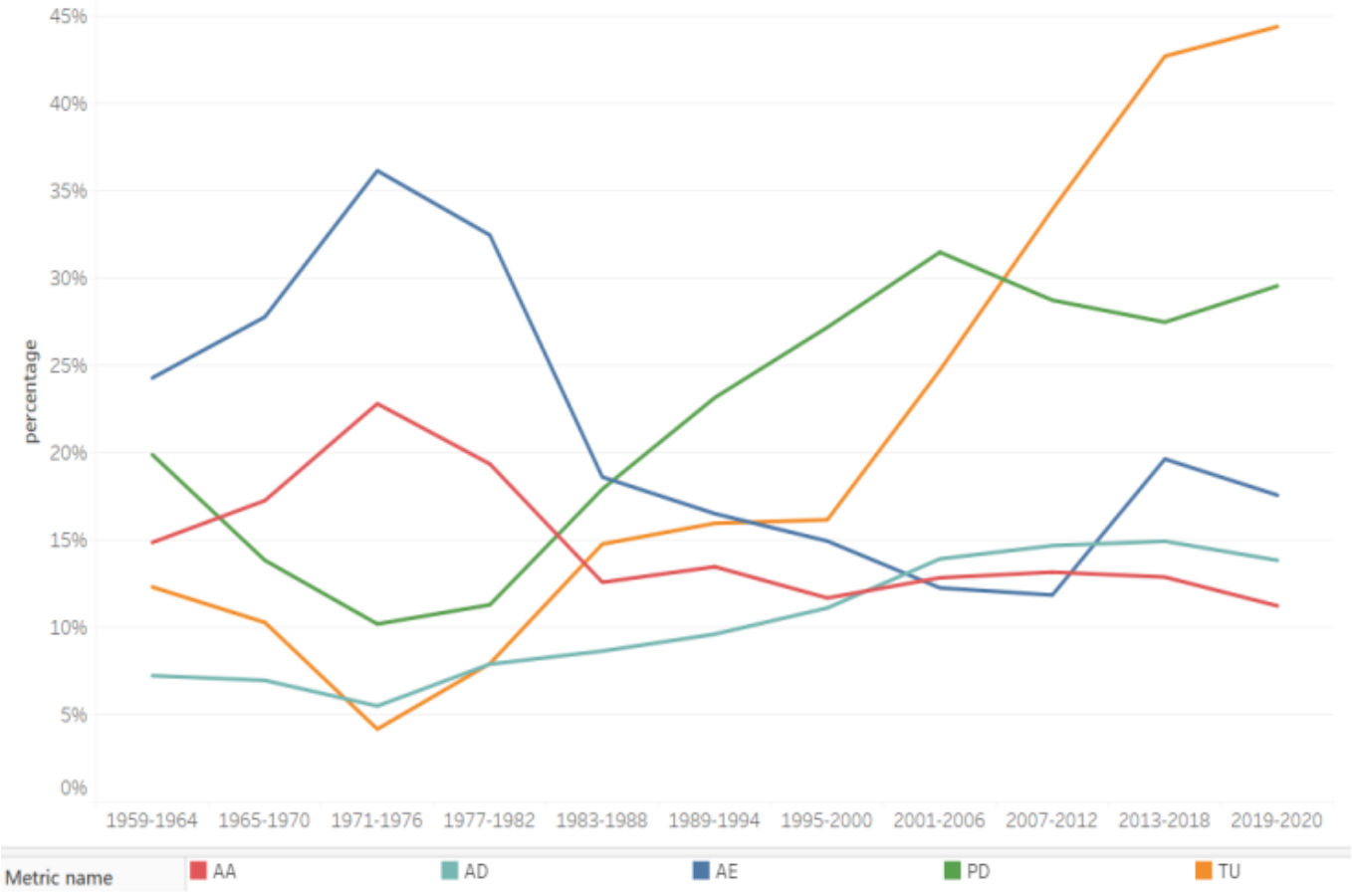

Figure 2

The proportion of the five main subtitles of metformin in different periods

