2019

Citationally Enhanced Semantic Literature Based Discovery

John David Fleig
Nova Southeastern University, john.david.fleig@gmail.com

Citationally Enhanced Semantic Literature Based Discovery

By
John David Fleig

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in
Computer Science

College of Engineering and Computing
Nova Southeastern University

2019

We hereby certify that this dissertation, submitted by John Fleig, conforms to acceptable standards and is fully adequate in scope and quality to fulfill the dissertation requirements for the degree of Doctor of Philosophy.

Junping Sun, Ph.D.
Chairperson of Dissertation Committee

6/6/2019
Date

Francisco J. Mitropoulos, Ph.D.
Dissertation Committee Member

6/6/2019
Date

Jaime Raigoza, Ph.D.
Dissertation Committee Member

6/6/2019
Date

Approved:

Meline Kevorkian, Ed.D.
Interim Dean, College of Engineering and Computing

6/6/2019
Date

College of Engineering and Computing
Nova Southeastern University

2019

An Abstract of a Dissertation Submitted to Nova Southeastern University
in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

Citationally Enhanced Semantic Literature Based Discovery

by
John David Fleig
June 2019

We are living within the age of information. The ever increasing flow of data and publications poses a monumental bottleneck to scientific progress as despite the amazing abilities of the human mind, it is woefully inadequate in processing such a vast quantity of multidimensional information. The small bits of flotsam and jetsam that we leverage belies the amount of useful information beneath the surface. It is imperative that automated tools exist to better search, retrieve, and summarize this content. Combinations of document indexing and search engines can quickly find you a document whose content best matches your query - if the information is all contained within a single document. But it doesn't draw connections, make hypotheses, or find knowledge hidden across multiple documents. Literature-based discovery is an approach that can uncover hidden interrelationships between topics by extracting information from existing published scientific literature. The proposed study utilizes a semantic-based approach that builds a graph of related concepts between two user specified sets of topics using semantic predications. In addition, the study includes properties of bibliographically related documents and statistical properties of concepts to further enhance the quality of the proposed intermediate terms. Our results show an improvement in precision-recall when incorporating citations.

Acknowledgements

"Be humble for you are made of dung,
Be noble for you are made of stars."
-Serbian proverb.

The journey to the stars is a long one, and it is never accomplished alone. There are many that have guided me to complete this body of work. Some are as a beacon in the night, illuminating the way ahead. Some whisper encouragement, lest one lose themselves in the unknown. And others yet whom have tempered the soul to weather the struggles. I would like to thank four groups, without whom this dissertation would not have been possible: my advisors past and present, my committee members, my colleagues, and those I consider family and friends.

I could not have finished this without the aid and advice from my advisor, Dr. Junping Sun. Despite his heavy workloads he always provided the support I needed to keep moving forward. I would also like to thank my advisors from my prior degrees: Dr. Larry Ramsey, Dr. Bradley Clymer, and Dr. Joel Saltz. I am only the researcher I am today because I have been molded by all four of you. Additionally, I would like to thank my dissertation committee members Dr. Jaime Raigoza and Dr. Francisco Mitropoulos for their invaluable feedback. Hopefully they can forgive me for the amount of reading with which they were subjected.

I would be remiss in my acknowledgements to not also mention those outside academia. To my colleagues at Chemical Abstracts - I thank you. Not only did I receive support in the form of computing resources and data access with which to conduct my research, but many have been great sources of information and advice.

Last but not least, I would like to express my gratitude to my family and friends. Through this journey I have not been there as much as I would like to have been. This dissertation would not have been possible without their warm love and continued patience. Finally, my sincere thanks to my partner and dear friend Chase Grozdina, for your unwavering support throughout this process.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Relevance and Significance

We are living within the age of information. Since the 1970's we have progressed from traditional industrialization into an age categorized by the creation, distribution, and manipulation of information as a significant factor to society. This Information Age continues to redefine not only our technology but the level of integration the technology impacts each of us on a daily basis. Already there are more mobile connected devices than there are people across the planet, and within a few years there will be 50 billion smart connected devices collecting, storing, and sharing information.

No matter where you go you'll find people on their cell phones, tablets, laptops, and computers creating data. Every second there are 6,000 tweets over Twitter and 40,000 Google searches. Every minute there are 125,000 photos uploaded to Facebook along with half a million comments posted. Every day there are 200 billion emails sent. Numerous other cameras, scanners, and sensors actively record anything that might be useful from the weather, to traffic patterns measured by the GPS movements of cellular phones, to heart rate information collected by fitness devices that is used within medical studies. We are drowning in a vast sea of information of which less than a single percent is used.

Over 100 million scientific papers have been published to date and approximately two million a year join their ranks. In some fields there may be tens of thousands of articles published every year. Despite the amazing abilities of the human mind, it is

woefully inadequate in processing such a vast quantity of multidimensional information. The small bits of flotsam and jetsam that we leverage belies the amount of useful information beneath the surface. As such, our tools to organize and examine this ever-expanding flow of information becomes a bottleneck to scientific progress. Traditional techniques such as document indexing and search engines can quickly find you a document whose content best matches your query - if the information is all contained within a single document. But it doesn't draw connections, make hypotheses, or find knowledge hidden across multiple documents. The volume of scientific literature being published may be overwhelming for an individual researcher, and it is nigh impossible to keep track of the developments in other research areas, which can cause logical connections between separate areas of knowledge to be overlooked (Weeber, Vos, Klein, Aronson, and Molema 2003).

1.2 Literature Based Discovery

Literature-based discovery (LBD) was first proposed by Don R. Swanson in 1986 through the *Raynaud Syndrome–Dietary Fish Oils* hypothesis (Swanson 1986). By reading portions of more than four thousand MEDLINE articles, Swanson discovered that dietary fish oils lowers blood viscosity, reduces platelet aggregation, and inhibit vasoconstriction. He also observed that a reduction in both blood viscosity and platelet aggregation, as well as the inhibition of vasoconstriction, appeared to prevent Raynaud disease (Cameron et al. 2015).

Swanson therefore theorized that dietary fish oil might prevent Raynaud's syndrome - this hypothesis was later confirmed clinically in 1989 (DiGiacomo, Kremer,

and Shah 1989). The significance of the observation was that at the time none of the literature had a direct association between fish oils and Raynaud disease but it was possible to derive an implicit indirect association via existing intermediate topics.

Although Swanson has been largely credited with the concept of LBD, the general concept predates him. Several decades earlier, American biologist James Peters stated that:

> When an author takes a series of apparently unrelated facts and ideas from two areas of investigation, combines them so that they make new sense, and develops a new hypothesis from the combination, he not only aids in the advance of both fields but also is quite likely to open up a new one... In Sutton's paper (Sutton 1902) you will see this development of relationships between the fields of cytology and heredity, which, at the time Sutton wrote, were considered to be fairly divergent from one another, in that no research techniques were shared... Sutton's paper can be considered the beginning of cytogenetics... This paper is a good model to follow in the preparation of a study involving synthesis and correlation (Peters 1959, p. 27).

This notion has been studied, such as reports from the US National Research Council of the National Academies. They indicate that many scientific discoveries involve drawing new connections between scientific domains (Feller and Stern 2007). Furthermore, many scientific breakthroughs can be characterized by linking disjoint collections of research papers (Chen 2009). Similarly, in a study on US Patent records in

between 1790 and 2010 they discovered that many previously reported inventions featured new combinations between already existing technologies and techniques (Youn, Strumsky, Bettencourt, and Lobo 2015).

To date there have been many connections and discoveries made due to the use of literature based discovery: *Fish oils* and *Raynaud's syndrome* (Swanson 1986), *migraines* and *magnesium* (Swanson 1988), *estrogen* and *Alzheimer's disease* (Smalheiser and Swanson 1996), *Parkinson's Disease* and *Crohn's disease* (Kostoff 2014), new treatments for cataracts (Kostoff 2008) and breast cancer (Li, Zhu, and Chen 2010), and others. Given that the number of new scientific papers added to the human body of knowledge each year is only going to continue to grow, it is essential that we improve our existing tools to better search, process, and visualize this expanding volume.

As a research field, LBD aims at exploring approaches to discover these hidden connections between pieces of already existing knowledge, either in automatic or semiautomatic fashion (Smalheiser 2012). Its main goal is to produce both interesting and novel knowledge that has yet to be published in scientific literatures (Swanson 2008).

1.3 Semantic-Based Literature Based Discovery

Most LBD systems produce a list of intermediate terms as their results. This serves to illuminate a potential relationship, but does nothing to illustrate the underlying mechanisms of those interactions. Semantic-based approaches are designed to provide this contextual relevance in addition to the association. Existing research in this area have made use of natural language processing (NLP) tools such as SemRep (Rindflesch and Fiszman 2003) and BioMedLEE (Lussier, Borlawsky, Rappaport, Liu, and Friedman

2006) to extract semantic propositions from free-text documents. These relationships can help to illuminate the mechanisms of interaction and the causality between concepts.

Here, not only is the correlation between the two concepts captured, but the contextual relevance of that relationship as well. For example, a predicate such as [`Serotonin - CAUSES - Contraction`] indicates that the concepts of *Serotonin* and *Contraction* show up together, and that *Serotonin* causes *Contraction*. As the predicates relate one concept to another it is straightforward to utilize these to construct graphs for visualization where nodes represent concepts and connecting edges as predicates.

Recent work by Cameron et al. has done just this, extracting semantic predicates via SemRep to form a directed predication graph (Cameron et al. 2015). As an example, Magnesium has been known to exert an inhibitory effect on platelet prostacyclin (Epoprostenol) interaction, and prostacyclin is known to affect migraines (Briel, Lippert, and Zahradnik 1985; Peatfield, Gawel, and Rose 1981). These relationships are illustrated as directed edges in Figure 1 where the concepts of *Magnesium*, *Epoprostenol*, and *Migraine Disorders* comprise the graph vertices.

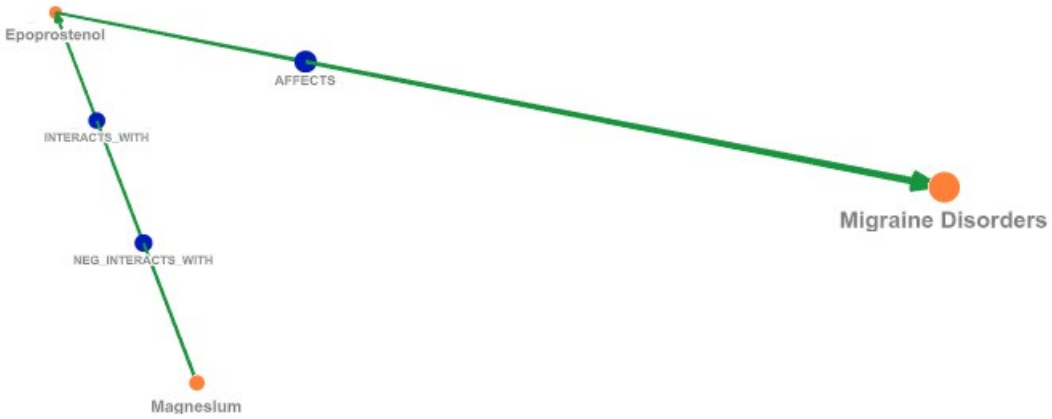


Figure 1. Example relationship between Magnesium and Migraine Disorders

1.4 Goals

A challenge of LBD is the reliance on having a collection of text documents with which to study, as frequently only the title and abstracts of a reference are available in large quantities. In the study by Cameron et al. they extracted predicates by using the full-text of documents as input to the SemRep tool to achieve their results (Cameron et al. 2015). In addition, Cameron noted in his dissertation that the "results collectively suggest that titles and abstracts alone, might NOT be sufficient for LBD" (Cameron 2014).

The goal of this dissertation is to measure the effectiveness of incorporating bibliographic citations into semantic-based LBD. When an author chooses to cite another document, they do so because there is some important piece of information that is foundational or related to their own work.

LBD approaches typically start with a literature search of all documents that contain those concepts specified by the user. By also utilizing references that are bibliographically related to our search terms, we are leveraging additional documents that are semantically related to our primary terms of interest that don't already include those terms.

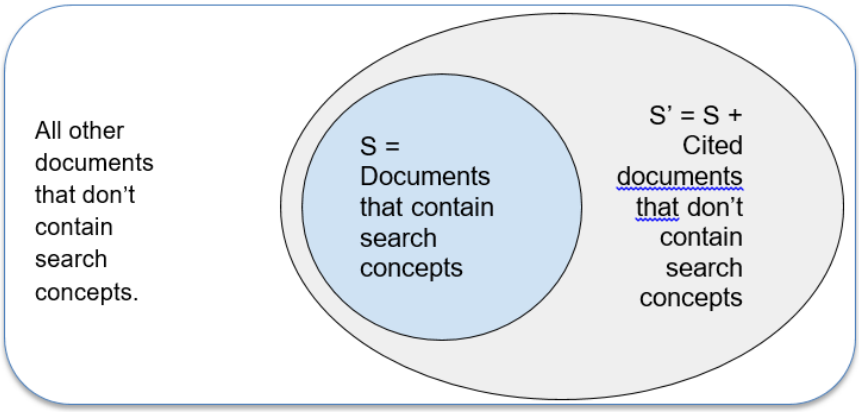


Figure 2. Content of References

Illustrated in Figure 2, the initial search returns a set of documents (S) that contain those concepts. We then expand this set, naming it S', to also include the cited documents of S. Any documents within S' that are not within S do not contain the search concepts but are related. Therefore this approach provides the benefit of potentially adding new intermediate terms and possible pathways from these documents with a low risk of cluttering up the results with totally unrelated knowledge.

This study proposes to accomplish this goal by modifying the approach by Cameron et al. (2015) to take into account cited documents. This will then be followed by performing a retrospective analysis against past LBD discoveries to investigate if the inclusion of citation information can retrieve better results such as revealing new relationships. The studies in chapter four show a comparison of the results of the expanded document set S' and the baseline set S.

1.5 Research Questions

This proposed dissertation research concerning the blend of bibliographic citations with semantic-based predicate graphs suggests the following research questions:

Quantitative Study: Many LBD approaches begin with an initial set of references which include one or more concepts of interest to the user. This is generally done to reduce the scope of the problem down from the full corpus to a smaller more manageable subset that is more likely to have points of interest. Cameron's approach constructs predication subgraphs using only those predicates that exist in that collection of documents. By expanding the initial set of documents via citations, how does this affect

the precision and recall of the returned answers? Many LBD studies use retrospective studies as a means of evaluation, comparing the results to those of the original discovery. We have performed several retrospective studies to investigate this question with positive results, discussed in Chapter 4.

Qualitative Study: Many concepts will appear related to other terms, but what are good ways to ignore concepts that provide little information? This can be accomplished statistically such as using IDF, which is a measure of how much information a word provides, or employing the use of existing knowledge like ontological hierarchies. We have used a combination of the IDF values for individual concepts calculated over the entire corpus and the concept's position within the MeSH ontology.

Quantitative Study: Our results are essentially a ranked order of intermediate concepts/paths, thus the quantitative results of the retrospective studies are shown as precision-recall curves. The use of citations expands the potential number of paths within the results, how does this affect the precision-recall curves? Our citationally enhanced studies have a better precision-recall curve. This is attributed to two cases. The first, and most important, case is due to new paths being generated from the addition of citations. The second case is caused by an increase in the IDF concept threshold.

1.6 Barriers and Issues

The potential of Swanson's research has been widely acknowledged, but so has the depth and complexity of the field when considering the expansive quantity of source documents and the potential numbers of interconnections (Hearst 1999; Weeber et al. 2001). There have been three large research challenges within LBD. The first is the

scalability of algorithms in dealing with very large volumes of information. The second is the heavy reliance on subject matter experts or domain-specific knowledge sources. Typically their help, whether automatic or manual, better refines a search for the researcher or helps filter the results. For example, every concept within the Metathesaurus has been assigned one or more semantic types ("Cell function", "molecular function", etc.). Thus a researcher could make use of this by only wanting to look at connecting concepts that match these additional criteria (Weeber et al. 2001). The utilization of domain-specific sources helps natural language processing (NLP) to tackle the variation and intricacies of natural language. Finally, the third difficulty is the lack of a standard means of comparing the successfulness of LBD approaches. By its very definition LBD attempts to uncover previously unknown details. This lack of a true gold standard has caused many to simply qualitatively compare results to a select few original studies (Yetisgen-Yildiz and Pratt 2008).

1.7 Summary

This expanding sphere of human knowledge has precipitated the need for researchers to narrow their focus as it is infeasible for an individual to keep abreast of the swelling volume of information. LBD is an attempt to help solve these challenges - dedicated to building tools to discover, visualize, and traverse the hidden connections of knowledge across many disciplines.

Semantic-based LBD techniques attempt to expand this by not only uncovering connections, but illustrating the meaning of those connections. Stating that *Fish Oils* are

related to *Raynaud's disease* is useful, but not as useful as the statement: *Fish Oils* are related to *Raynaud's disease* due to their effect on *blood viscosity*.

Research into summarization techniques, discussed in chapter 2, has shown that there are semantic correlations between a document and another references that cite it. This dissertation explores the usefulness of incorporating bibliographic citations into implementations of semantic LBD.

Chapter 2

Review of the Literature

2.1 Introduction

Hypothesis generation is a key step in the process of making scientific discoveries as much time and effort can be wasted just to disprove an initial hypothesis only to repeat the cycle again. Traditionally, this step relies on prior knowledge, experience, and even intuition (Srinivasan 2004). Much of what was involved with Swanson's work was reading and assimilating the literature and then using good judgment to explore related topics until finally happening upon a plausible discovery. Automating this knowledge discovery needs to focus on extracting useful patterns or relationships within large data sets.

Literature-Based Discovery, is, in essence, a field focused on a blend of text mining, information retrieval, and summarization in an attempt to answer questions regarding the relationship between two concepts. Much of the work in these areas is relevant and is covered within this chapter first, followed by a review of current research within the field of LBD.

2.2 Text Mining

The process of discovering patterns in large data sets is termed data mining. Data mining typically focuses on structured data stored within a database - looking for trends, identifying clusters, frequently occurring sets, and so forth. A subtopic, text mining,

makes use of many data mining techniques but instead looks for patterns among the unstructured raw text of document collections.

In the simplest form a document collection, or corpus, can be any grouping of text-based documents from a few to millions. An example of a well-used document collection suitable for text mining is PubMed – the National Library of Medicine's online repository of biomedical research papers. To date PubMed contains over 28 million document abstracts and is widely used by computer scientists in employing text mining (Feldman and Sanger 2007).

As a research area, text mining incorporates techniques from other areas such as information retrieval (IR) systems, natural language processing (NLP), and machine learning combined into three different tasks: Information Retrieval, Question Answering, and Text Summarization.

2.3 Information Retrieval

Information Retrieval can be defined as "finding material (usually documents) of an unstructured nature (usually text) that satisfies an information need from within large collections (usually stored on computers)" (Manning, Raghavan, and Schütze 2008). The task itself is to find a subset of a larger collection which contains the desired information. Most IR systems commonly extract meaningful words, concepts, or text fragments and index the entire corpus in advance to avoid having to linearly scan all text in response to a query.