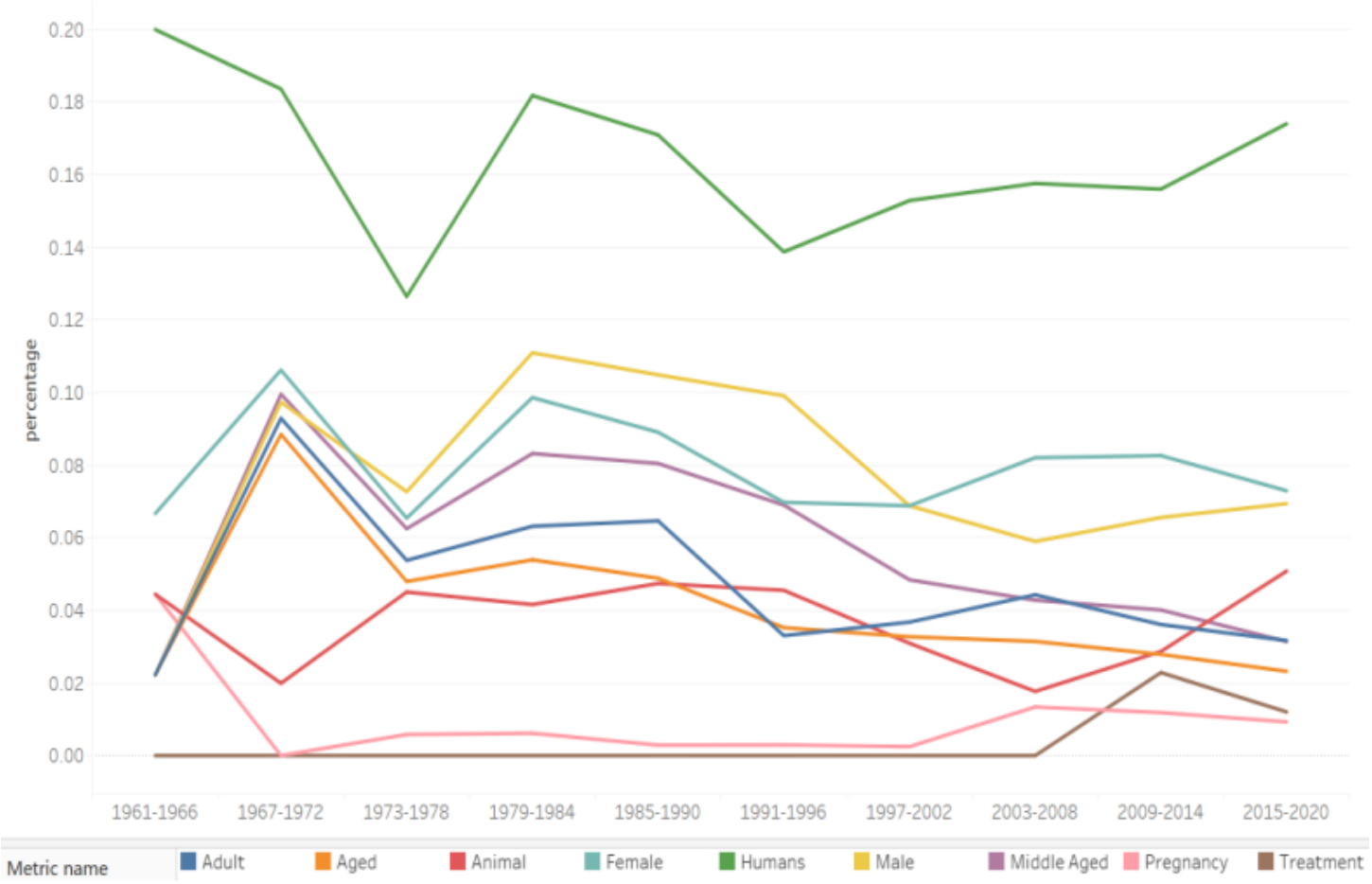

Figure 3

Percentage of inspection label frequency every 5 years to total document inspection label frequency every 5 years