The IR system also provides a search system containing a user interface or query language. A user can express their needs by means of search primitives that can be

understood by the system. The user query is then translated into a query that can be executed by the underlying system. For example, a user may request information about *Fish Oils* via the provided interface. If the IR system makes use of a Relational Database Management System (RDBMS) this might be converted into a SQL query that specifically looks for the given text within both the titles and abstract fields: `SELECT DocumentIds From SomeTable WHERE (Title LIKE '%Fish Oils%' OR Abstract LIKE '%Fish Oils%')`. A second example that is frequently utilized for text processing is Apache Solr coupled with Apache Lucene. A similar query using a solr-based system might look like: `Title:"Fish Oils" OR Abstract:"Fish Oils"`.

Once the system queries are executed against the indexed corpus, matching results are returned to the user ordered by relevancy. Despite the power of such systems one drawback is that users must look through potentially large volumes of results to find relevant information. The combination of document indexing and search engines can quickly find you documents whose content best matches your query - if the information is all contained within a single document. This can be problematic when the information desired is spread across multiple documents. A user must therefore already know exactly what they are looking for and sift through the multitudes of results or continually reformulate their queries in an ad hoc fashion to search for all the pieces to the puzzle. An area of research known as Question Answering has techniques to help address these limitations.

2.4 Question Answering

Unlike the traditional Information Retrieval approaches, Question Answering (QA) attempts to provide users with precise answers to their questions instead of providing a large number of documents that are potentially relevant. To do so, QA systems consist of three distinct phases: question processing, document processing, and answer processing. These phases use a combination of Information Extraction and Natural Language Processing techniques (Athenikos and Han 2010).

The question processing phase works much like Information Retrieval systems in that they interpret the user's input and translate it into an internal query. However, there may be additional linguistic processing of the user query to determine the type of question being asked and the expected answer type. Additional steps may also include Named Entity Recognition (NER) to identify keywords or concepts.

In the second phase, Document Processing, the internal query is used to fetch a set of resulting documents. The system attempts to select specific sections from the resulting documents (termed candidate answer passages) that are most relevant to the query. To accomplish this many QA systems make use of semantic knowledge in a combination of named entity recognition and word/query disambiguation.

Named entities can be identified via the use of domain specific ontologies such as Medical Subject Headings (MeSH) which is the National Library of Medicine's (NLM) controlled vocabulary thesaurus used with PubMed, and will be described in more detail later. Systems like this are useful to both identify key terms for indexing articles but also to identify when there may be more than one name for the same thing, such as *Epoprostenol* and *Prostacyclin*.

Another type of useful semantic source is WordNet, a large lexical database of English. Within WordNet different parts of speech are grouped into sets of cognitive synonyms (Athenikos and Han 2010). Consider an information retrieval search of the word “iron”, here, possible results may include metals, golf clubs, and laundry services. Identifying how a word is being used (noun vs verb) and the words with which it keeps company (context) can help determine relevancy of search results.

Inclusion of semantic information can ensure that when a user searches for *Epoprostenol* that the results also include documents that talk about *Prostacyclin*, or that a search about *Magnesium deficiency*, might also include results with *hypomagnesaemia*. In addition, a search of “iron oxide” likely means a user is not interested in oxides and their use within laundry.

In the answer processing phase the candidate answers are then matched against the expected answer types specified and then ranked according to matching scores as determined by the search.

As a whole, Question Answering is considered an effective, but more challenging task than information retrieval. It necessitates additional processing during corpus indexing and processing of user queries. Furthermore, semantic knowledge systems tend to be domain specific and thus cannot necessarily be applied across diverse texts. When considering complex problems it can be challenging for a user to identify the best queries to use, these techniques attempt to do some of that work for the user.

2.5 Text Summarization

Text summarization is designed to combat the problem of large result sets by providing a concise overview of the information from one or multiple documents in either a textual or graphical summary (Mishra et al. 2014; Mani 1999). In general, automatic text summarization can be categorized into several different types, each described based on its input type (single or multi document), the output type (extraction or abstraction), and the approach (Statistical, Latent semantic analysis, Natural Language Processing, and others) (Yogan 2016).

Most of the initial research in this area focused on single document extractive summarization (Aliguliyev 2009; Ko and Seo 2008). Extractive summarization works by identifying the most important sections of the original text and compiling them together. With the alternative, abstractive summarization, the system attempts to paraphrase the content by means of natural language processing (Ganesan et al., 2010; Khan et al., 2015; Allahyari et al. 2017). There has been less work in this area due to its complexity, but there has been a growing interest in multi-document summarization. This is likely due to the volume of information available and that relevant information is often scattered among multiple sources (Mishra et al. 2014).

2.5.1 Statistical Text Summarization

Statistical implementations are typically based on the early works of Luhn, Spärck Jones, and Edmundson. Luhn proposed that more important words will be repeated within a document more often than others - this measurement became known as Term Frequency (Luhn 1958). Karen Spärck Jones later created a second measure called

Inverse Document Frequency (IDF) to account for the idea that if a word shows up repeatedly in many documents then it loses meaning by being too general (Spärck Jones 1972). A common approach is to use the combination of both, entitled Term Frequency Inverse Document Frequency (TF-IDF). Consider a single document that contains "Hypomagnesemia" many times. In this case, it is likely that it is a central point of that document. Whereas the word "patients", which despite appearing in large numbers across many documents, has far less useful meaning.

$$\text{TF-IDF(w,d)} := \text{TermFrequency(w,d)} \times \text{InverseDocFrequency(w)} \tag{2.1}$$

$$\text{TF-IDF(w,d)} := \text{TermFrequency(w,d)} \times \log\left(\frac{|\text{Corpus}|}{\text{DocFrequency(w)}}\right)$$

Although there are multiple variations, a simple form of the TF-IDF is shown in Equation 2.1. The first variable, *TermFrequency*, is a function of term *w* and document *d*. For a given term and document, it is the number of times that *w* appears within *d* divided by the total number of terms contained within *d*. The second variable, *InverseDocFrequency*, is the log of the total number of documents within the corpus divided by the number of documents that contain the term *w*.

As an example, suppose our corpus consists of two documents (Table 1). The first document, d_1, contains the text "example: this is an example" and the second, d_2, "this is a second sample". The term frequency of a term is calculated for a given document, and is equal to the number of times that particular term appears within that document divided by the total number of terms for the document.

Table 1. Example term counts for two documents

d_1		d_2	
Term	Term Count	Term	Term Count
this	1	this	1
is	1	is	1
an	1	a	1
example	2	second	1
		sample	1

The term frequency of the term "example" for each document is calculated as follows:

$$\text{tf("example", d1)} = \frac{\text{term count("example", d1)}}{\text{total terms(d1)}} = \frac{2}{5} = 0.4$$

$$\text{tf("example", d2)} = \frac{\text{term count("example", d2)}}{\text{total terms(d2)}} = \frac{0}{5} = 0$$

The inverse document frequency is calculated once for the entire corpus and represents the ratio of documents that include that particular term. For the term "example", it would calculated as follows:

$$\text{idf("example")} = \log\left(\frac{\text{\# of documents}}{\text{\# of documents with term}}\right) = log\left(\frac{2}{1}\right) = 0.301$$

If we examine the idf of the term "this" we find that since every document in our corpus contains this word, thus the idf equates to zero:

$$\text{idf("this")} = \log\left(\frac{\text{\# of documents}}{\text{\# of documents with term}}\right) = log\left(\frac{2}{2}\right) = 0$$

This implies that the word is not very informative as it appears in all documents. The measure of tf-idf appears often within statistical text mining when considering individual terms, but doesn't account for sentence level constructs.

An intuitive approach to determine the importance of a sentence is to first identify the characteristics that highlight the relevance of the sentence. In 1969, Edmundson defined features which reflect the relevance of a sentence such as the sentence position and the presence of title or cue words (Edmundson 1969). For title words, the idea is simple: as the title is the most succinct summary of the document by its authors, sentences that contain words that also occur within the title are likely to be more important. This is in contrast to cue words, which are not important by themselves but rather signify that something else is more likely to be important. Adjectives and adverbs like "significant", "irrelevant", "impossible", and others are applied in text to highlight the neighboring word or phrase to the reader. Thus, if a list of such words is provided to a summarization system it can be used within the scoring process. Each sentence is then ranked based upon some scoring algorithm such as in Equation 2.2.

$$\text{Score} = \sum_{i=1}^{n} wi \times fi \tag{2.2}$$

where w_i is the weight of a particular feature and f_i is the score of that feature. Each sentence will receive a score based upon all of its features, and the document summary is produced from the highest valued content. Note that although specific words may be called out (such as cue words, stop words, title words, etc.) only statistical properties, relative positioning, and existence are examined in those methods.

2.5.2 Latent Semantic Analysis

Latent Semantic Analysis (LSA) is a method that attempts to identify the underlying meaning of a document. In other words, it extracts the relationships between a set of documents and the terms they contain. This might be simple if a given term described only a single concept, and a concept could always be described by a single term. But natural languages have many situations where this is not the case. Thus the idea behind LSA is that "the aggregate of all the word contexts in which a given word does and does not appear provides a set of mutual constraints that largely determines the similarity of meaning of words and sets of words to each other" (Landauer, Foltz, and Laham 1998). For example, the word "bank" in English can refer either to a financial institution or to the land alongside a river. If a document contains the word along with other words such as loans and interest then it likely refers to a financial institution, but if the document also contains terms such as fishing and rivers then it refers to the second.

The algorithm first constructs an *m* x *n* matrix whose *m* rows represent each unique term and *n* columns represent each document. Each entry within the matrix, $a_{i,j}$ is the weight of the term *i* within sentence *j*, computed by TF-IDF. The next step applies a linear algebra technique to the matrix called single value decomposition (SVD) which factors the original matrix into several derived from its eigenvectors. If desired, dimensionality reduction can also be applied during the SVD to retain only the *k* principal eigenvectors. The result of this produces three new matrices U, Σ, and V^T (See Eq. 2.3).

$$M = U \Sigma V^T \tag{2.3}$$

If the original matrix, M, is *m* x *n* then U will be an *m* x *m* matrix, *Σ* will be an *m* x *n* diagonal matrix, and V^T is an n x *n* matrix. The U matrix contains the representation of

each term within a *concept* space, the V^T matrix contains the representation of each document within the *concept* space, and the Σ matrix represents the overall weight of each concept.

Although LSA has been mainly used in the context of Information Retrieval, returning answers that contain the concept of a user's query if not the exact terms, it has also been used within document summarization. Initially proposed by Gong and Liu, instead of applying the LSA algorithm to terms within documents, it can instead be applied to individual sentences within a document (Gong and Liu 2001). In their work the matrix $D = \Sigma V^T$ describes how much any given sentence represents a topic where $d_{i,j}$ is the weight of concept *i* within sentence *j*.

2.5.3 Natural Language Processing

Natural Language Processing (NLP) attempts to recreate a more human-like understanding of language. Each NLP implementation typically focuses on one or more levels of language processing such as Phonology, Morphology, Lexical, Syntactic, Semantic, Discourse, and Pragmatic (Liddy 2001). When applied to text processing generally only lexical, syntactic, and semantic techniques are considered. This is mainly due to the fact that phonology studies audible sounds, morphology analyzes the compositions of individual words, discourse examines text at a multi-sentence granularity, and pragmatic levels require a level of world knowledge to interpret the context of words.

Lexical and syntactic parsing allows for the identification of words and phrases, parts of speech tagging, and the ability to perform disambiguation tasks. In many cases

approaches frequently make use of domain-specific knowledge sources to help in this regard such as using controlled vocabularies to generate index terms or to produce a summary. For example, within the biomedical area the UMLS Metathesaurus (Lindberg et al. 1993) is used to link similar names for the same concept from nearly 200 different vocabularies allowing systems to recognize that two papers might be talking about the same subject (e.g. one document using the term “dyspnea” while another states “shortness of breath”) (Verma et al. 2007).

Semantic NLP approaches can be more challenging as they attempt to discern the meaning of the text. Not only does this leverage vocabularies to identify known terms but also requires some understanding of the relationships between them. One tool of note, SemRep, extracts semantic predications (subject-predicate-object triples) from raw text (Rindflesch and Fiszman, 2003). SemRep was developed within the biomedical research domain and uses domain knowledge provided by the UMLS, where its generated predications contain “textual content with semantic predications consisting of UMLS Metathesaurus concepts as arguments and UMLS Semantic Network relations as predicates” (Kilicoglu et al. 2008). With the use of the tool, sentences are transformed into any number of assertitional predications. These predications have been subsequently used to build abstractive summarizations in the forms of graphs (Fiszman, Rindflesch, and Kilicoglu 2004). Within the graphs, each vertex represents a concept while the edges represent the relationships between two concepts.

2.6 Literature Based Discovery

LBD is considered a more complex aspect of text mining or information retrieval (Berry and Castellanos 2004; Smalheiser 2012). Not only does it involve identification and extraction of salient information from text but seeks to uncover novel cross-document associations that have never been publicly published (Ganiz et al. 2005; Swanson 2008).

2.6.1 Literature Based Discovery Modes

The majority of LBD approaches are built upon a paradigm now referred to as the ABC discovery model (Swanson 1987; Weeber et al. 2005; Smalheiser 2012). This model states that new knowledge can be discovered between two concepts (A, C) from non-interacting literatures, if hidden associations involving some intermediate concept (B) can be uncovered (Swanson 1986). This is simply the transitive property of relations; in other words, if an object A is associated with a second object B and that object B is associated with a third object C, then it can be inferred that first object, A, is associated with object C in some manner (Figure 3). These "B" concepts are frequently called intermediate terms.

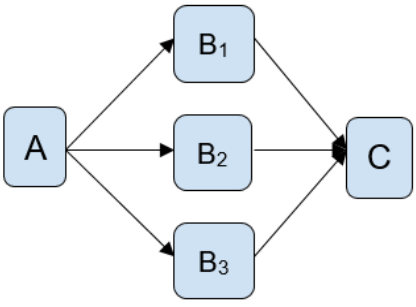


Figure 3. ABC Model

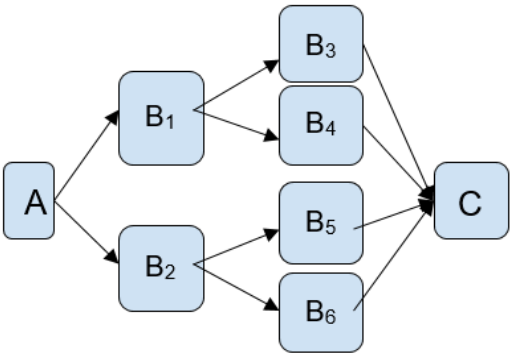


Figure 4. AnC Model

The ABC model was later generalized by Wilkowski et al. where they proposed an extension where several intermediate concepts may be required to connect the A and C concepts (Wilkowski et al. 2011). This extension of Swanson's ABC model is referred to as the AnC model wherein n = $\{B_1, B_2, \ldots, B_m\}$ (Figure 4). The work of this dissertation, and the work by Cameron et al. (Cameron et al. 2015), are based off Wilkowski's AnC model.

Although Swanson's work provided an explanation for his observations it did not propose any framework for finding such connections. As such, subsequent LBD research has begun to investigate better, and more automated, techniques and tools.

The ABC and AnC models have been implemented within LBD research in two modes: open and closed discovery. Open discovery is the embodiment of hypothesis generation and has been implemented in various approaches (Srinivasan 2004; Pratt and Yetisgen-Yildiz 2003; Weeber et al. 2001; Wren et al. 2004; Frijters et al. 2010). It assumes that the user begins with only a single concept (A) from their research questions and they propagate outward via associations to an unknown target concept C by means of one or more intermediate objects (Figure 5).

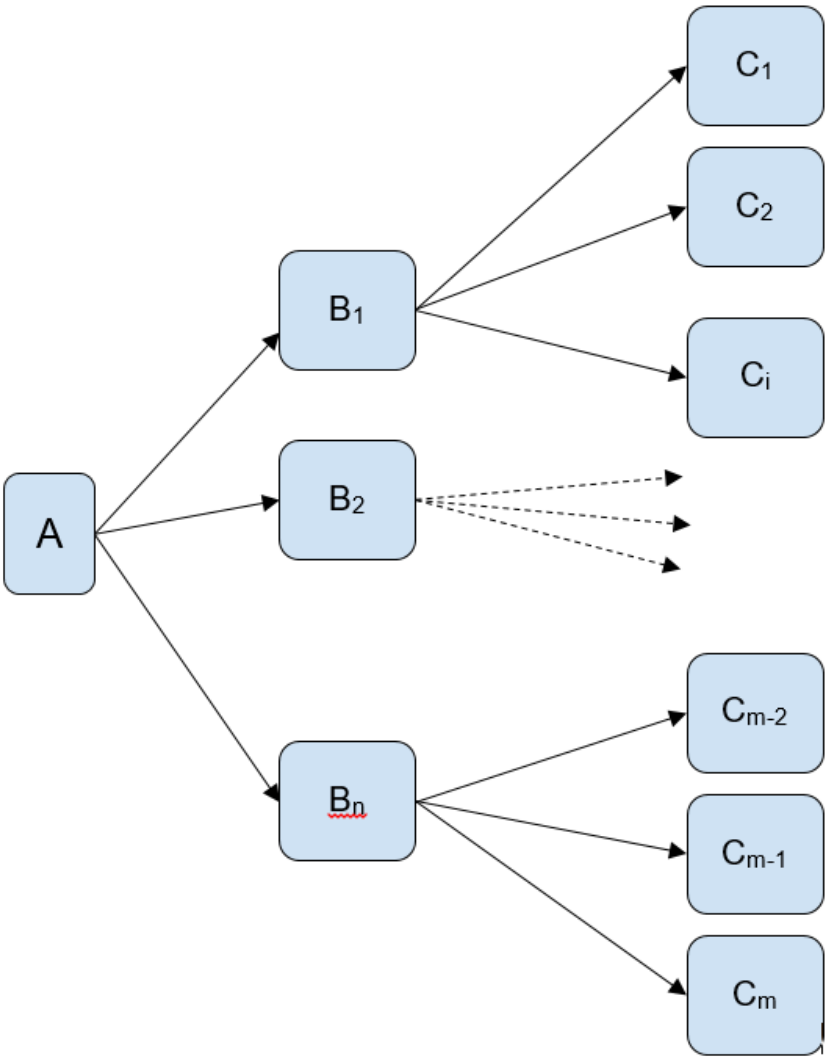


Figure 5. Open Discovery Mode

Closed discovery, in comparison, is characterized by testing one's hypothesis and has also been widely implemented (Smalheiser, Torvik, and Zhou 2009; Srinivasan 2004; Hristovski et al. 2006; Hu et al. 2005; Gordon, and Dumais 1998; Ahlers et al. 2007; Cameron et al. 2013; Cameron et al. 2014; Lindsay and Gordon 1999; Weeber et al. 2001). Within a closed discovery approach, typically both the source concept A and target concept C are known, but not necessarily the intermediate(s). These are instead

discovered via an exploration from both ends to determine potential interesting intermediates (Figure 6).

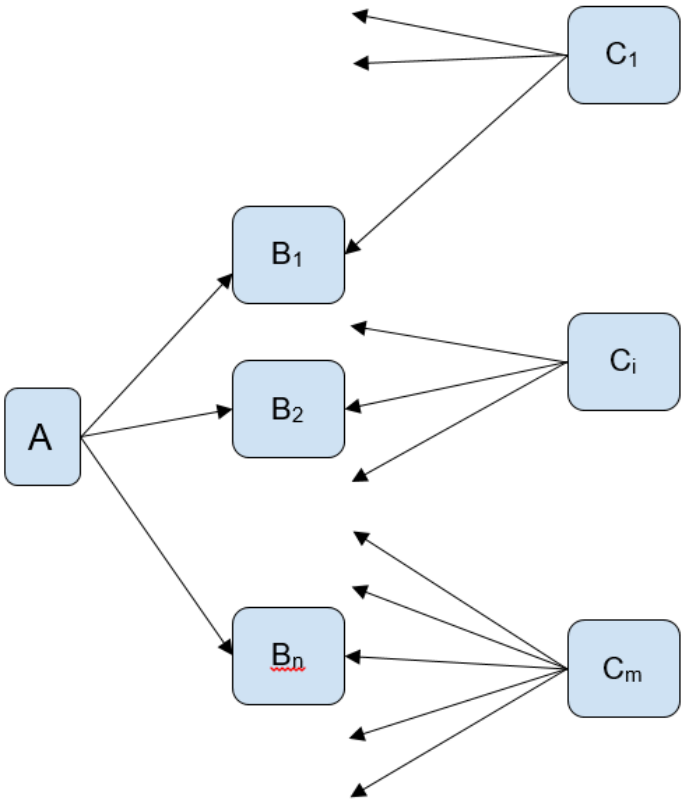


Figure 6. Closed Discovery Mode

Regardless whether an LBD implementation is designed for open or closed discovery the algorithms used to determine intermediate candidates have included a variety of approaches including statistical, knowledge-based, semantic-based, citation-based, and others.

2.6.2 Statistical Literature Based Discovery

The earliest LBD approaches established links between disjoint knowledge by looking for the most frequently co-occurring intermediate terms across documents utilizing statistical frequencies or distributions (Gordon and Lindsay, 1996; Lindsay and

Gordon, 1999; Gordon, Lindsay, and Fan 2002). The conventional wisdom was that discoveries are likely to arise from logical connections between terms or concepts that frequently (co)occur within literature. The intermediate terms are then ranked based on some relative frequency and presented to a researcher. While many of these techniques have been used to successfully discover (Swanson 1986; Swanson 1988), or recover previously discovered ones using a different approach (Gordon et al. 2002; Gordon & Dumais 1998; Gordon and Lindsay 1996; Lindsay and Gordon 1999), several problems exist. The reliance on statistical measures from text will favor co-occurring terms and potentially miss associations from less frequent terms (Kostoff et al. 2009). Furthermore, the statistics will be flawed in regards to Biomedical and Chemical entities that can be expressed in a variety of forms such as chemical names (Krallinger, Rabal, Lourenço, Oyarzabal, and Valencia 2017). In addition, these approaches rely on the researcher's domain expertise and their ability to eliminate noise (terms that are too general or irrelevant) and refine their parameters.

Most existing approaches for LBD use feature extraction to determine what a document is "about". For example, a common approach for this type of textual analysis makes use of counts and frequencies of tokens (words or phrases) as they appear. If a particular token appears frequently within a particular document, it likely signifies importance to the overall meaning of that document. Furthermore, understanding the frequency across a collection of documents is also key. These relative frequencies give an indication of a term's statistical importance (Gordon and Lindsay 1996). A typical metric used is TF-IDF (Sparck Jones 1972). Secondarily, the statistical approach can establish links between different papers by looking for the most frequently co-occurring terms or

concepts in those documents. This is achieved by computing the statistical distributions and frequencies of the terms without consideration of their semantics (Lindsay and Gordon 1999).

Understanding the contextual information of a document may also illustrate an implicit association of concepts. If there is a large number of sources that associate two topics A and B together, and also a great number of sources that associate two topics B and C together. Then, it is plausible there is some association between A and C as well. Some developed strategies and tools to generate semantic associations between terms or compiled manual annotations. (Srinivasan 2004; Yetisgen-Yildiz and Pratt 2006; Ahlers 2007; Frijters et al. 2010) Other work has utilized some of those strategies and direct mining to formulate testable hypotheses in specific areas through a series of manual steps. (Weeber et al. 2003)

The primary limitation of approaches using term co-occurrence is that although it can find intermediate concepts, it does not provide the relationship of those concepts. For example, the concepts of *Reynaud Syndrome* and *platelet aggregation* might be found, but not that *platelet aggregation* causes *Reynaud Syndrome*. In addition, these approaches break down if there is more than one level of separation between the initial and final set of concepts. That is to say that there might be relevant information that exists in longer chains of concepts that are semantically connected.

2.6.3 Knowledge Based Literature Based Discovery

A criticism of purely statistical approaches is that domain expertise is necessary to help refine a search, filter datasets, or extract plausible final results. In many cases it is

entirely left up to the researcher to employ such knowledge when interpreting results. However, within domain-specific knowledge repositories some of this expertise is "baked in" ahead of time so to speak.

The knowledge-based approach gains its strength from the usage of additional knowledge-based resources such as vocabularies, databases, and ontologies. It is also quite common to find the usage of natural language processing as a basis of extraction (Weeber, Kors, and Mons 2005).

Knowledge-based approaches can prioritize terms according to specific semantic types, removing very general ones and targeting others (Hu, Yoo, Song, Zhang, and Song 2005). Thus a more refined set of potential intermediate terms are generated which satisfy a set of researcher's provided predefined semantic types.

An additional benefit that arises from these techniques is that variations in terminology and nomenclature can be recognized - that is to say, synonyms (different terms to designate the same concept) and homonyms (the use of words/phrases with multiple meanings). "For instance, IL-12, IL12, interleukin 12, CLMF, cytotoxic lymphocyte maturation factor, and natural killer cell stimulatory factor all refer to the same concept: Interleukin-12" (Weeber et al. 2003). Purely statistical approaches won't draw connections between the variations, but normalizing extracted terms by means of a thesaurus or ontology allow them to be treated as a single term.

Fundamentally, knowledge sources can be subdivided into two broad categories based on the type of knowledge they contain. These two categories are: 1) definitional knowledge and 2) assertional knowledge.

2.6.3.1 Definitional Knowledge

Definitional knowledge refers to information that can be considered facts within a particular domain. In many cases these sources exist in the form as vocabularies, ontologies, or a database of facts. One such example within the biomedical domain is the UMLS Metathesaurus.

The UMLS Metathesaurus is a large biomedical thesaurus from the National Institute of Health that is organized by concept and links similar names for the same concept from nearly 200 different vocabularies and is comprised of close to 3 million unique concepts.

Figure 7 shows a small portion from this dataset. In this snippet the first segment is the unique UMLS Concept identifier (C#) while other portions provide the corresponding identifier in other vocabularies such as MeSH, represented in these examples as D#. For example, in the figure *Raynaud Disease* is identified by C0034734 and *Raynaud Phenomenon* is identified by C0034735 within the MetaThesaurus but are both associated with the same MeSH concept D011928.

```
C0034734|ENG|S|L1446659|PF|S0974677|N|A3500400||M0018534|D011928|MSH|MH|D011928|Raynaud Disease|0|N||
C0034735|ENG|P|L1970983|PF|S0974685|N|A3486562||M0449202|D011928|MSH|PEN|D011928|Raynaud Phenomenon|0|N||
```

Figure 7. Sample entries from the UMLS Metathesaurus

A second frequently used source is Medical Subject Headings (MeSH), which is a controlled vocabulary of biomedical terms from over 5000 biomedical journals organized into 16 hierarchical structures - each belonging to a different category such as Anatomy, Organisms, Diseases, Chemicals, and others. MeSH has been utilized within LBD research to map text to known structured knowledge and to provide implicit context between co-occurring concepts. It also has been widely used to index, catalog, and search

MEDLINE articles. To accomplish this, MeSH descriptors are manually assigned to scientific articles by domain experts. “The quality of these assignments is considered high and relatively good indicators of the semantics of the content of the article to which they are assigned” (Cameron 2014).

2.6.3.2 Assertional (Semantic) Knowledge

Assertional knowledge refers to statements, predicates, and assertions that have been mined from scientific literature. Within the biomedical domain, processes are “inherently composed of interactions between various types of entities” (Bakal, Talari, Kakani, and Kavuluru 2018). Usually these statements are captured by an NLP tool, such SemRep from the National Institute of Health, as binary interactions connecting a subject entity to an object entity by means of some relation or predicate (subject-relation-object). SemRep matches subject and object terms with UMLS Metathesaurus concepts and the predicates from a list of types within the UMLS Semantic Network. An example, shown in Table 2, where the tool extracts a list of predicates from the given source text.

Table 2. SemRep generated predicates from example source text

Source Text: “We used hemofiltration to treat a patient with digoxin overdose that was complicated by refractory hyperkalemia”
Extracted predicates: ● Hemofiltration-TREATS-Patients ● Digoxin overdose-PROCESS_OF-Patients ● hyperkalemia-COMPLICATES-Digoxin overdose ● Hemofiltration-TREATS(INFER)-Digoxin overdose

SemMedDB is a repository by the NIH consisting of semantic predications extracted from the titles and abstracts of Medline documents by the SemRep tool.

2.6.4 Semantic Based Literature Based Discovery

Traditional knowledge-based implementations that utilize definitional knowledge can provide improvements over standard statistical approaches, but they do not provide insights into the meaning of these associations. For example, in regards to the treatment of diseases, "drug therapies are often used effectively, even though the exact cause of action may be either poorly understood or unknown" (Ahlers, Hristovski, Kilicoglu, and Rindflesch 2007). Although providing a new novel association is 'better than nothing', LBD systems should strive to offer insight into the meaning of relationships as well.

Several attempts have been made to employ semantic predicates to perform automatic abstractive summarization (Fiszman and Rindflesch 2004; Plaza, Díaz, and Gervás 2011). These approaches have chosen to represent their summaries using concepts rather than words and presenting the result as an interconnected graph. Traditionally graph-based summarization methods usually represent the documents as graphs where the vertices correspond to text units such as words, phrases, or sentences, and the edges represent a form of similarity (Plaza, Díaz, and Gervás 2011).

In 2015, Cameron et al. devised a context-driven LBD approach based upon the earlier graph-based semantic summarization methods. Their method was a closed-discovery approach: given a pair of concepts as user input their system automatically generates a graph representing the semantic associations between the two concepts. These graphs are then used to find semantically related pathways from one concept to another.

One of the main strengths of this technique is the ability to automatically extract subgraphs that allow researchers to interpret the meaning of the displayed predication pathways. In the study by Cameron et al. subgraphs were constructed with the use of a hierarchical agglomerative clustering algorithm to group edges based upon their similarity. For example one subgraph might illustrate how *Fish Oils* is connected to *Raynaud's Disease* by way of *Blood Viscosity* effects, while a second subgraph illustrates this connection via *Platelet Aggregation* (Cameron et al. 2015).

2.6.5 Citation Based Literature Based Discovery

Bibliometric analyses in the field of scientometrics have shown that "the most productive authors are the ones engaged in the most innovative and varied research [and] in addition, these researchers draw from a larger pool of knowledge, indicating that they may be moving more often from topic to topic." (Milojevic 2012).

Studying the structure of how documents are interrelated is usually is focused on citations. Several examples of which include direct citations, bibliographic coupling, and co-citations. A direct citation is the simplest case where paper A cites paper B. Bibliographic coupling link exists between two papers A and B when each cites a common paper C. Lastly a co-citation link exists between two papers A and B when a third paper C cites both of them. Additionally, there are other possible ways to consider two references to have some type of link either involving citations or by other factors such as having authors in common.

Kostoff was the first to explore bibliographic coupling within the field of LBD.

Kostoff's approach begins like many other LBD algorithms using a pair of terms (A,C) and retrieving all the documents related to each. He then augments the list of records to include additional bibliographically coupled documents before performing some statistical based LBD algorithms. Results of his research indicate that shared references between two disjoint literatures could harbor many useful linking terms. The primary drawback of the technique is the labor intensive procedure and lack of automation (Kostoff 2014).

2.7 Citational Coupling

Zhang, Ding, and Milojević state that it "has become a convention for scientists and researchers to refer to earlier work that relates to, inspires or is used by their own work" (Zhang, Ding, and Milojević 2013). This is in reference to an earlier document by Nicolaisen wherein certain basic assumptions are made about citations. That is, the citation of a document (1) implies the use of that document and (2) reflects the merit of that document (Nicolaisen 2007).

Sentences that contain a citation to another document are typically referred to as citation sentences, or citances. These sentences typically highlight the most important aspects of the paper being cited. This includes the research problem it addresses, a methodology, or the results. (Abu-Jbara and Radev 2011). The author of an article writes their abstract as a summarization to convey the central ideas from their perspective. Research has shown that citing sentences are a "collaborative summary that indicates what other researchers found relevant, interesting, or novel about the article" (Elkiss et al. 2008). In bioscience literature it is common for the text around citations to state known

biological facts with reference to the original papers that discovered them (Nakov, Schwartz, and Hearst 2004).

Some studies have shown that the "information contained in the abstract and in the citances overlap to a large extent" (Divoli 2012). To that end efforts have been put into creating automatic summarization by using the citation sentences to build a second "abstract" for a given document constructed from all the papers that have cited it. This overlap between the meaning of the citation sentences and the abstracts of the cited articles goes both directions. Many times full-text is not necessarily available, and only titles and abstracts can be utilized. In these situations, it stands that the abstracts of cited references can be used as an indicator of a document's body.

2.8 Summary

Regardless of their implementation (i.e. statistical, semantic, etc.), LBD approaches are a blend of text processing techniques including text mining, information retrieval, question answering, and text summarization. Each seeking to understand and extract salient details about documents to present the most likely connected information. Studies into summarization techniques have shown that strong connections between documents and those they cite such that they are semantically related. This paper therefore theorizes that bibliographic citations can improve LBD implementations when incorporated.

Chapter 3

Methodology

3.1 System Design

The foundation for the proposed work relies on replicating the research performed by Cameron and then incorporating citationally related documents to improve the results. The foundation portion was completed first going so far as to produce a single graph of all uncovered predicates between two concepts focusing on the *Fish Oils - Raynaud Disease* study. This work was further extended wherein cited documents were added to the corpus and citational information was included as an initial study to determine if the hypothesis had merit. Due to the positive initial results subsequent retrospective experiments were conducted.

A standard three tier architecture is utilized to process the requests (Figure 8). Upon receiving the request from the client, the web tier constructs an appropriate internal query for our in house search engine. The distributed search engine executes the queries (Apache Solr and otherwise) across multiple instances over an HDFS cluster and returns the query results.

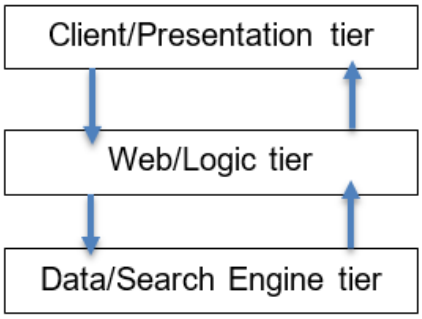


Figure 8. Three tier architecture

Chemical Abstracts, a division of the American Chemical Society, has multiple software products that allow researchers to search patents, journals, and other sources for chemically related information, including MEDLINE. Forms were added in an existing product's user interface where a researcher simply types in the concept(s) of interest.

3.2 Query Specification

Users need some method of querying the LBD system. For closed discovery approaches they will need to specify at least one 'A' concept and at least one 'C' concept for the AnC methodology described earlier. Whereas for open discovery only a single concept is needed as a starting point to explore.

3.2.1 Query Specification (Open)

For open discovery, our interface allows the user to select a single concept from within the corpus and the initial predicate type of interest such as *affects*, *causes*, *treats*, and others (Figure 9).

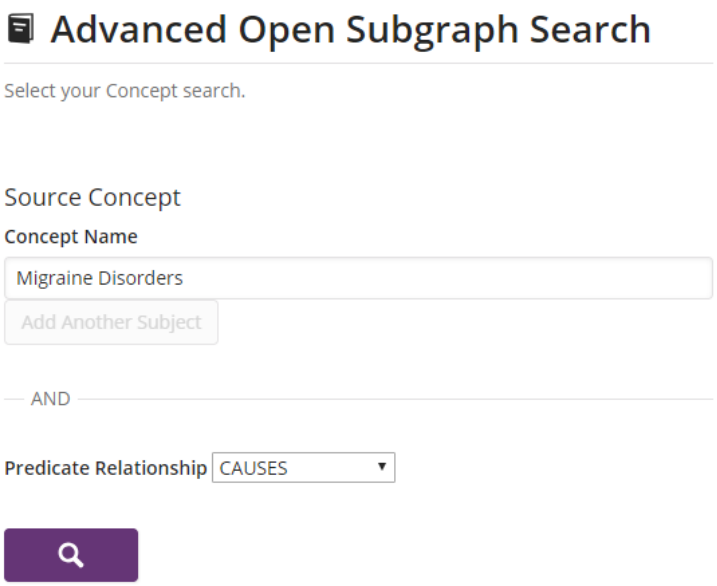


Figure 9. Example open subgraph query

3.2.2 Query Specification (Closed)

Following the AnC pattern the design of our closed discovery system requires a query in the form of two sets of concepts. Our user interface for the closed query specification can be seen in Figure 10. The first set (Source) corresponds to the "A" set, and the second (Destination) to the "C" set. There are also additional configuration options to specify the desired path length, if citational information should be used, and if the user wishes to only return paths that pass through a particular intermediate.

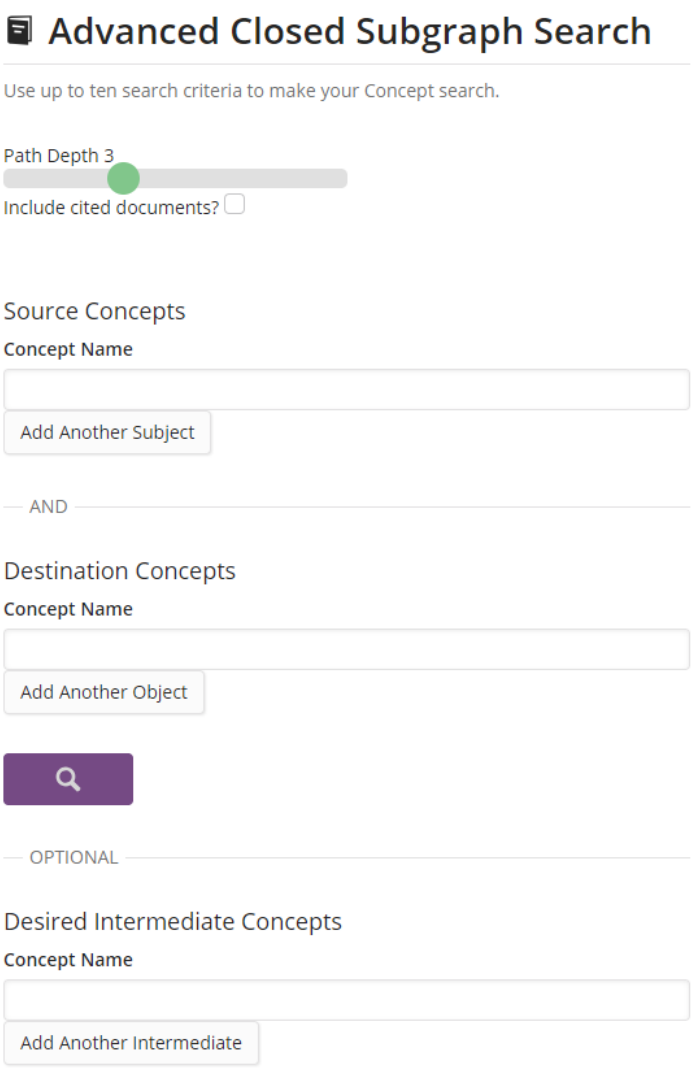


Figure 10. Example closed subgraph query

3.3 Preprocessing

Cameron's closed discovery implementation (Cameron et al. 2015) is the basis for our approach, where the essential idea is to search all documents within the corpus that contain at least one of the terms provided (MeSH or UMLS) and to use semantic predications from those documents to build graphs. To accomplish this we need to have several distinct types of data: 1) we need to process a collection of MEDLINE documents and identify what concepts are contained within those documents, 2) we need to process

document citations to expand our document results, and 3) we need to identify what predicates are contained within those documents.