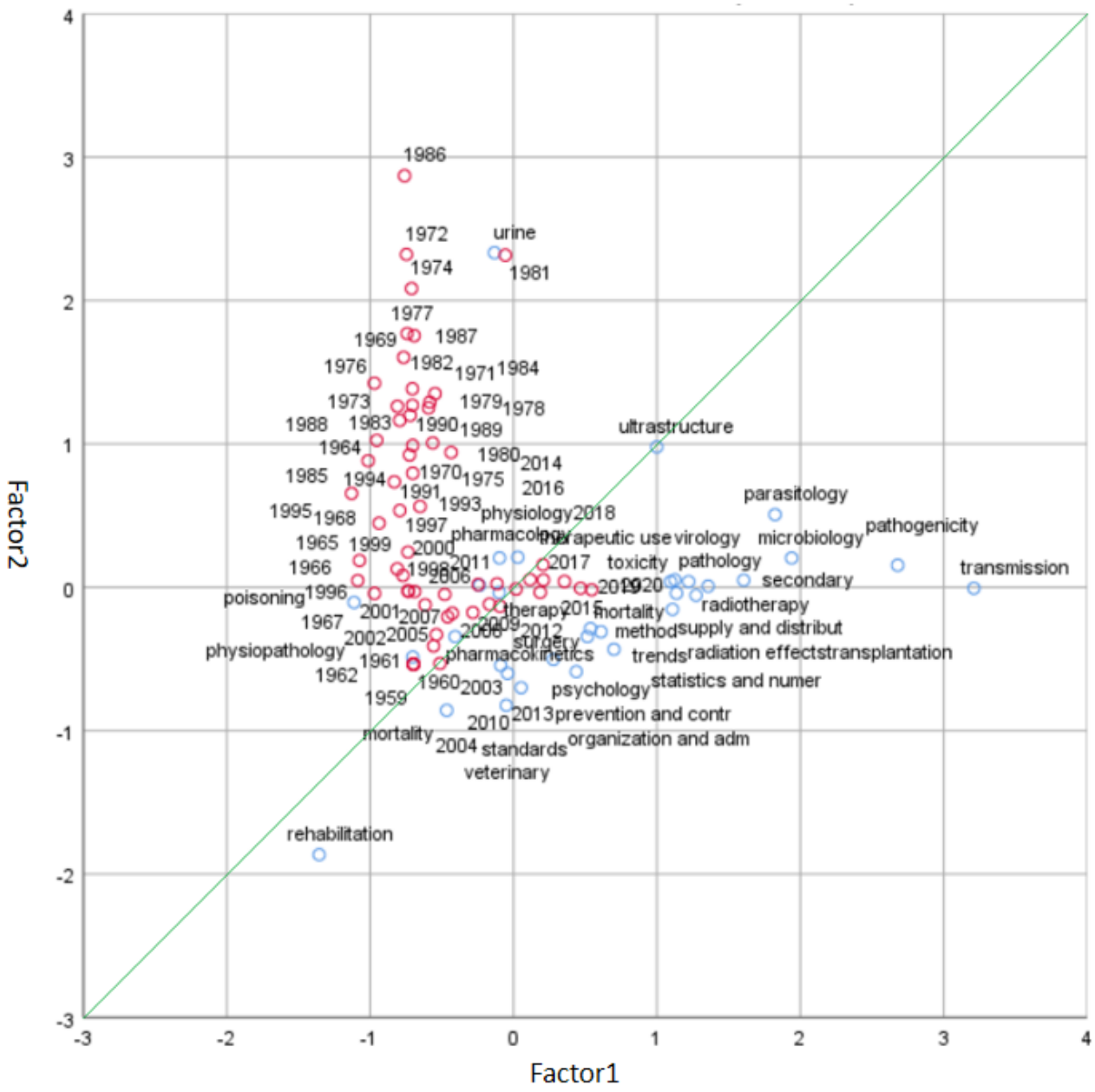

Figure 4

Correspondence analysis for metformin years and subtitles

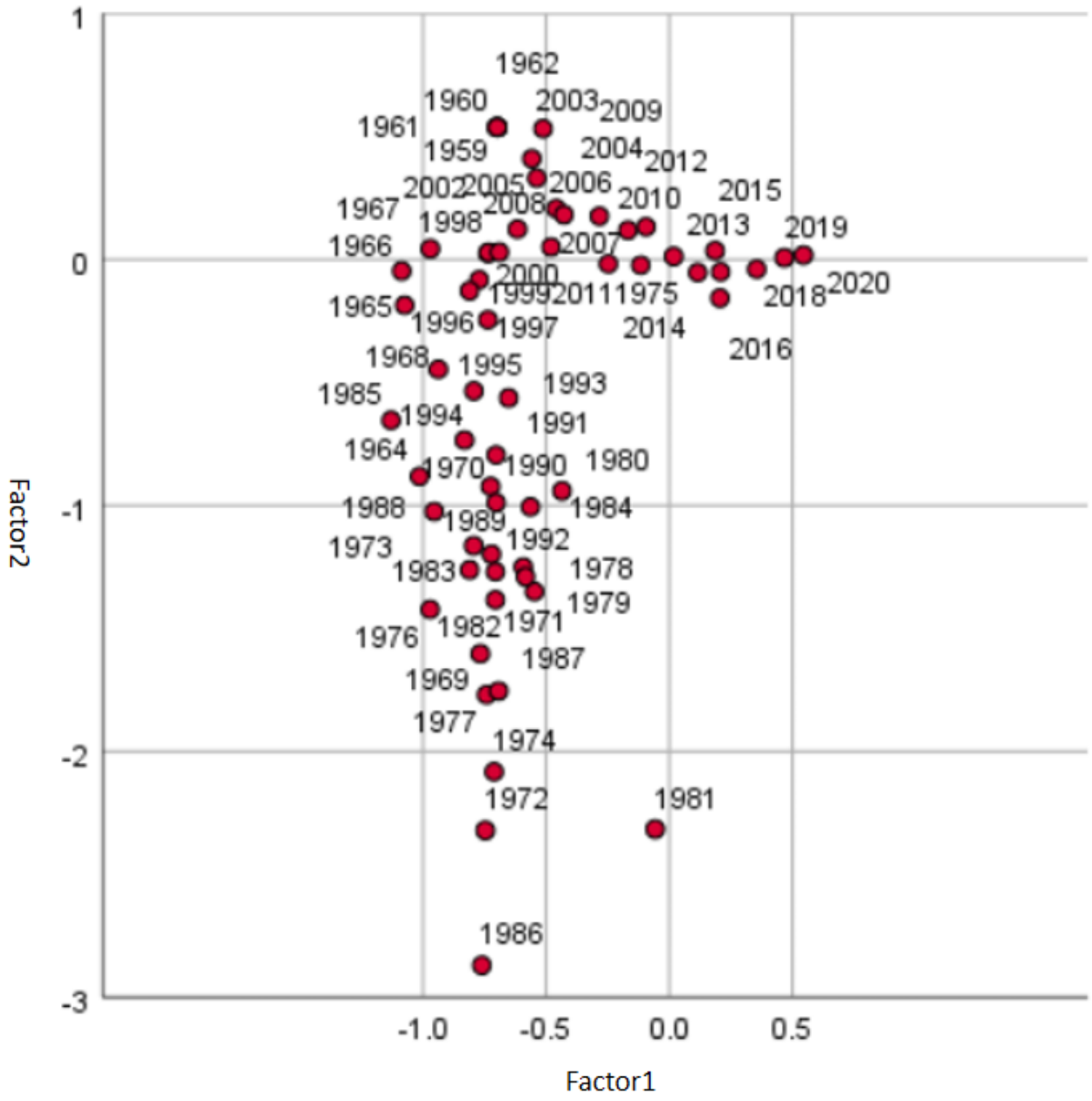

Figure 5

Correspondence analysis for Metformin years