All MEDLINE (currently over 29 million) documents along with their basic information were already available at Chemical Abstracts in the form of xml files available via HDFS. This includes the title, abstract, authors, and indexed MeSH concepts as defined within PubMed. Documents within PubMed are only abstracts and, in general, do not contain information on their references. According to the National Institute of Health 37% of records added in the last five years contain full text (stored in PubMed Central) that is freely available. In these cases, the cited references were also available within the xml records, unfortunately, for our studies we found that less than one in ten fit this category. As such, when citational associations were not available this information was added manually for each study. Additionally, in a small number of cases the PubMed record was also missing an abstract - for these, the abstracts were gathered from Publisher websites. This is indicated appropriately for each document within the appendix corresponding to each research study.

The predicate information was acquired from the National Institute of Health as an export of the SemRep tool. It is periodically run against the entire MEDLINE database which contains titles and abstracts, and the download is provided via their website. The export can be placed into the form of csv text files for processing. In cases where the abstract was missing from the MEDLINE database, the tool was run manually on the abstract/introduction when found and the information added. Figure 11 shows a small portion from this dataset, where each entry identifies a unique relationship between two

concepts from the UMLS Metathesaurus and which document and sentence it was derived from.

```
"5306041","32399885","1","1296760","STIMULATES","C0024467","Magnesium","bacs","true","C
0033567","Epoprostenol","bacs","true"

Field 1: "5306041" - Auto-generated primary key for each unique predication
Field 2: "32399885" - Foreign Key to a table of sentence information
Field 3: "1" - The number of times the predication is extracted from the sentence
Field 4: "1296760" - Medline Document ID
Field 5: "STIMULATES" - Predicate Semantics
Field 6: "C0024467" - Subject UMLS Concept ID
Field 7: "Magnesium" - Subject Concept Name
Field 8: "bacs" - Subject Semantic Type
Field 9: "true" - Subject Novelty based on distance from root in UMLS Metathesaurus
Field 10: "C0033567" - Object UMLS Concept ID
Field 11: "Epoprostenol" - Object Concept Name
Field 12: "bacs" - Object Semantic Type
Field 13: "true" - Object Novelty based on distance from root in UMLS Metathesaurus
```

Figure 11. Sample entry from the SemRep DB export

In our implementation, as is typical in search engines, we make use of an offline compilation process to pre-process our data (Figure 12). Using Hadoop/Cascading we extract data from the formatted text (XML or CSV), parse the information, and then treat individual records as tuples. To be able to execute our queries, we then need to identify the classifications of data and how they are associated with each other. Thus, we need to identify documents, MeSH concepts, UMLS concepts, predicates, and their underlying relationships. By constructing this information during offline compilation we can build necessary solr indexes for searching and pre-defined mapping relationships such as "Document A cites Document B" which allow the queries to project from a given result set to another. Later, using our internal syntax we can perform an online solr query for a particular concept identifier then cross reference for all documents related to the results of the query.

```
// Extracts Tuples from source files
Pipe referencePipe = new ReferenceExtraction(referenceInput); //each Reference
Pipe predicatePipe = new PredicateExtraction(predicateInput); //each Predication
Pipe meshPipe = new MeshExtraction(meshInput);                //each MeSH Concept
Pipe umlsPipe = new UmlsExtraction(metaThesaurusInput);       //each UMLS Concept

/* Calculates the Context for each Reference. That is, the MeSH Tree #s of all MeSH
terms for that reference.*/
Pipe referenceContextPipe = new ReferenceContext(referencePipe, meshPipe);

/* Each UMLS Concept already has its MeSH Concept ID from the MetaThesaurus entry where
it was extracted if it exists. However, here we merge in the corresponding MeSH
TreeNumbers.*/
umlsPipe = new UmlsMeshMergeAssembly(umlsPipe, umlsPipe);

//Calculate the IDF for each UMLS Concept as they appear from predications.
umlsPipe = new CalculatePredicateIDF(umlsPipe, predicatePipe);

/* Determine the context of a predicate as the summation of the reference contexts from
all references where that predicate has been extracted.*/
predicatePipe = new PredicateContextAssembly(predicatePipe, referenceContextPipe);

//For each predicate merge in the MeSH TreeNumbers and UMLS Concept IDF scores.
predicatePipe  = new PredicateMeshAssembly(predicatePipe, umlsPipe);

//Construct solr Searchable/Indexed objects for query-time searching later.
Pipe predicateIndexedEntry = new ConstructPredicateEntity(predicatePipe);
Pipe referenceIndexedEntry = new ConstructReferenceEntity(referencePipe);
Pipe umlsIndexedEntry = new ConstructUmlsEntity(umlsPipe);
Pipe meshIndexedEntry = new ConstructMeshEntity(meshPipe);

//Construct relationships to allow projections between object types.
Pipe umlsToMeshRelationshipPipe = new ConstructUmlsToMeshRelationship(umlsPipe);
Pipe predToDocRelationshipPipe = new ConstructDocToPredRelationship(predicatePipe);
Pipe predToUmlsRelationshipPipe = new ConstructUmlsToPredRelationship(predicatePipe);
Pipe meshToDocRelationshipPipe = new ConstructMeshToDocRelationship(referencePipe);
Pipe DocToDocRelationshipPipe = new ConstructCitationRelationship(referencePipe);
```

Figure 12. Example Hadoop code to prepare searchable digest

3.4 Closed Search Algorithm

Closed search is the most common among LBD implementations (Ahlers, Hristovski, Kilicoglu, and Rindflesch 2007; Cameron et al. 2013; Gordon and Dumais 1998; Hristovski, Friedman, Rindflesch, and Peterlin 2006; Lindsay and Gordon 1999; Smalheiser, Torvik and Zhou 2009; Srinivasan 2004), namely because it applies the most constraints and, as such, makes the results more narrow and meaningful. Our general approach follows the same basis as described by Cameron et al in their paper (2015).

- Step 1 - User specifies “A” and “C” sets of concepts, a maximum path length k, and a date Dt.
- Step 2 - Define the set D as all pubmed IDs published before Dt which contain any concept from set A or set C.
- Step 3 - Define G and R as an empty graphs and the set S as empty.
- Step 4 - For each pubmed id d ∈ D, add all predicates from d as edges into the graph G.
- Step 5 - For all pairs of concepts $(a,c) \in A \times C$, determine all possible paths between concept *a* and concept *c* of maximum length k inside G. All these to R.
- Step 6 - Divide the graph R into smaller subgraphs, adding each into S.
- Step 7 - Sort the collection S of subgraphs by ranking.

Next we discuss our approach to the closed query formulation, leaving a discussion regarding subgraph generation and scoring algorithms for later in the chapter.

3.4.1 System Query Construction

Our query process is summarized in the steps below, diverging from a generalization of Cameron’s algorithm only by adding steps 3 and 4 where the list of documents may optionally be expanded to also include any additional references that the initial set has cited. For each resulting document in step 5, we want to collect all defined semantic predicates.

- Step 1 - User specifies “A” and “C” sets of concepts.
- Step 2 - Define the set S as the set of all documents within the corpus that contain at least one of the terms.
- Step 3 - (Optional) Define the set S’ as the set of all documents cited by documents within the set S.
- Step 4 - Define the resulting set of documents S’’ = S ∪ S’
- Step 5 - Define the set of predicates P as the set of all predicates defined for each document in S’’
- Step 6 - Create resulting graphs from the predicates within P.

Steps one through five are formulated within a system query. Using the *Fish Oil-Raynaud Syndrome* study as an example, Figure 13 shows the portion of the system query which searches for the UMLS and MeSH concepts.

```
query = 'UMLS_CONCEPT_NAME: "Fish Oils" | UMLS_CONCEPT_NAME: "Eicosapentaenoic Acid" |
UMLS_CONCEPT_NAME: "Raynaud Disease" | UMLS_CONCEPT_NAME: "Raynaud Phenomenon" '
exploreQuery("subgraphTextQueryId") { solr(q: query) }
//Search UMLS Concepts for the above query terms
entity("umlsconcept") {
  identifier 'UMLS_QUERY_ID'
  search("solr") { queryId subgraphTextQueryId" }
}
//Find corresponding MeSH terms for the provided search concepts (UMLS)
entity("mesh") {
  identifier 'MESH_QUERY_ID'
  constraint {
    type "PROJECTION"
    projection {
      relationshipName "umls_to_mesh"
      sourceIdentifier "UMLS_QUERY_ID"
    }
  }
}
```

Figure 13. Step 1 – Search for concepts

When executed, the results return a total of four UMLS concepts, and three MeSH concepts. This is because both *Raynaud Disease (C0034734)* and *Raynaud Phenomenon*

(C0034735) within the UMLS are mapped to the single MeSH concept *Raynaud Disease (D011928)* (Figure 14).

```
Retrieved umlsconcept results: (4)
[UMLS_CONCEPT_ID:[C0000545], UMLS_CONCEPT_NAME:[Eicosapentaenoic Acid],
MESH_ID:[D015118], INV_DOCUMENT_FREQ:[8.333311917991818]]
[UMLS_CONCEPT_ID:[C0016157], UMLS_CONCEPT_NAME:[Fish Oils], MESH_ID:[D005395],
INV_DOCUMENT_FREQ:[8.1199863723497 22]]
[UMLS_CONCEPT_ID:[C0034734], UMLS_CONCEPT_NAME:[Raynaud Disease], MESH_ID:[D011928],
INV_DOCUMENT_FREQ:[9.758878686545028]]
[UMLS_CONCEPT_ID:[C0034735], UMLS_CONCEPT_NAME:[Raynaud Phenomenon], MESH_ID:[D011928],
INV_DOCUMENT_FREQ:[8.87448928343275]]

Retrieved mesh results: (3)
[DESCRIPTOR_IDENTIFIER:[D005395], DESCRIPTOR_NAME:[Fish Oils],
MESH_TREE_NUMBER:[D10.627.430]]
[DESCRIPTOR_IDENTIFIER:[D011928], DESCRIPTOR_NAME:[Raynaud Disease],
MESH_TREE_NUMBER:[C14.907.617.812]]
[DESCRIPTOR_IDENTIFIER:[D015118], DESCRIPTOR_NAME:[Eicosapentaenoic Acid],
MESH_TREE_NUMBER:[D10.212.302.380.410.385    D10.251.355.255.200    D10.251.355.337.290
        D10.627.430.450.390]]
```

Figure 14. Results of step 1 system query

Although MEDLINE documents have MeSH concepts assigned to them directly, content curators will only mark those concepts which are deemed significant. This is in contrast to SemRep predications that use UMLS concepts, which are extracted via NLP. As either approach might miss the presence of a topic, the second step requires fetching the appropriate references for our search terms from either of these initial sets. The query is a simple projection (Figure 15).

```
//Find which predicates contain the given UMLS Concepts
entity("predicate") {
  identifier 'PREDICATE_QUERY_ID'
  constraint {
    type "PROJECTION"
    projection {
      relationshipName "predicate_to_umls"
      sourceIdentifier "UMLS_QUERY_ID"
    }
  }
}
//Find set of references S which contain EITHER the MeSH term or a predicate
//which contained the UMLS Concept.
entity("reference") {
  identifier 'REFERENCE_QUERY_ID'
  constraint {
    type "PROJECTION"
    projection {
      relationshipName "reference_to_mesh"
      sourceIdentifier "MESH_QUERY_ID"
    }
  }
  constraint {
    type "PROJECTION"
    projection {
      relationshipName "predicate_to_document"
      sourceIdentifier "PREDICATE_QUERY_ID"
    }
  }
  or()
}
```

Figure 15. Step 2 – Search for the set S of references

For each retrospective study we collected together the MEDLINE articles cited by the original LBD paper. We will refer to these as core papers as they were directly cited by the LBD researcher. Then, for each core paper, all cited MEDLINE papers were also added to our corpus. Although each of the studies focused on different areas, some overlap was found. This is due to both the inherent interconnectedness of biomedical science as well as later LBD papers citing earlier work.

Our example query results in a total of 193 documents from our corpus (Figure 16). Those marked in red are from the *Fish Oils-Raynaud Disease* study and comprise the majority of the hits as expected. Those in blue come from a study around *Magnesium-Migraine Disorders*, those in green come from a study about the *Obesity Paradox*, and

those in yellow are from the *Somatomedins-Arginine* papers. Additionally, bold pubmed id numbers are from a core paper and those in italics were cited from a core paper but are directly hit by the search terms.

6109	*561947*	*2983350*	*4106870*	**6097237**	*6166048*	**6300604**	**6376801**	**6812750**	*7362667*
53042	*591484*	*2985986*	*4121220*	**6098049**	*6168223*	**6301111**	*6377247*	*6816332*	**7364865**
58309	*594750*	*2989401*	*4218716*	**6098051**	*6192302*	**6302714**	*6383036*	*6821892*	*7378283*
72898	*616031*	*2996169*	*4327362*	**6100033**	**6209510**	**6303363**	*6383760*	**6890719**	*7427564*
73104	*636997*	**2997286**	*4544972*	*6102181*	**6229551**	**6307322**	**6432198**	*7025341*	*7435433*
89498	*661616*	*3010003*	*4604746*	*6104008*	*6247744*	*6308046*	**6449756**	**7031981**	*7442530*
121610	*698554*	*3030173*	*4693099*	*6104986*	*6247778*	**6314583**	*6469703*	**7037038**	*7459607*
215027	*760690*	*3115623*	*4740483*	*6106100*	*6247894*	*6316965*	**6540787**	*7039582*	*12854830*
218223	*762741*	**3157318**	*4792656*	*6107739*	*6252872*	**6318123**	*6540986*	*7082969*	*13851885*
277163	*849637*	*3551081*	*4792657*	*6114257*	*6266305*	**6320840**	**6613908**	**7161779**	*13884560*
376042	*878907*	*3711250*	*4808710*	*6117898*	*6266735*	**6320945**	*6617097*	*7195627*	*14061587*
388720	*881094*	**3724238**	*5006758*	*6118628*	*6278536*	**6321621**	**6636033**	*7208950*	**14289442**
419455	*884936*	**3797213**	*5083413*	*6123019*	*6281800*	*6329189*	**6639916**	*7212523*	*16963250*
428147	*952739*	**3883365**	*5090397*	**6128596**	*6281802*	**6330926**	**6707529**	*7240991*	
465096	*1069432*	**3888229**	*5410395*	**6130329**	*6285445*	*6337664*	*6712540*	**7259326**	
471766	*1135634*	**3977414**	*5414890*	*6131248*	*6293041*	**6339921**	**6775675**	*7263863*	
519354	*1162080*	**3990714**	*5488384*	**6134122**	*6294902*	*6339937*	*6782927*	*7275566*	
531220	**2412071**	**4002440**	*5547282*	*6135105*	*6295686*	**6340424**	**6788326**	*7284246*	
550156	**2536517**	**4015748**	*5791479*	**6136879**	**6298902**	**6352267**	*6803303*	**7295999**	
554069	*2857265*	**4082084**	*6087007*	*6158764*	*6298968*	**6365102**	*6810413*	*7311739*	

Figure 16. Results of step 2 system query

Later, in chapter four, when we discuss each retrospective study we will examine the effectiveness of closed discovery and ensure that no results are skewed to this overlap.

The next step of our process gathers all of the cited references from these 193 documents. The query is shown in Figure 17 and yields a total of 520 documents within our corpus.

```
entity("reference") { //The set of references S' = cited by the set of references S.
  identifier 'CITED_REF_QUERY_ID'
  constraint {
    type "PROJECTION"
    projection {
       relationshipName "reference_to_cited_reference"
       sourceIdentifier "REFERENCE_QUERY_ID"
    }
  }
}
```

Figure 17. Step 3 – Search for the set S' of cited references

In total this makes 561 unique references as many papers have overlapping references. The fifth step of the query takes this list of references and collects the predicates that were extracted from those documents via SemRep (Figure 18). For our example study this amounted to 2097 predicates.

```
entity("predicate") { //Find the set of unique predicates that come from either S or S'
  identifier 'SUBGRAPH-QUERY-RESULTS'
  constraint {
    type "PROJECTION"
    projection {
      relationshipName "predicate_to_document"
      sourceIdentifier "REFERENCE_QUERY_ID"
    }
  }
  constraint {
    type "PROJECTION"
    projection {
      relationshipName "predicate_to_document"
      sourceIdentifier "CITED_REF_QUERY_ID"
    }
  }
  or()
}
```

Figure 18. Steps 4 and 5 – Search for predicates from the union of S and S'

3.4.2 Closed Graph Generation and Display

Our system query results will be a list of all predicates that have been derived from the documents related to the original query. For closed discovery, graph algorithms are then applied to distill all possible paths up to the specified length between the concepts of the A set and the concepts of the C set. Then, one or more directed predications graphs are created in which nodes are UMLS concepts and edges are UMLS predicates. An example graph can be seen in Figure 19.

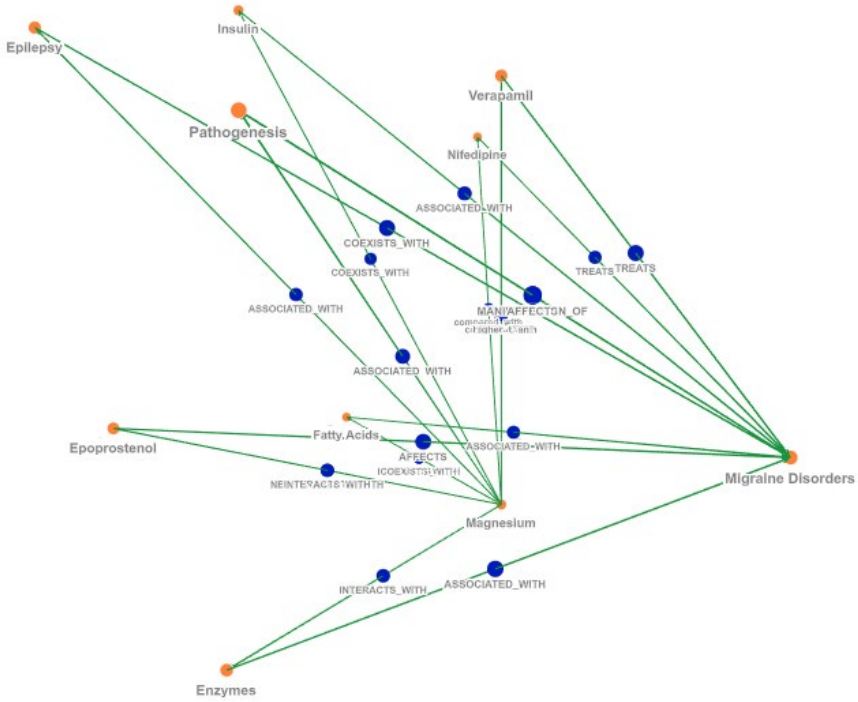


Figure 19. All paths of length two between Magnesium and Migraine Disorders

For ease of implementation, the points of the graph are distributed randomly in a three dimensional box with edges drawn between their corresponding vertices. This was accomplished using the THREE.js JavaScript framework. The image can be rotated, zoomed, and navigated to help fully understand all of the predicate relationships. This

particular graph is a candidate graph containing all possible paths between the two sets. Each graph edge represents one or more predicates that associate two concepts. The blue dots along the edge represent each type of connection such as *treats*, *causes*, *isa*, etc. Clicking on one of the blue dots navigates to a new page with every document that generated the predicate in question. Additionally when there are multiple predicate types that connect two concepts, multiple blue dots may appear.