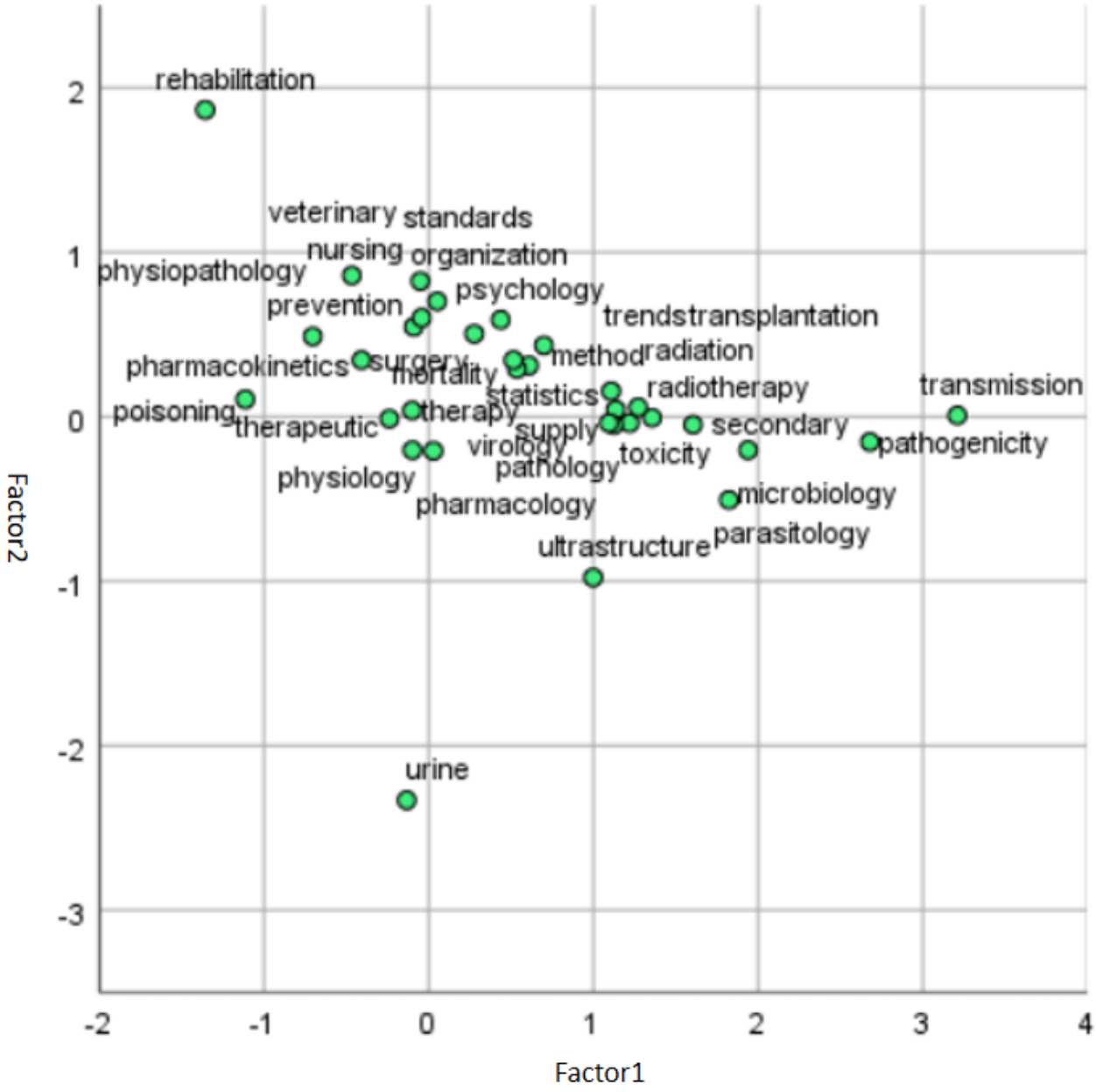

Figure 6

Correspondence analysis for metformin subtitles