It should be obvious that representing all of the resulting paths within a single graph will quickly become unusable with larger numbers of concepts and graph edges. As such, it is necessary to divide this resulting graph into multiple subgraphs and impose criteria to rank the results. This will be discussed later in this chapter.

3.5 Open Search Algorithm

The closed discovery process is based around having some hypothesis to be explored. This creates the initial search concepts which generate the resulting connection graphs. At that point the user is left to explore the concepts which appear and the listings of references which produced the predicate graph edges. An open discovery approach needs to be more dynamic and flowing as its purpose is to help uncover possibilities. As such we find that there needs to be a difference between the initial search starting the open discovery and the subsequent exploratory queries.

3.5.1 Initial Open Discovery System Query Construction

Our exploratory approach, or open discovery, is started where a user specifies a single concept. However, this could prove to be intractable by itself as there are likely to

be a large number of other concepts that appear in a predication with the selected concept. Thus, the user must also specify an initial relationship of interest, such as *treats* or *causes*. The initial search is described as follows:

- Step 1 - User specifies a single concept C and relationship R.
- Step 2 - Define the set of predicates P as the set of all predicates that have the relationship R and where one of the concepts is C.
- Step 3 - Create resulting display graph from the predicates within P.

Here, as we are interested in predicates that specifically contain the provided concept and relationship there is no need to include additional documents via citational association or to gather up other predicates within the documents. Using the same example as before, perhaps we wish to start our exploration for things that are known to *treat* the condition known as *Raynaud Disease*. The query for this is simple and is shown in Figure 20.

```
query = '(SUBJECT_NAME: "Raynaud Disease" | OBJECT_NAME: "Raynaud Disease") &
(PREDICATE_DESCRIPTOR: "TREATS") '
exploreQuery("subgraphTextQueryId") { solr(q: query) }
//Search Predicates directly for the above query terms
entity("predicate") {
  identifier 'SUBGRAPH-QUERY-RESULTS'
  search("solr") { queryId subgraphTextQueryId" }
}
```

Figure 20. Open search query for predicates

The results of the query will be a series of predicates, each with one of the concepts as the specified search concept, *Raynaud Disease*.

3.5.2 Open Graph Generation and Display

The resulting predicates are shown directly connected to the user's single search concept (Figure 21).

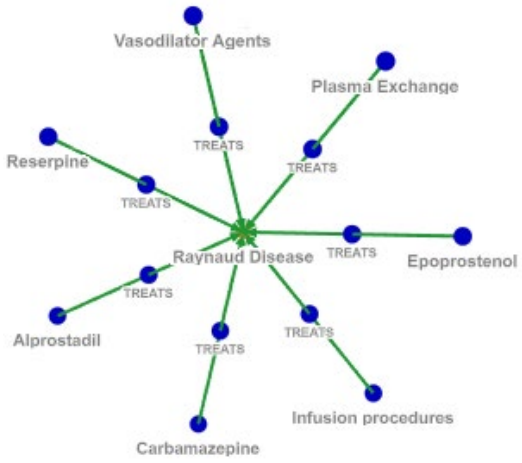


Figure 21. Example open discovery for Raynaud Disease

Unlike the closed discovery graphs, here, concepts can also be selected to further exploration. If the user starts their open discovery with *Raynaud Disease*, they can then select *Epoprostenol* as an exploration target. It then makes sense that the system will then display *Epoprostenol* in the center along with a new arrangement of related concepts (Figure 22). A breadcrumb trail is displayed showing the user their exploration path, also providing a quick method to return back to a previous concept or conduct a closed discovery search between their initial search concept and the current one.

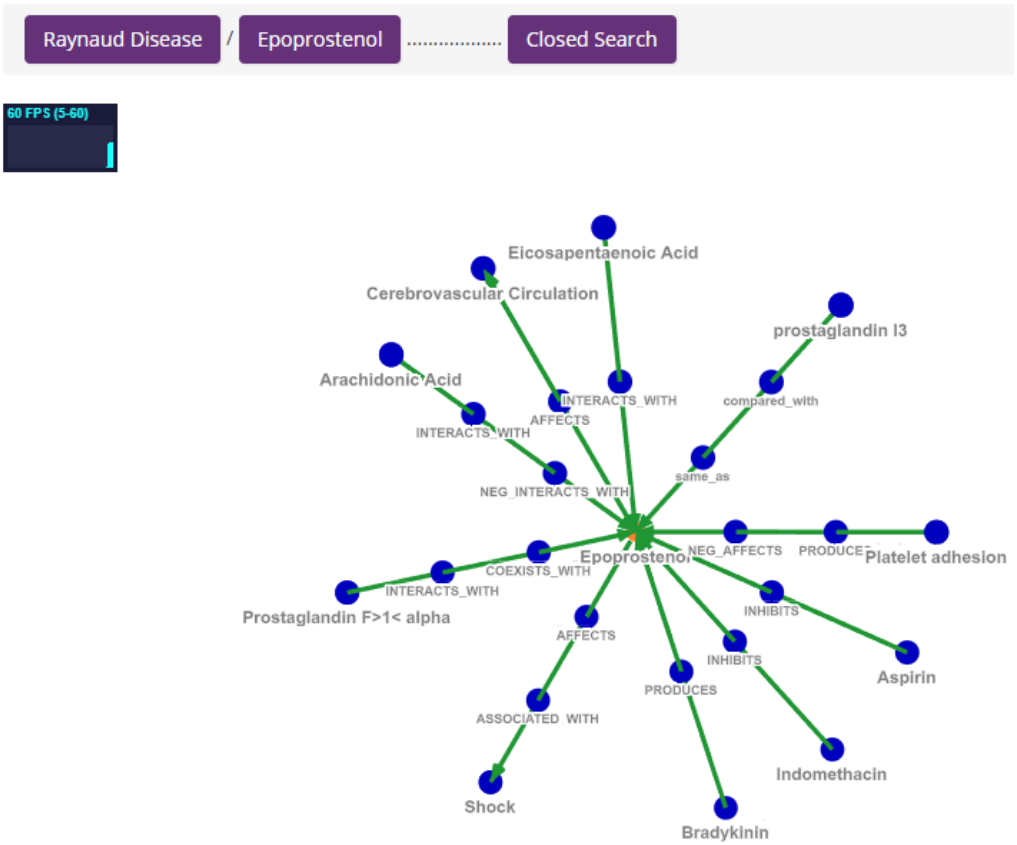


Figure 22. Open discovery for Raynaud Disease –> Epoprostenol

3.5.3 Subsequent Open Discovery System Query Construction

Subsequent exploration queries could continue to search for "all" predicates containing a single concept each time. However, this will be little coherence between the results from one search to the next subsequent search. By selecting a subsequent concept the user is essentially exploring the connection to that concept, not just the concept. Thus, it is important to take into consideration the result context when performing the next search. Our algorithm is as follows:

- Step 1 - Define the concept C’ as the previously selected concept within this open discovery exploration.
- Step 2 - User selects concept C within results of an open discovery search.
- Step 3 - Define the set S as the set of all documents within the corpus that contain a predicate that associates C with C’.
- Step 4 - Define the set S’ as the set of all documents cited by documents within the set S.
- Step 5 - Define the resulting set of documents S’’ = S ∪ S’
- Step 6 - Define the set of predicates P as the set of all predicates defined for each document in S’’.
- Step 7 - Define the set of predicates P’ as the set of all predicates within P where one of the concepts is the selected concept C.
- Step 8 - Create resulting display graph from the predicates within P’.

The initial search (Figure 23) looks for predicates that connect concept C and C’ together.

```
query = '(SUBJECT_NAME: ""Epoprostenol" & OBJECT_NAME: "Raynaud Disease") |
(SUBJECT_NAME: "Raynaud Disease" & OBJECT_NAME: "Epoprostenol") '
exploreQuery("subgraphTextQueryId") { solr(q: query) }
//Search predicates for the above query terms
entity("predicate") {
  identifier 'BASE_PREDICATE_QUERY_ID'
  search("solr") { queryId subgraphTextQueryId" }
}
```

Figure 23. Steps 1-2 – Search for predicates that connect concept C’ and C

The rest of the query steps follow a similar series of steps to our closed discovery implementation, collecting the predicates from the documents and cited documents. These can be seen together in Figure 24.

```
//Find the set of documents S that contain the predicates
entity("reference") {
  identifier 'REFERENCE_QUERY_ID'
  constraint {
    type "PROJECTION"
    projection {
      relationshipName "predicate_to_document"
      sourceIdentifier "BASE_PREDICATE_QUERY_ID"
    }
  }
}
//Find the set of references S' which are cited by the set of references S.
entity("reference") {
  identifier 'CITED_REF_QUERY_ID'
  constraint {
    type "PROJECTION"
    projection {
       relationshipName "reference_to_cited_reference"
       sourceIdentifier "REFERENCE_QUERY_ID"
    }
  }
}
//Find the set of unique predicates that come from either S or S'
entity("predicate") {
  identifier 'SUBGRAPH-QUERY-RESULTS'
  constraint {
    type "PROJECTION"
    projection {
      relationshipName "predicate_to_document"
      sourceIdentifier "REFERENCE_QUERY_ID"
    }
  }
  constraint {
    type "PROJECTION"
    projection {
      relationshipName "predicate_to_document"
      sourceIdentifier "CITED_REF_QUERY_ID"
    }
  }
  or()
}
```

Figure 24. Rest of predicate query for subsequent open discovery

3.6 System Query Post-Processing

During the execution of our retrospective studies, it was noticed that a number of concepts and predicates would appear within the results that had little to no value. The SemRep tool produces a large variety of predications, each with a semantic type that associates the two concepts. Some types such as *treats*, *stimulates*, or *causes* provide useful meaning, while others such as *administered_to*, *diagnoses*, and *location_of* are of less usefulness for our purposes. Additionally, concepts such as *Patients* or *Human* which

are so pervasive within the literature also provide little to no meaning. Two types of filters were placed during post-processing to account for these situations: one for predicates and one for concepts.

3.6.1 Concept Filtering

The standard approach for examining the "value" of concepts/words is to start by calculating the Inverse-Document Frequency (IDF) of each concept. This is normally computed during the compilation of the corpus. The difficulty for our study arises from two aspects. The first is the smaller total number of documents within our corpus (thousands of documents rather than millions). The second comes from the challenge of using NLP to associate words from titles and abstracts to the UMLS concepts within our results. Due to these restrictions the IDF was instead calculated based upon the predications data from SemRep for all of MEDLINE – consisting of 85.7 million predicates extracted from 16 million documents. For a given UMLS concept, if it exists within a predicate, then it is counted as existing within that document.

From Equation 2.1, the more times a particular term shows up across the corpus, the smaller the IDF value will be. For a given closed search, each predicate returned from a system query contains two concepts. The set of unique concepts returned can then be examined. An examination of the IDF scores for the *Magnesium – Migraine Disorders* study is illustrated in Figure 25. Table 3 then shows the lowest ten concepts, all with little semantic meaning.

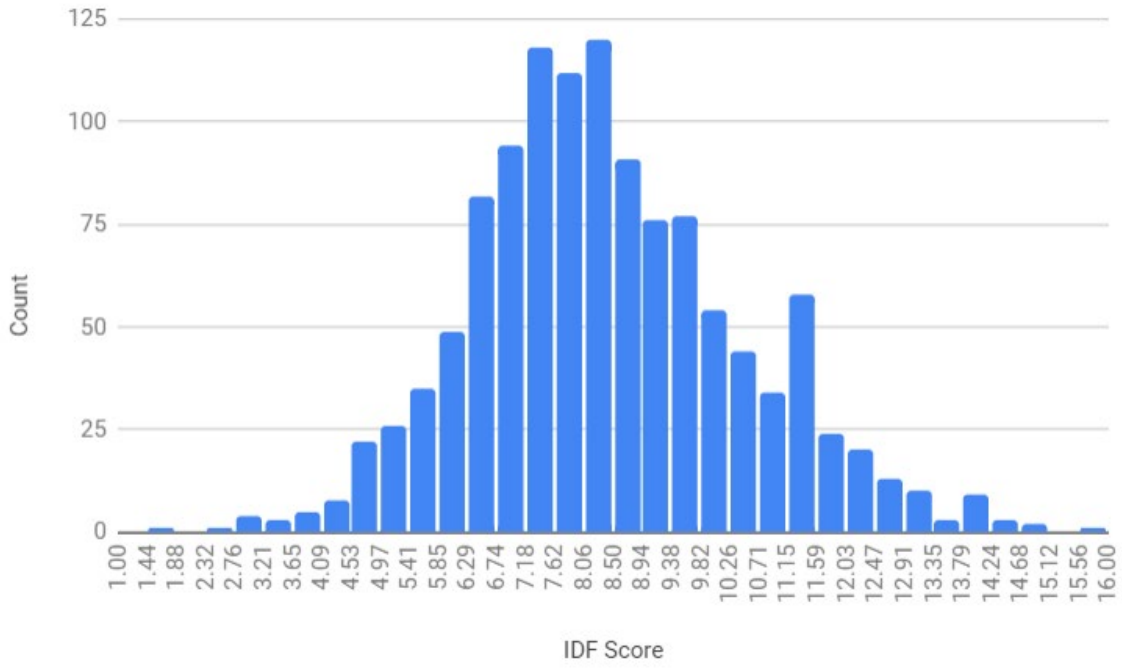


Figure 25. Histogram for concepts returned for Magnesium – Migraine Disorders

Table 3. Lowest concepts by IDF score returned for Magnesium – Migraine Disorders

Concept	IDF score
Patients	1.449559652
Therapeutic procedure	2.345444877
Human	2.821365442
Cells	2.903629008
Rattus norvegicus	2.996794117
Disease	3.147157075
Child	3.273016623
Woman	3.545474848
Mus	3.584664126
Pharmaceutical Preparations	3.64006119

The web tier can cull any predicates that contain a concept with a score lower than a given threshold. For experimental purposes we utilized a value that was two standard deviations below the mean score.

A second approach to remove uninformative relationships involves MeSH tree numbers. Each MeSH concept is organized into a hierarchy consisting of sixteen root categories listed in Appendix F. Category A contains anatomic terms, B is for organisms, C for diseases, D for drugs/chemicals, etc. With each category descriptors are arranged hierarchically from most general to most specific. As an example, the MeSH tree number for *Raynaud Disease* is C14.907.617.812, laying in category C because it is classified as a disease. Above it in the hierarchy is C14.907.617 for peripheral vascular diseases, C14.907 for vascular diseases, and C14 is for Cardiovascular diseases (Figure 26). A concept's tree number can help identify how general or specific it is, and some LBD studies have made use of this to remove suggestions that were too general (Zhang, Fiszman, Shin, Wilkowski, and Rindflesch 2013; Fiszman, Rindflesch, and Kilicoglu 2004; Swanson, Smalheiser, and Torvik, V. I. 2006).

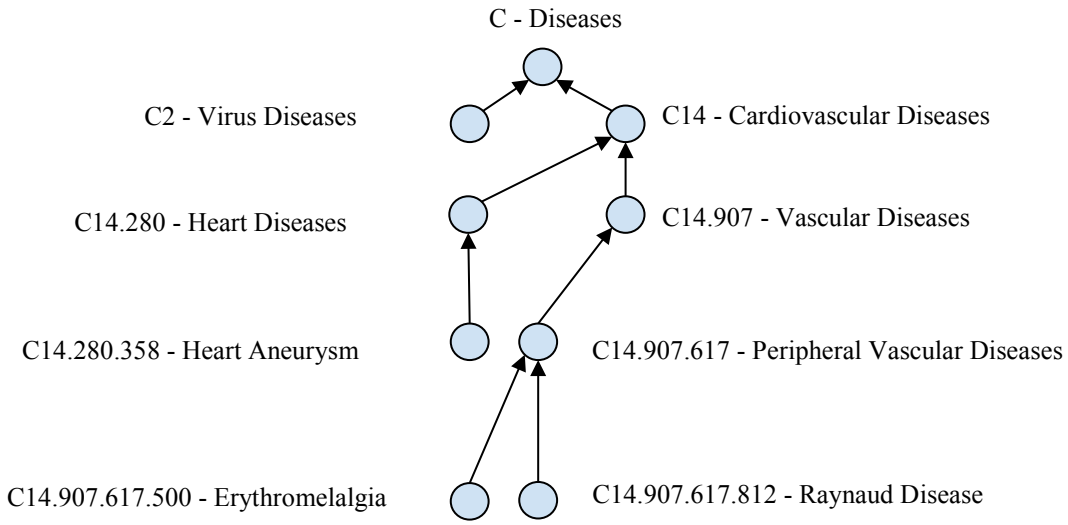


Figure 26. Sample MeSH Tree

Our primary results rely on semantic predications which relate two UMLS concepts. As not all UMLS concepts have an associated MeSH descriptor, they will also not have a corresponding MeSH Tree number. For those that do, having a shorter tree number indicates a more general concept and potentially is less useful. An examination of the average IDF values as a function of the MeSH tree depth shows just this. For the three studies *Fish Oils-Raynaud Disease*, *Magnesium-Migraine Disorders*, and *Somatomedins-Arginine* the list of predicates resulting from the search query was captured. For each study the list of unique concepts was extracted and divided into groups by MeSH tree number length. If a concept had more than one tree number, the closest integer average length was used.

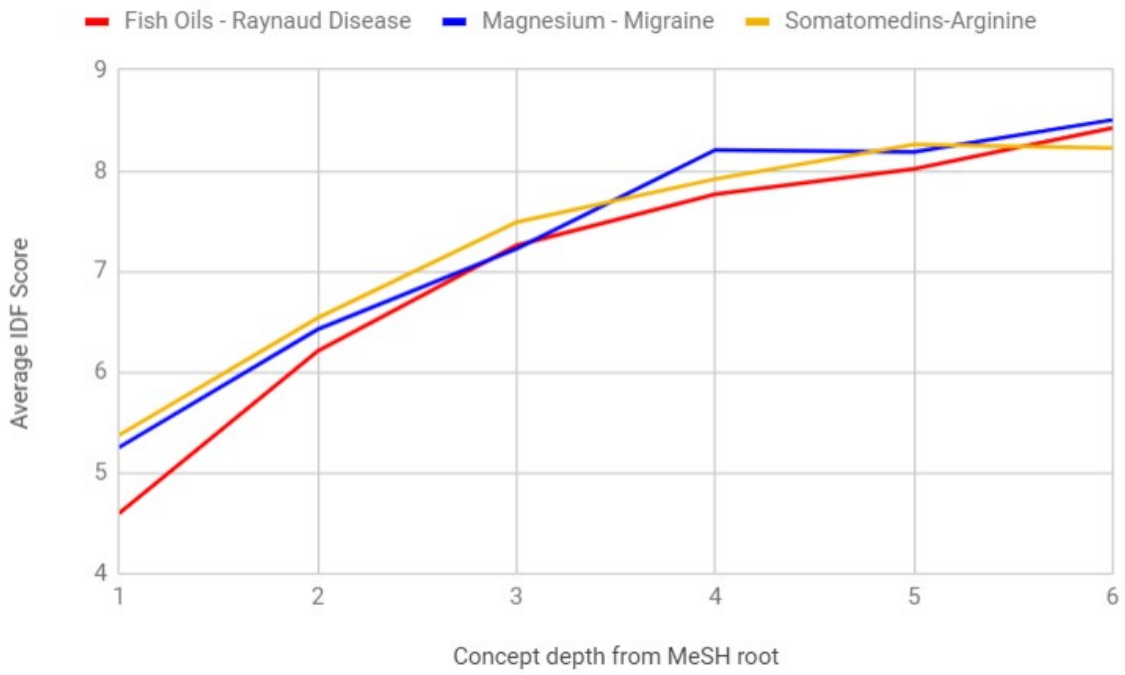


Figure 27. Average IDF value of concepts as a function of MeSH tree depth

As can be seen in Figure 27, more general concepts (i.e. shorter tree number length) have a smaller average IDF score than more specific terms. A closer examination

of the concepts shows us more concrete examples to determine an appropriate cutoff. Table 4 shows a listing of concepts with a tree depth of one for the same three studies along with their UMLS concept identifier and the MeSH Tree Number.

Table 4. Concepts with MeSH tree depth of one for three studies

C0043250	Injury wounds	[C26]	C0013227	Pharmaceutical Preparations	[D26]
C0007613	Cell physiology	[G04]	C0007634	Cells	[A11]
C0007222	Cardiovascular Diseases	[C14]	C0087111	Therapeutic procedure	[E02]
C0023779	Lipids	[D10]	C0751282	Macromolecular Complexes	[D05]
C0040300	Body tissue	[A10]	C0005515	Biological Factors	[D23]
C0027361	Persons	[M01]	C0021053	Immune System Diseases	[C20]
C0543467	Operative Surgical Procedures	[E04]	C0007004	Carbohydrates	[D09]
C0013227	Pharmaceutical Preparations	[D26]	C0007613	Cell physiology	[G04]
C0007634	Cells	[A11]	C0043251	Wounds and Injuries	[C26]
C0087111	Therapeutic procedure	[E02]	C0023779	Lipids	[D10]
C0029224	Organic Chemicals	[D02]	C0004611	Bacteria	[B03]
C0014130	Endocrine System Diseases	[C19]	C0011900	Diagnosis	[E01]
C0027765	nervous system disorder	[C10]	C0040300	Body tissue	[A10]
C0007004	Carbohydrates	[D09]	C0027361	Persons	[M01]
C0007613	Cell physiology	[G04]	C0006826	Malignant Neoplasms	[C04]
C0007222	Cardiovascular Diseases	[C14]	C0543467	Operative Surgical Procedures	[E04]
C0023779	Lipids	[D10]	C0013227	Pharmaceutical Preparations	[D26]
C0006826	Malignant Neoplasms	[C04]	C0007634	Cells	[A11]
C0543467	Operative Surgical Procedures	[E04]	C0087111	Therapeutic procedure	[E02]

As before, those in red are from the query results of the *Fish Oils-Raynaud Disease* study, those in blue come from a study around *Magnesium-Migraine Disorders*, and those in yellow are from the *Somatomedins-Arginine* papers. A glance over these concepts shows that all of them are too general to be useful as potential intermediates as we are interested in more specific biochemical connections. The same is also true for those concepts with a depth of two. At a depth of three we find a mixture of useful concepts such as *Hypertension* (C0020542) [C08.381.423] and less useful like *Human Volunteers* (C0020155) [M01.774.500]. The above analysis led to the culling of any predicates that contain a concept with either a tree depth of one or two.

3.6.2 Predicate Filtering

Most of the post-processing is accomplished via an examination of the concepts within each predicate. However, each predicate also has one of thirty semantic types which characterizes the relationship between the two concepts (Table 5). In addition, all but three of the types can appear in a negated form that have the prefix "NEG_". The semantic meaning of these connections may not be applicable for all studies attempting to uncover hidden knowledge.

The types *administered_to, location_of*, *process_of*, *method_of*, *part_of,* and *uses* consistently appear at the bottom of IDF scoring of predicates by group. This is caused by the prevalence of more generic concepts within such groupings: such as "water - administered_to - patients", "heart - location_of - fatty acids", "anemia - process_of - patients", and so on. In addition, some types such as *diagnoses* and *occurs_in* are less useful to us because such relations discuss procedures, in the case of *diagnoses*, and

population groups in the other. For our studies, where we wish to examine biochemical interactions, we filter out predicates with these eight semantic types.

Table 5. Predicate Semantic Types

Clinical	Biologic	General
ADMINISTERED_TO	ASSOCIATED_WITH	AFFECTS
COMPLICATES	AUGMENTS	CAUSES
DIAGNOSES	CONVERTS_TO	COEXISTS_WITH
MANIFESTATION_OF	DISRUPTS	COMPARED_WITH
OCCURS_IN	INHIBITS	HIGHER_THAN
PREDISPOSES	INTERACTS_WITH	ISA
PREVENTS	STIMULATES	LOCATION_OF
PROCESS_OF		LOWER_THAN
TREATS		METHOD_OF
		PART_OF
		PRECEDES
		PRODUCES
		SAME_AS
		USES

For closed discovery, the remaining predicates are then used to form a graph which can be very large and challenging to explore. Paths are computed from any source concept within (A set) to any target concept (C set), for path lengths up to the specified length.

Recall the earlier description of document sets S and S' (Figure 2). By definition the document set S contains all documents which contain at least one concept from the

union of the sets A and C. Therefore, all predicates that contain either an A or C concept must also come from those documents in S.

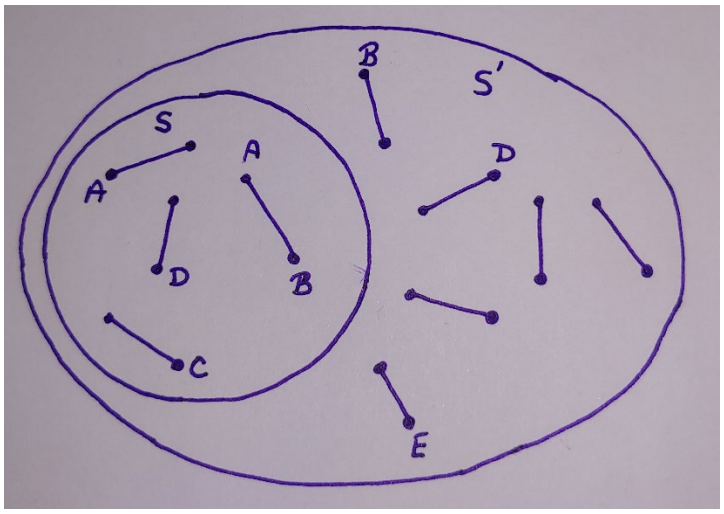


Figure 28. Predicates within document sets S and S'

A revisualization of this is shown in Figure 28. Predicates containing either an A or C concept must be within S, but past that there are no restrictions (i.e. all other possible concepts and predicates that connect those concepts may be in S, S', both, or neither). When we consider paths of length two with a single intermediate concept [A–B_1–C], then predicates will originate from papers containing the original search terms as each of the two predicates contain at least one search concept (Figure 29).

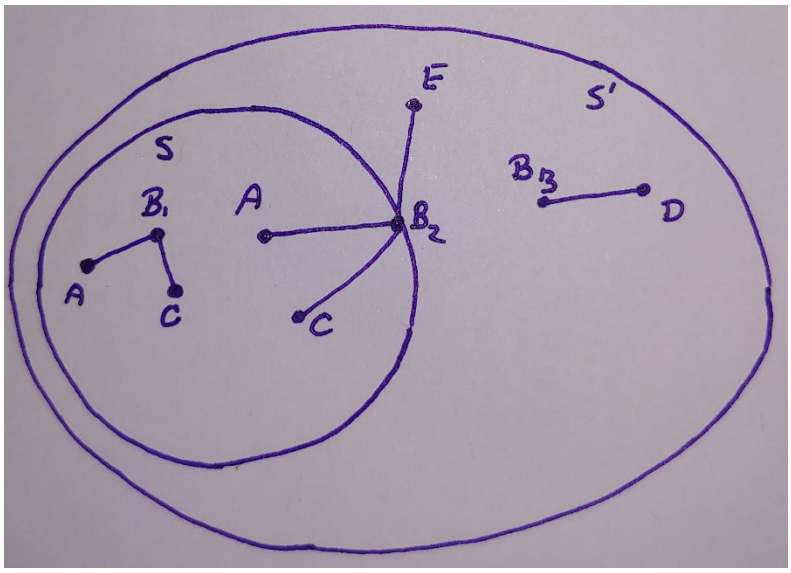


Figure 29. Paths of length two within document sets S and S'

Once you consider paths of length three [A–B_m–B_n–C], you allow for one predicate that does not contain a search term [B_m–B_n]. By observation the starting predicate [A–B_m] and the ending predicate [B_n–C] must be found within S, and therefore all four concepts must also be found within S. However, the connecting predicate [B_m–B_n] can also be found outside S – which in our case, will be S' (Figure 30).

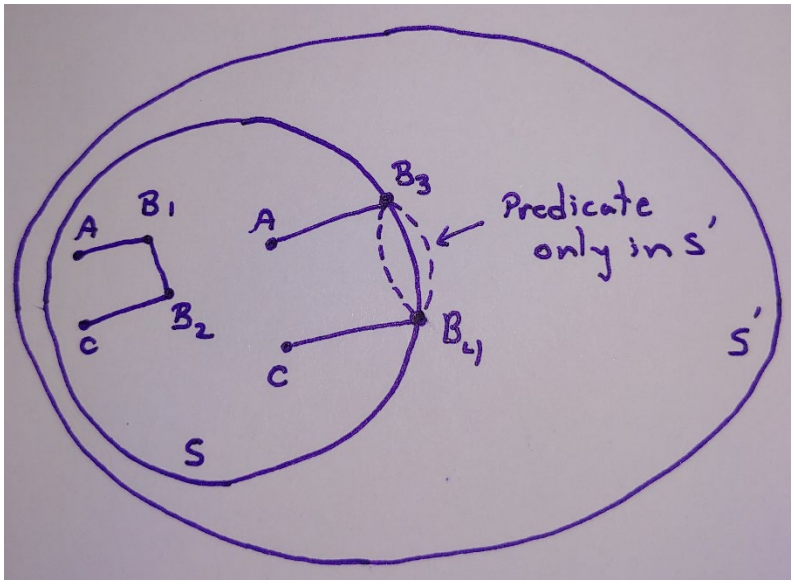


Figure 30. Paths of length three within document sets S and S'

We can further extend this analogy to paths of lengths four [$A-B_m-B_n-B_o-C$] which can be seen in Figure 31. It is at this point possible for a path to be formed that includes a concept that is not present within any of the documents of S. This analogy can be continued to further increase the possible number paths found, however, increasing path lengths will, at some point, cause each pair of concepts in the graph database to be connected as greater and greater numbers of predicates are included. As such, there is an expectation that the number of paths will grow quickly and the average value of those paths to decrease.

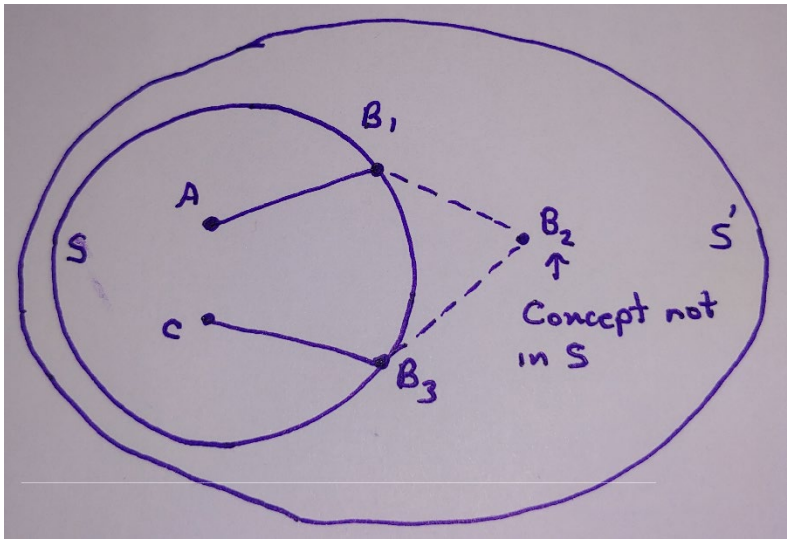


Figure 31. Paths of length four within document sets S and S'

Using the *Fish Oil-Raynaud Disease* study as an example, the total number of paths found and the average path IDF score were calculated as the maximum path length was varied from two to five. The average IDF score for a given path was calculated as the average of the IDF scores of each of the concepts that comprised the path. The average was then computed across all paths returned (Figure 32).

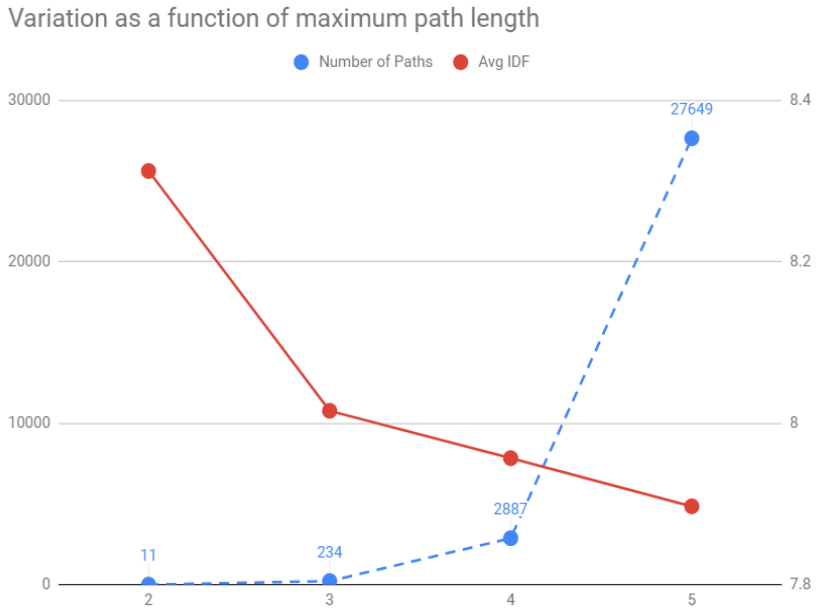


Figure 32. Paths found and average path IDF varying with maximum path length

From this it can be seen that the graph complexity (number of paths) grows quickly as a function of path length. Additionally, the average IDF value for those paths declines as additional 'low value' connections are made. These two facts match the intuitive idea that a more direct connection has more meaning.

The studies detailed in the next chapter are based upon previous work within the field. The seminal papers from which the discoveries were originally made tend to provide supporting evidence in the form of citations. Considering the ABC model for simplicity, in many cases the authors point out those documents supporting where either 'half' of the connection is derived (e.g. [A–B] or [B–C]). If a perfect set of semantic predicates were to be generated from all of those documents, then merely examining paths of length two would be sufficient to recover the connections. However, if our predicates are not a perfect representation of the semantic knowledge contained within those documents, then we want to make use of their citations as an alternative means of

summarization. As a compromise between graph complexity and desiring the information gain from citations, our trials are conducted utilizing path lengths of three by default, two when possible.

3.7 Subgraph Generation

As we've seen, even with a small corpus a search may result in hundreds or thousands of possible paths. This has caused researchers to explore approaches to filter, compare, and organize their concepts or predicates.

3.7.1.1 Degree Centrality

Graph theory defines degree centrality as a measure of connectedness of the vertices within a graph. By definition, a vertex with more edges than other vertices has higher degree centrality. Researchers have hypothesized that this equates to the importance of the corresponding concepts (Wilkowski et al. 2011; Zhang et al. 2011; Zhang, Fiszman, Shin, Wilkowski, and Rindflesch 2013). In their 2011 study "Degree centrality for semantic abstraction summarization of therapeutic studies" Zhang et al. used this method on SemRep predications to summarize the most important information from a collection of documents. They achieved better results when compared to term frequencies. Wilkowski et al. also used degree centrality when comparing possible paths, reducing large graphs by retaining only paths which contained highly connected concepts.

3.7.1.2 Fact Checking

When considering knowledge in general, statements of facts are in the same form as our predications: (subject, predicate, object) triples. If these then existed within a perfect knowledge graph, fact checking would be as trivial as checking for the existence of an edge. Unfortunately, our graphs of knowledge not only have limited information but also suffer from missing or incorrect assertions (Shi and Weninger 2016). When considering the knowledge contained within a large body of scientific papers there will exist duplicate or even contradictory assertions, not to mention that what can be regarded as true in one moment may change over time.

The frequency in which a given piece of knowledge appears within a corpus has been used as a measure of its validity for information retrieval (Kan, McKeown, and Klavans 2001; Zhang et al. 2013). In a newer study, NLP has been used to attempt to extract a factuality measure along with the predications Kilicoglu, H., Rosemblat, G., & Rindflesch, T. C. (2017). The basis of their work is formed on the idea that most biomedical knowledge claims are expressed in the form of hypotheses or speculation rather than explicit statements of fact. Additional linguistic analysis examine the words used (e.g. “Nifedipine increase blood flow”, “our findings support the hypothesis”, “may benefit”, “does not support”, etc.) and attempts to assign a factuality score: Fact, Probable, Possible, Counterfact, Conditional, etc.). These approaches attempt to order possible results by the degree of measured or implied factuality.

3.7.1.3 Concept Similarity

One of the strengths of utilizing predicates is the inherent semantic meaning contained with a path relating two concepts. A reasonable goal would be to cluster

together predicates that are more closely related to that higher level concept. For example, in the *Fish Oils-Raynaud Disease* study we could expect one grouping around the topic of *Platelet Aggregation* and a different one around the topic of *Blood Viscosity* – similar to how latent semantic analysis attempts to produce collections of terms around higher order themes.

One of the simplest measures to compare concepts is a technique based on the reciprocal of the shortest path length between a pair of concepts within an ontology (Rada, Mili, Bicknell, and Blettner 1989; Lee, Kim, and Lee 1993). For instance, in Figure 26 the similarity score between *Raynaud Disease* and *Vascular Disease* is the same as between *Erythromelalgia* and *Vascular Disease* because the distance is the same.

A weakness of this approach lies in the assumption that the link or association between a concept and its parent are uniform distances. In 1995, Resnik introduced an equally simply but alternative measure based upon the concept of information content. Resnik stated that "one key to the similarity of two concepts is the extent to which they share information in common" (Resnik 1995). That is the connecting path between the concepts must go through a concept that is abstract enough to contain them both. For our earlier example (Figure 26) *Erythromelalgia* and *Raynaud Disease* are connected through *Peripheral Vascular Diseases* while *Virus Diseases* and *Raynaud Disease* are only connected through the root *Diseases*.