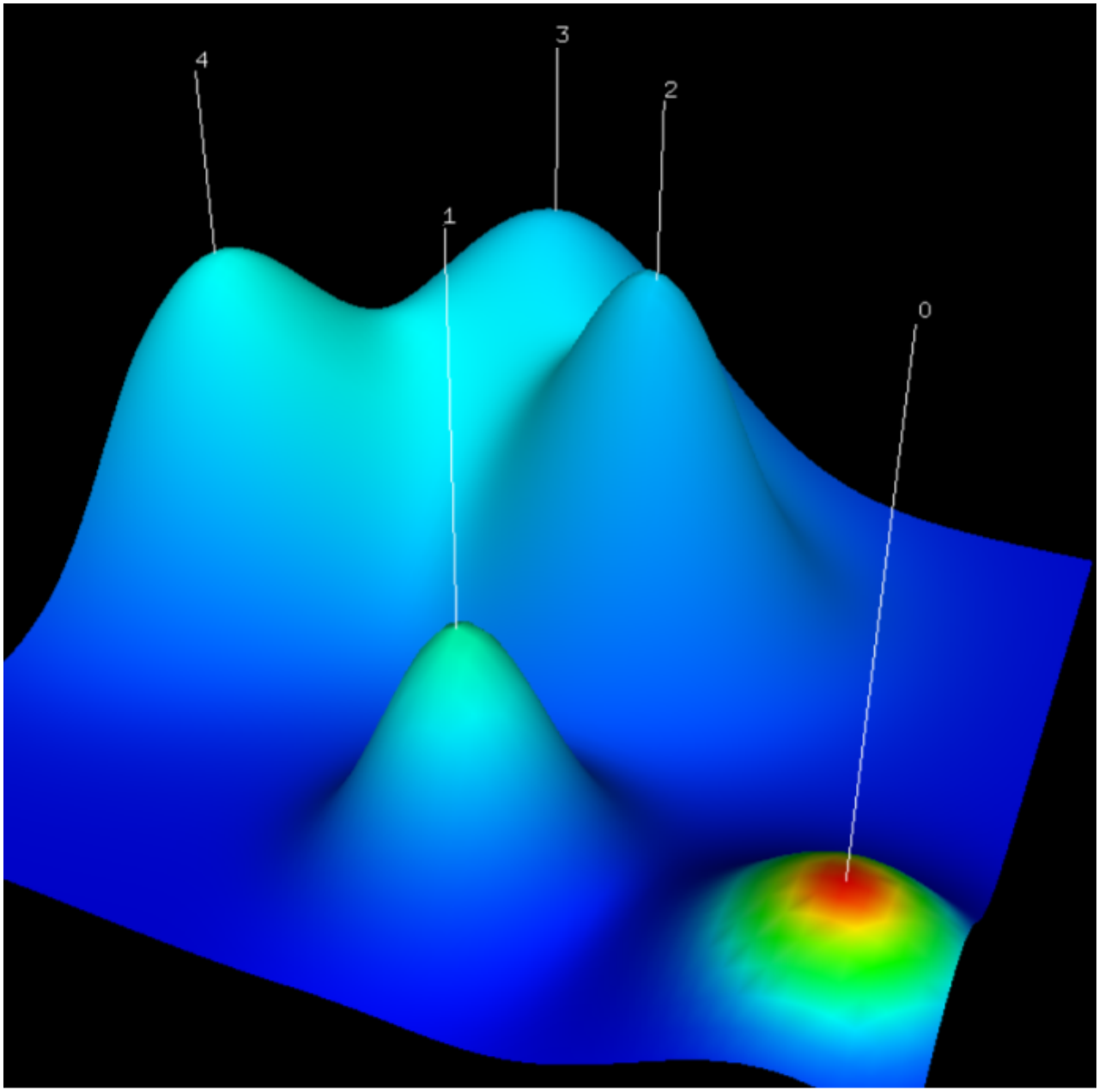

Figure 7

Cluster analysis for years

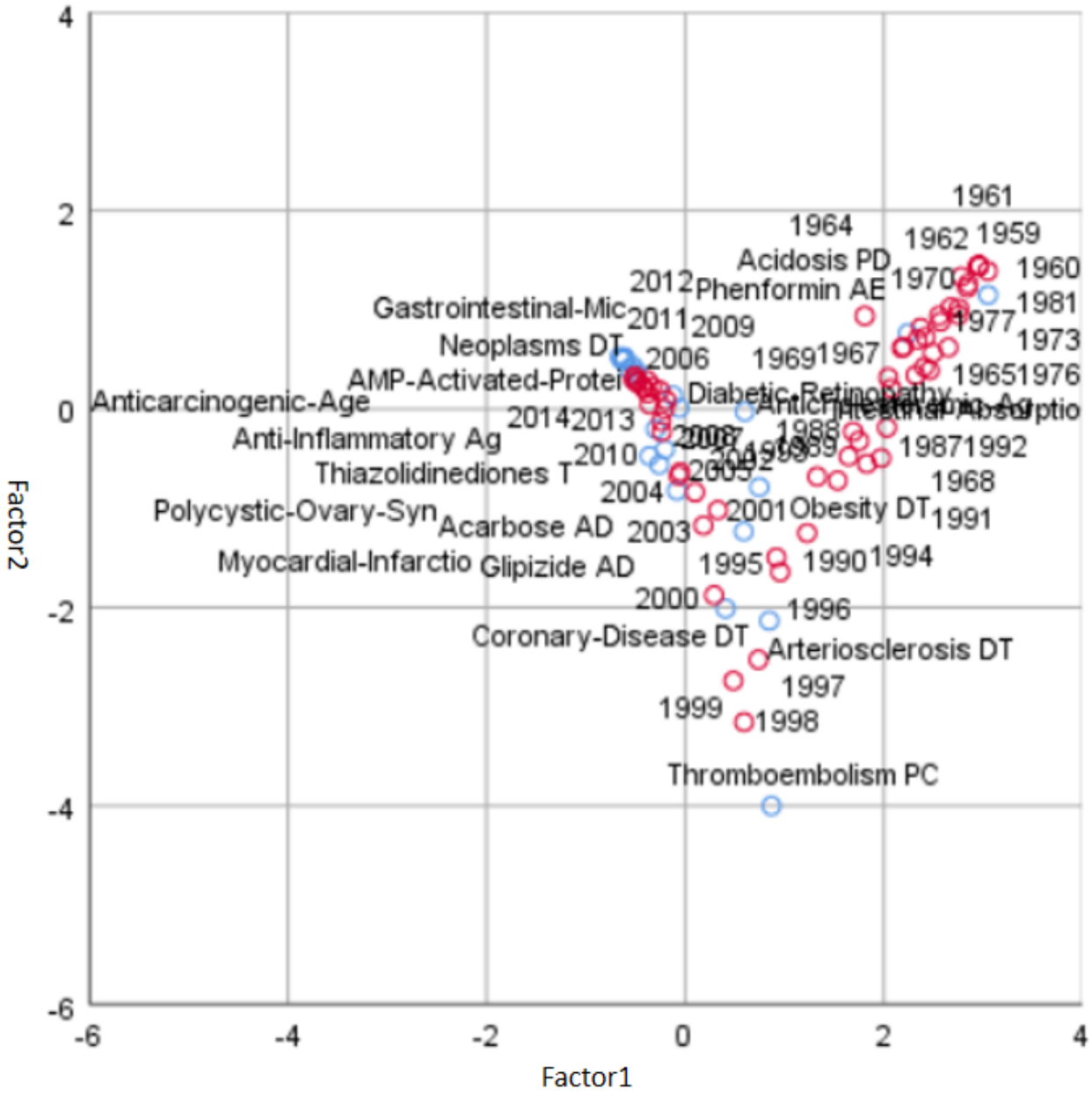

Figure 8

Correspondence analysis between indexing term joint subtitles and years