Resnik also suggested the association of probabilities with the concepts to avoid the unreliability of edge distances. Each concept was assigned a value that equates to the probability of encountering an instance of that concept. In information theory, *information content* is defined as the amount of information gained when a random

variable or signal is sampled (Ross 2014). In this case, the information gained for a concept (E) with probability P is simply:

$$I(C) := -\log(P) \tag{3.1}$$

This should make intuitive sense: as the probability increases (the more abstract the concept), the lower its potential information content (Figure 33).

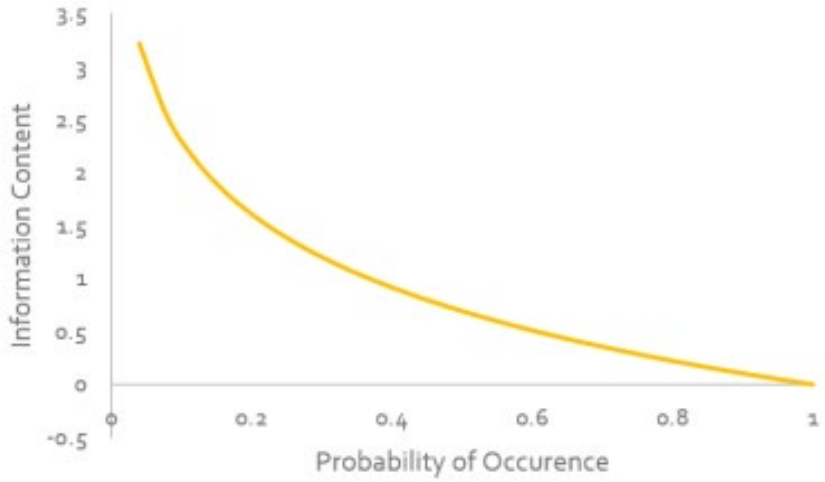


Figure 33. Information Content as a function of probability

Resnik, taking into account multiple inheritance, defined his similarity as the information content of the least common subsumer of both concepts. In a more general sense, the semantic similarity between two concepts is proportional to the amount of information they share (Resnik 1995; Jiang and Conrath 1997; Lin 1998).

3.7.2 Path Relatedness

A fundamental requirement for clustering paths based upon semantic meaning is the ability to define how similar two paths are to each other. According to Cameron,

assertions in scientific literature which are related will usually occur in similar contexts, and that the context of a path is the aggregation of the context of each assertion or predication within the path (Cameron 2014).

To define the context of a single semantic predication Cameron made two assumptions. The first is that as MEDLINE articles are manually assigned a list of MeSH descriptors, these form a semantic summary of the full text of the article comprised of concepts. In addition, the predications themselves form a relational semantic summary by associating one concept to another. As both capture the implicit context of the reference, the concept-level semantic summary (MeSH descriptors) and relational semantic summary (predications) are interchangeable, albeit not equal, perspectives.

The essence of the second assumption is succinctly stated by the English linguist John Rupert Firth: "You shall know a word by the company it keeps" (Firth 1957). In other words, that the "context of an individual predication across the entire corpus could be represented more comprehensively, as the aggregation of the MeSH descriptors assigned to each document in which it occurs" (Cameron 2014).

Cameron continued this reasoning to define the context of a path as the summation of the contexts of its predicates. While standard distributional approaches would assign frequency vectors as the context of each predicate and then to use cosine similarity to compare paths, this is a poor fit as the goal is to cluster paths based upon relatedness and not similarity. Additionally, because the concept descriptors are arranged in a hierarchical structure, concepts will have one or more tree numbers that indicate their relative positioning within the tree. If two different concepts share some common parents they will share some semantic meaning. For example, the MeSH term *Fish Oils*

(D005395) has a tree number of D10.627.430. The D signifies *Chemicals and Drugs*, D10 is the identifier for *lipids*, and D10.627 identifies *oils*. Thus, another concept such as *Fatty Acids, Omega-3* with a tree number of D10.627.430.450 signifies that it is closely related to the concept of Fish Oils.

The Sørensen–Dice coefficient, or Dice similarity, is a statistic used to measure the similarity between two sets (Eq 3.2).

$$dice(X, Y) = \frac{2 \times |X \cap Y|}{|X| + |Y|} \tag{3.2}$$

$$dice(m_i, m_j) = \frac{2 \times |A(m_i) \cap A(m_j)|}{|A(m_i)| + |A(m_j)|} \tag{3.3}$$

This measure has long been applied to terms within a taxonomy leveraging information measures and the least common parent that subsumes both terms (Lin 1998). Cameron's suggestion was that the dice similarity between two MeSH descriptors m_i and m_j could be computed based upon the sets $A(m_i)$ and $A(m_j)$ respectively where A(m) is the set of all ancestors of the MeSH descriptors m (Eq 3.3). For example, the dice similarity between *Fish Oils* (D10.627.430) and *Fatty Acids, Omega-3* (D10.627.430.450) is calculated as follows:

A(*Fish Oils*) = {D10.627.430, D10.627, D10}

A(*Fatty Acids, Omega-3*) = {D10.627.430.450, D10.627.430, D10.627, D10}

A(*Fish Oils*) ∩ A(*Fatty Acids, Omega-3*) = {D10.627.430, D10.627, D10}

dice (*Fish Oils, Fatty Acids, Omega-3) = (2 x 3) / (4 + 3) = 6/7*

The range of similarity values is [0,1] where the maximum similarity occurs when the two descriptors are equal. Cameron further defines a normalized dice similarity score by using a similarity threshold ($\tau_{sim} = 0.75$) in an effort to maximize the weight of *in-context* descriptors and minimize the weight of *out-of-context descriptors* (Eq 3.4).

$$dice_N(m_i, m_j) = \begin{cases} 1 & if\ \ dice(m_i, m_j) > \tau_{sim} \\ 0 & otherwise \end{cases} \tag{3.4}$$

$$sr''(p_i, p_j) = \sum_{(a,b) \in C(p_i) \times C(p_j)} dice_N(a, b) \tag{3.5}$$

The semantic relatedness $sr''(p_i,p_j)$ between two paths p_i and p_j, is then computed as the sum of the pairwise normalized dice similarity between the context sets $C(p_i)$ and $C(p_j)$ (Eq 3.5). Take the following two paths from our results explained later as an example:

p_i	Eicosapentaenoic Acid	INTERACTS_WITH	Epoprostenol	TREATS	Raynaud Phenomenon
p_j	Eicosapentaenoic Acid	STIMULATES	Epoprostenol	TREATS	Raynaud Disease

The first predicate ($pred_1$), *Eicosapentaenoic Acid* - INTERACTS_WITH - *Epoprostenol* appears only in [PMID6252872] which is in our corpus because it is cited by three core references [PMID6302714, PMID6318123, PMID6321621]. The second predicate ($pred_2$), *Epoprostenol* - TREATS - *Raynaud Phenomenon* appears in three core papers [PMID3883365, PMID6788326, PMID7037038]. The third predicate ($pred_3$), *Eicosapentaenoic Acid* - STIMULATES - *Epoprostenol* comes from [PMID6098049]. The last predicate ($pred_4$), *Epoprostenol* - TREATS - *Raynaud Disease* appears in the core reference [PMID3883365] and one of the papers it cites [PMID7025341].

The context set for each of these predicates is simply the aggregate of the MeSH descriptors of all the documents from which it is present. Table 6 shows the context sets for these four predicates along with the document identifiers where each concept originated. The context set for each path is then simply the unique set of MeSH descriptors aggregated from its predicates. For p_i, this is the 27 unique terms resulting from the union of $C(pred_1)$ and $C(pred_2)$, and for p_j, this is the 26 unique terms resulting from the union of $C(pred_3)$ and $C(pred_4)$.

For the 702 pairs only 32 cross the similarity threshold (Table 7). Of these 18 are self matches, leaving only 14 that are more interesting. An example calculation of one pair is as follows for the case m_i = *Epoprostenol* [D10.251.355.255.550.550.500] and m_j = *Prostaglandins* [D10.251.355.255.550]:

A(*Epoprostenol*) = {D10.251.355.255.550.550.500, D10.251.355.255.550.550, D10.251.355.255.550, D10.251.355.255, D10.251.355, D10.251, D10}

A(*Prostaglandins*) = {D10.251.355.255.550, D10.251.355.255, D10.251.355, D10.251, D10}

A(*Epoprostenol*) ∩ A(*Prostaglandins*) = A(*Prostaglandins*)

dice (*Epoprostenol, Prostaglandins) = (2 x 5) / (7 + 5) = 10/12 = 0.833*

Table 6. Context set for example predicates

$C(pred_1)$	$C(pred_2)$	$C(pred_3)$	$C(pred_4)$
Arachidonic Acids (PMID6252872)	6-Ketoprostaglandin F1 alpha (PMID3883365)	Arachidonic Acids (PMID6098049)	6-Ketoprostaglandin F1 alpha (PMID3883365)
Blood Vessels (PMID6252872)	Adult (PMID3883365,PMID7037038)	Eicosapentaenoic Acid (PMID6098049)	Adult (PMID3883365)
Eicosanoic Acids (PMID6252872)	Aged (PMID7037038)	Endothelium (PMID6098049)	Ambulatory Care (PMID7025341)
Eicosapentaenoic Acid (PMID6252872)	Epoprostenol (PMID3883365, PMID6788326, PMID7037038)	Epoprostenol (PMID6098049)	Buffers (PMID7025341)
Epoprostenol (PMID6252872)	Erythrocyte Deformability (PMID3883365)	Fatty Acids, Unsaturated (PMID6098049)	Clinical Trials (PMID7025341)
Fatty Acids, Unsaturated (PMID6252872)	Erythrocytes (PMID6788326)	Humans (PMID6098049)	Epoprostenol (PMID3883365, PMID7025341)
Humans (PMID6252872)	Female (PMID3883365, PMID6788326, PMID7037038)	Indomethacin (PMID6098049)	Erythrocyte Deformability (PMID3883365)
Platelet Aggregation (PMID6252872)	Humans (PMID3883365, PMID6788326, PMID7037038)	Platelet Aggregation (PMID6098049)	Female (PMID3883365, PMID7025341)
Prostaglandins (PMID6252872)	Infrared Rays (PMID7037038)	Umbilical Veins (PMID6098049)	Fingers (PMID7025341)
Thromboxanes (PMID6252872)	Infusions, Parenteral (PMID7037038)		Hand (PMID7025341)
Umbilicus (PMID6252872)	Male (PMID3883365, PMID6788326, PMID7037038)		Humans (PMID3883365, PMID7025341)
	Middle Aged (PMID7037038)		Injections, Intravenous (PMID7025341)
	Prostaglandins (PMID6788326, PMID7037038)		Lupus Erythematosus, Systemic (PMID7025341)
	Prostaglandins E (PMID6788326)		Male (PMID3883365)
	Radiometry (PMID7037038)		Prostaglandins (PMID7025341)
	Raynaud Disease (PMID3883365, PMID6788326, PMID7037038)		Raynaud Disease (PMID3883365, PMID7025341)
	Scleroderma, Systemic (PMID3883365, PMID6788326)		Scleroderma, Systemic (PMID3883365)
	Thermography (PMID7037038)		Thermography (PMID7025341)
	Time Factors (PMID7037038)		Time Factors (PMID7025341)

Table 7. MeSH descriptors pairs m_i x m_j that exceed τ_{sim}

m_i	m_j	$\sum(dice_N(m_i,m_j))$
Epoprostenol	Epoprostenol, Prostaglandins	2
Eicosapentaenoic Acid	Eicosapentaenoic Acid, Arachidonic Acids, Prostaglandins	3
Thromboxanes	Arachidonic Acids	1
Fatty Acids, Unsaturated	Fatty Acids, Unsaturated	1
Humans	Humans	1
Arachidonic Acids	Eicosapentaenoic Acid, Arachidonic Acids, Prostaglandins	3
Platelet Aggregation	Platelet Aggregation	1
Prostaglandins	Eicosapentaenoic Acid, Arachidonic Acids, Prostaglandins, Epoprostenol, 6-Ketoprostaglandin F1 alpha	5
Erythrocyte Deformability	Erythrocyte Deformability	1
Female	Female	1
Raynaud Disease	Raynaud Disease	1
Male	Male	1
6-Ketoprostaglandin F1	6-Ketoprostaglandin F1, Prostaglandins	2
Scleroderma, Systemic	Scleroderma, Systemic	1
Adult	Adult	1
Time Factors	Time Factors	1
Middle Aged	Adult	1
Thermography	Thermography	1
Aged	Adult	1
Prostaglandins E	Prostaglandins, Epoprostenol, 6-Ketoprostaglandin F1 alpha	3

Cameron noted that a broad range of scores may exist for this approach and additionally suggested using a log reduction in addition to the normalized dice similarity scores. This was accomplished by adding an additional step to compute the relatedness between a single MeSH descriptor to the context set of descriptors for a path (Eq 3.6).

$$sr'(m_i, C(p_j)) = \sum_{m_j \in C(p_j)} dice_N(m_i, m_j) \tag{3.6}$$

$$sr''_L(p_i, p_j) = \sum_{m_i \in C(p_i)} ln(1 + sr'(m_i, C(p_j))) \tag{3.7}$$

Using this new equation for the semantic relatedness between two paths (Eq 3.7), and data calculated in Table 7, for our example we see that:

$$sr''_L(p_i, p_j) = (14 \times ln(2)) + (2 \times ln(3)) + (3 \times ln(4)) + (1 \times ln(6))$$

$$sr''_L(p_i, p_j) = 17.852$$

3.7.3 Cluster Similarity

If the desired goal is to achieve one or more result graphs with semantically related paths, then there is an additional measure we require. Equation 3.7 measures how much two single paths are related to each other, but another statistic is required when comparing one cluster to another. Hierarchical clustering algorithms measure this similarity in one of four ways (Manning, Raghavan, and Schütze 2008).

The first, single-link similarity, simply uses the similarity between the pair most similar between the two clusters. The second, complete-link similarity, uses the pair most dissimilar. The third is centroid, or average inter-similarity across all pairs between the two clusters. Lastly, the fourth is the average of all similarities between all pairs including those pairs from the same cluster. Cameron noted that centroid similarity is

ideal for the merging of two groups of paths as it will maintain the greatest amount of relatedness. As such he defines the inter-cluster similarity between two buckets B_a and B_b each containing one or more paths:

$$sim_{inter}(B_a, B_b) = \frac{\sum_{(p_i, p_j) \in B_a \times B_b} sr''_L(p_i, p_j)}{|B_a| \times |B_b|} \tag{3.8}$$

3.7.4 Subgraph Clustering

Cameron used a hierarchical agglomerative clustering (HAC) algorithm to construct coherent subgraphs based on the context relatedness of the individual paths (Cameron 2014). Initially starting with a number of buckets equal to the total number of possible paths, they iteratively calculate the relatedness between a set of paths from one bucket to each other. If the score exceeds a threshold the buckets are combined. Repeating the process until no more buckets are merged. This can result in one or more subgraphs whose paths share greater semantic meaning, and thus should appear less cluttered. The steps for Cameron's algorithm are shown in Table 8.

Table 8. Hierarchical Agglomerative Clustering algorithm used by Cameron et al.

Step 1: For each path generate an empty bucket and place the path within. Call this collection *buckets*

Step 2: Compute path relatedness threshold as two standard deviations above the mean of the path relatedness scores.

Step 3: Computer the inter-cluster similarity score between each pair of buckets b_1 and b_2 in *buckets* compute the similarity score (Equation 3.8). For each pair whose similarity score exceeds the threshold, mark the two buckets for merging.

Step 4: For each pair marked for merging, create a new bucket consisting of all paths from the two buckets. Place this new bucket in *mergedBuckets*. If a bucket has not been marked for deletion or merging, add it to *mergedBuckets* by itself.

Step 5: Repeat this algorithm until number of buckets in the previous iteration is equal to the number of buckets from the current one - that is, there are no more changes.

As noted by Cameron et al. (2015) the time to complete this step could take from minutes to hours and produced between one and two hundred graphs and singleton paths for each of their studies. Although this can produce more semantically meaningful results, this is problematic for two reasons. One, there are two sets of results: subgraphs consisting of more than one path and singletons. If we wish to ascertain quantifiable measurements about the efficacy of the system in terms of precision and recall then only a single ordered list of the results should be returned to the researcher.

Secondly, there may be many paths that are extremely similar. This can arise from several different cases: differences in predicate semantic types only, differences only in

the starting and/or ending concepts, and paths that are a longer version of an existing path. Despite their similarities, they will have slightly different contexts which could cause them to not be combined together during clustering. But to the user viewing the end results, the extra graphs will seem redundant and of little use.

When searching the *Fish Oils-Raynaud Disease* study with a path length of two, there are eleven paths found (Table 9). The first case can be seen when comparing the first and third listed paths. The paths differ only in that with one *Epoprostenol* TREATS *Raynaud Phenomenon* and in the other, it is ASSOCIATED_WITH.

If we consider these cases as similar and treat them as a single merged path, then after combining a single path is left as a placeholder for the grouping, and for our *Fish Oils* example this leaves only three paths remaining (Table 10).

Zhang et al. describes a similar approach, where at a higher level, a predicatation can be considered as a metapredication in the form {Semantic Concept Group - Predicate Group - Semantic Concept Group} (Zhang et al. 2013). In their study predicate types are grouped together (e.g. INHIBITS, STIMULATES, and INTERACTS_WITH into an "Interaction" group). They use these grouping to form graph clusters.

Table 9. Fish Oils – Raynaud Disease paths with max length two

Eicosapentaenoic Acid	INTERACTS_WITH	Epoprostenol	TREATS	Raynaud Phenomenon
Eicosapentaenoic Acid	INTERACTS_WITH	Epoprostenol	TREATS	Raynaud Disease
Eicosapentaenoic Acid	INTERACTS_WITH	Epoprostenol	ASSOCIATED_WITH	Raynaud Phenomenon
Eicosapentaenoic Acid	NEG_INHIBITS	Epoprostenol	TREATS	Raynaud Phenomenon
Eicosapentaenoic Acid	NEG_INHIBITS	Epoprostenol	TREATS	Raynaud Disease
Eicosapentaenoic Acid	NEG_INHIBITS	Epoprostenol	ASSOCIATED_WITH	Raynaud Phenomenon
Eicosapentaenoic Acid	INHIBITS	Thromboxane A2	CAUSES	Raynaud Phenomenon
Eicosapentaenoic Acid	INTERACTS_WITH	Thromboxane A2	CAUSES	Raynaud Phenomenon
Eicosapentaenoic Acid	STIMULATES	Epoprostenol	TREATS	Raynaud Phenomenon
Eicosapentaenoic Acid	STIMULATES	Epoprostenol	TREATS	Raynaud Disease
Eicosapentaenoic Acid	STIMULATES	Epoprostenol	ASSOCIATED_WITH	Raynaud Phenomenon

Table 10. Condensed Fish Oils – Raynaud Disease paths with max length two

Eicosapentaenoic Acid	*	Epoprostenol	*	Raynaud Phenomenon
Eicosapentaenoic Acid	*	Epoprostenol	*	Raynaud Disease
Eicosapentaenoic Acid	*	Thromboxane A2	CAUSES	Raynaud Phenomenon

The second case can be seen in the first and second lines of Table 10. Here, an assumption is made that the researcher using the system considers the starting set of concepts together conceptually and the same for the target set of concepts. In other words, as long as the “core” of the path (everything not including the start and end) is

identical, we can consider the paths the same and plot them together. Figure 34 shows an illustration of these eleven paths being condensed into two aggregate groupings.

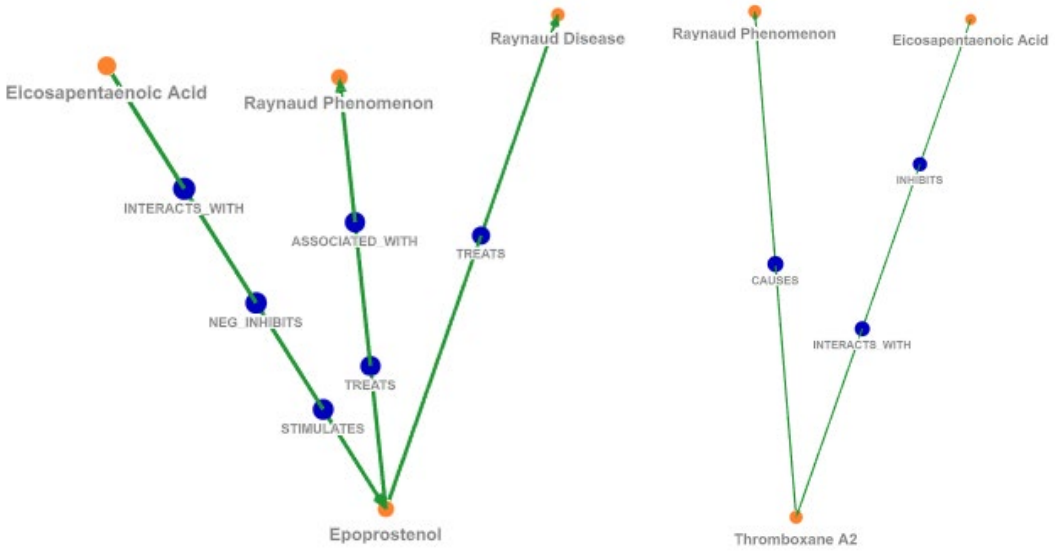


Figure 34. Example combined predicates for Fish Oils – Raynaud Disease

The third case arises with longer path lengths. For example, take the following two paths (Table 11). The first path is just a longer version of the second path. As such, we use the shorter path for subgraph creation and add the longer path into whatever results include the shorter.

Table 11. Sample of Fish Oils – Raynaud Disease paths with max length three

Eicosapentaenoic Acid	INHIBITS	Arachidonic Acid	STIMULATES	Thromboxane A2	CAUSES	Raynaud Phenomenon
Eicosapentaenoic Acid	INHIBITS	Thromboxane A2	CAUSES	Raynaud Phenomenon		

Using these three checks as additional post-processing we then can form the subgraphs with the resulting paths (Table 12):

Table 12. Clustering algorithm

Step 1: For each path generate an empty bucket and place the path within. Call this collection *buckets*

Step 2: Compute path relatedness threshold as two standard deviations above the mean of the path relatedness scores.

Step 3: Remove the first bucket, b_1, from *buckets*. Compare b_1 with the remaining buckets in *buckets* by computing their similarity score. For each bucket b_2 whose similarity score with b_1 exceeds the threshold, add all paths contained within b_2 to *listToCombine*.

Step 4: Create a new bucket consisting of all paths contained within b_1 and from *listToCombine*. Place this new bucket in *listsToAdd.* If a bucket has not been marked for merging, add it to *leftOvers* by itself.

Step 5: If *buckets* is not empty go back to step three.

Step 6: Add all buckets within *listToAdd* to *buckets*.

Step 7: If any buckets were combined, go back to step three.

The results of this algorithm consists of an unordered set of subgraph results. Each result will consist of one or more paths as some will have been combined semantically.

3.8 Result Ordering

With many possible subgraph results it is highly important to determine a sort order. Paths with the shortest length are likely to be of primary importance to the researcher using an LBD system. As such, the primary sorting criteria of the resulting subgraphs is the minimum path length contained. If there is a more direct relation between two concepts, a researcher is more likely to be interested in this than a path that goes through several intermediate concepts.

The second sorting criteria is an ordering based upon the average frequency of occurrence of the found path predications. The more documents a predication has appeared within, generally implies a greater accepted belief of that statement. (Kan, McKeown, and Klavans 2001; Zhang, Fiszman, Shin, Wilkowski, and Rindflesch 2013)

3.9 Result Evaluation

Put succinctly, evaluating LBD systems is a "fundamentally challenging task because if they are successful, by definition they are capturing new knowledge that has yet to be proven useful" (Yetisgen-Yildiz and Pratt 2008). This difficulty has been overcome by many researchers by replicating the historical discoveries to evaluate the performance of their systems (Lindsay and Gordon 1999; Weeber, Klein, de Jong-van den Berg, and Vos 2001; Srinivasan 2004; Hristovski, Friedman, Rindflesch, and Peterlin 2006; Cameron et al. 2015; Song, Heo, and Ding 2015). These historical studies have become known as the gold-standard for LBD systems and include many from Swanson and Smalheiser including *Fish Oils – Raynaud Disease*, *Magnesium – Migraine*

Disorders, *Alzehimer's Disease – Estrogen*, *Somatomedin C – Arginine, Alzeimer's Disease – Indomethacin*, and others.

In general, these systems have reported qualitative success based solely upon the presence of the original connections within their results - usually claiming to have recovered some number of the intermediate concepts found in the original paper. This approach ignores the fact that in some cases these results are a few out of potentially hundreds of results (ignoring the rest), nor does it give a quantitative measure of the information retrieval (Yetisgen-Yildiz and Pratt 2008).

3.9.1 Information Retrieval Metrics

At its heart LBD is a blend of summarization and information retrieval. A standard measure of IR system performance is not only the ability to return information relevant to the search, but also to not return non-relevant information. This has been quantized in two popular metrics: precision and recall (Baeza-Yates and Ribeiro-Neto 1999).

For a standard document-based information retrieval system, precision is defined as the proportion of the relevant documents returned compared with the total number returned (i.e. what % of the returned results are correct). Recall is defined as the proportion of relevant documents returned compared with the total number of relevant documents in the system (i.e. what percentage of the correct results are returned). In contrast to these, LBD systems return terms instead of documents. As such, precision and recall for the evaluation of an LBD system are calculated with the following formulas:

$$P_i = \frac{||T_i \cap G_i||}{||T_i||} \tag{3.9}$$

$$R_i = \frac{||T_i \cap G_i||}{||G_i||} \tag{3.10}$$

Here, T_i is the set of intermediate terms generated by the LBD system for a given starting term(s) *i*, and G_i is the set of intermediate terms as defined by the gold standard (Yetisgen-Yildiz and Pratt 2008).

Within information retrieval systems a well-known challenge is the trade-off between the two metrics. A system designed for high precision can result in low recall, and vice versa. An additional measure was created to help address this issue, F-measure, which is a combined version of the two:

$$F = \frac{\left(1+\beta^2\right) + \times R \times P}{\left(\beta^2 \times R\right) + P} \tag{3.11}$$

For F-measure, R is the recall, P is the precision, and β is a weighting factor between the two. The most commonly used case sets this to one, assigning equal value to precision and recall.

3.9.2 Ranked Retrieval

Precision, recall, and the F-measure are all set-based measures calculated using a complete unordered set of results. With ordered (ranked) result sets, subsets of retrieved documents can be examined using the top *k* retrieved documents. For example, if a search had a total of seven ordered results (d_0 through d_6) and a maximum of three relevant

results, we can calculate both precision and recall after considering each result. These figures are shown in Table 13.

Table 13. Example Ranked Retrieval Data

Ranked Result		**Precision**	**Recall**
d_0	Relevant	1.0	0.333
d_1	Relevant	1.0	0.667
d_2	Not Relevant	0.667	0.667
d_3	Not Relevant	0.5	0.667
d_4	Not Relevant	0.4	0.667
d_5	Not Relevant	0.333	0.667
d_6	Relevant	0.429	1.0

For each of these, the values of precision and recall can be plotted to provide a *precision-recall curve*. These curves have a distinctive saw-tooth shape due to the fact that if the $(k+1)^{th}$ result retrieved is non-relevant then recall is the same as for the top k documents, but precision has dropped. And if it is relevant, then both precision and recall increase causing the curve to increase up and to the right (Figure 35).

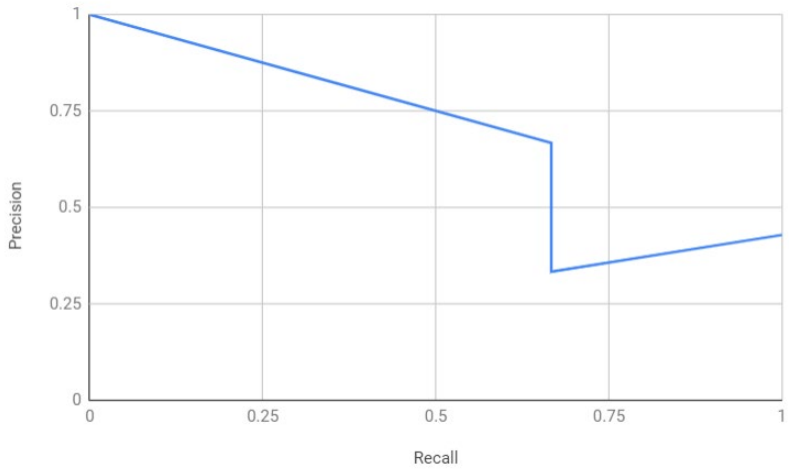


Figure 35. Sample Precision – Recall Curve

A common combination of precision and recall is the F_1 measure, defined as the harmonic mean of precision and recall (Eq. 3.12).

$$F1 = \frac{2 * precision * recall}{precision + recall} \tag{3.12}$$

$$F1@k = \frac{2 * precision@k * recall@k}{precision@k + recall@k} \tag{3.13}$$

This can then be further extrapolated for ranked recall to a measure of "F_1 at k" (Eq 3.13) which is a calculation of the F_1 score as a function of the top k returned results.

3.9.3 Natural Language Challenges

Suppose we try to strictly apply our measures of precision and recall. Using the *Magnesium-Migraine Disorders* study as an example, we run into an inherent problem. The usefulness of any term or predicate depends ultimately on the context of the articles within which it occurs. "Interpreting that context and its usefulness requires, in general, expert knowledge and human judgment" (Swanson, Smalheiser, and Torvik 2006). For example, magnesium acts like a Calcium antagonist, having been compared to both Verapamil and Nifedipine - two calcium channel blockers known to help treat migraines (Figure 36).

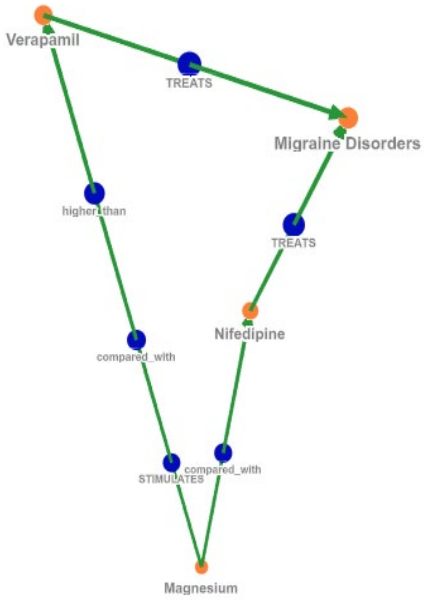


Figure 36. Magnesium as a Calcium antagonist

Although our gold standard LBD studies identify a list of intermediates, such as *Calcium Channel Blockers* between *Magnesium* and *Migraine Disorders*, these are usually meant to identify "higher-level" concepts that relate the two. And here two classic problems arise in natural languages: *synonymy* and *polysemy*. Synonymy refers to a case where two different words (say *Calcium Channel Blockers* and *Verapamil*) have the same meaning. Polysemy, on the other hand, refers to the situation where a term has multiple meanings (e.g. bank). To account for this in our calculations of precision and recall the following rules are used:

1. Multiple terms that are associated with a "gold-standard" intermediate are all counted as a single hit/true positive. (i.e. If the results contain *Calcium Channel Blockers, Verapamil, Nifedipine*, etc. then these all count as a single true positive for the idea of *Calcium Channel Blockers).*

2. If no terms are found to associate with a “gold-standard” intermediate, this is counted as a single miss/false negative.
3. Each term appearing in the answers that cannot be associated with a “gold-standard” intermediate will count as a false positive.

Yet another difficulty that may arise is that the semantics of terms may evolve over time. For example, the term “virus” which initially only referred to an infectious biological agent now also is used to refer to those of a digital origin with similar properties (Xun, Jha, Gopalakrishnan, Li, & Zhang 2017). When creating associations the context of the surrounding words can matter to better identify the semantic meaning of the term.

3.10 Summary

This chapter presents the system design of a performant LBD system and semantic-based search algorithms for both closed and open discovery approaches. A key contribution of this dissertation is the inclusion of citations within these algorithms. This is followed by a discussion on several filters based upon concept IDF scores and the MeSH term hierarchy that must be applied to cull terms which are too common or too general. Lastly we discuss how to generate subgraph results along with an evaluation approach to qualitatively compare the results with and without the use of citational references.

Chapter 4

Results

Knowledge rediscovery, or retrospective analysis, is the standard for evaluating LBD systems. Initially the goal was to construct a LBD framework with an implementation of Cameron's algorithms using the *Fish Oils – Raynaud Disease* study as a testbed. This was then expanded to make use of citational links between documents. Because these had positive results, additional studies were performed with the same algorithms to test the results to see if they were truly indicative of better results. To date, we have performed five projects using retrospective studies. These projects were based on the original studies: the *Magnesium – Migraine Disorders* discovery by Swanson, the connections between *Somatomedins* and *Arginine* also by Swanson, the association between *Testosterone* and *Sleep* by Miller, and Kostoff's research on the common factors of *Parkinson's disease* and *Crohn's disease*.

4.1 Retrospective Study: Fish Oils and Raynaud Disease

In this retrospective study the corpus consisted only of the papers that Swanson cited in his paper (Table A-1) as well as documents that those papers cited. A listing of the key predicates along with their corresponding sentences are recorded by document identifier within Table A-2. Appendix A also includes the topmost results of the search, first using citations and then again without.

We seeded our search with two concepts as sources: *Fish Oils* (C0016157) and *Eicosapentaenoic Acid* (C0000545), and two concepts as targets: *Raynaud Disease* (C0034734) and *Raynaud Phenomenon* (C0034735). This is analogous to the search criteria as made by Cameron et al.

In the original study it was stated that dietary fish oil might prevent Raynaud's syndrome. "This is because dietary fish oils (1) inhibit Platelet Aggregation, (2) increase the flow of blood (by reducing Blood Viscosity), and (3) have a regulatory effect on muscle (thereby preventing vasoconstriction and stimulating vasodilation)" (Cameron 2015). As such, rediscovery approaches consider these to be the primary three intermediate topics that should be found.

4.1.1 Platelet Aggregation

It is unsurprising that within our top results, the topic of *Platelet Aggregation* appeared several times under the guise of different MeSH terms, namely Figure A-2, Figure A-3, and Figure A-5. Particularly within biochemistry many things will be closely interrelated, and we can examine these connections via separate paths. The first two connect *Fish Oils* and *Eicosapentaenoic Acid* via *Prostaglandins* to *Epoprostenol* to *Raynaud Phenomenon* (Figure A-2 and Figure A-3). A third then connects *Eicosapentaenoic Acid* to *Platelet aggregation* to *Alprostadil* to *Raynaud Phenomenon* (Figure A-5).

4.1.2 Fish Oils to Raynaud Phenomenon Path via Prostaglandins

As shown in Figure 37, this path is present within our results. The stronger predicate `[Fish Oils INHIBITS Prostaglandins]` reported within Cameron's results, does not exist within the export of SemRepDB and thus must be derived from the full-text.

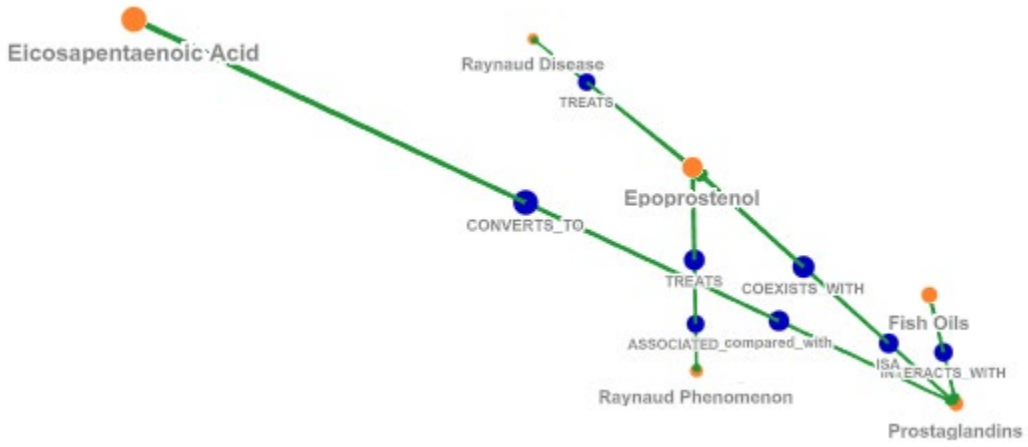


Figure 37. Prostaglandins Subgraph

In fact, none of the core papers have a predicate relating the two. However, multiple papers [`PMID6302714, PMID6303363, PMID6301111`] cite Culp et al. [`PMID7208950`] whose abstract states "The effect of altering the abundance of precursors and inhibitors of prostaglandin formation by dietary supplements of fish oil…" (Culp 1980). This generates a lesser version of the predicate `[Fish Oils INTERACTS_WITH Prostaglandins]`, connecting the two concepts with INTERACTS_WITH rather than INHIBITS.

The second predicate [`Prostaglandins ISA Epoprostenol`] also does not exist in the core non full-text documents - nor does any predicate relating the two. A core paper [`PMID6301111`] cites both [`PMID378308`] by Bayer et al. and [`PMID6773615`] by

Pickard et al. In the abstract of the first paper, the authors state "due to the formation of platelet aggregates it is concluded that PGI2 is a most potent anti-aggregatory prostaglandin" (Bayer 1979), while in the abstract of the second paper the authors state "These results support our suggestion that a prostaglandin, in particular PGI2, is required for hypercapnia to produce full cerebral vasodilatation" (Pickard 1980). In both cases, the predicate [`Prostaglandins ISA Epoprostenol`] is derived, as Epoprostenol is also known as PGI2 or prostacyclin.

The next predicate, [`Epoprostenol TREATS Raynaud Phenomenon`] has multiple predicates within the non full-text core documents. It is present in [`PMID6788326, PMID7037038, and PMID3883365`].

The analysis indicates that this particular path would not be found using only non full-text versions of the core documents, but is present in our results because our algorithm included the cited references.

4.1.3 Eicosapentaenoic Acid to Raynaud Phenomenon Path via Prostaglandins

This path is largely the same as the previously described one, aside from the difference of the first graph edge (Figure 37). Here, the predicate [`Eicosapentaenoic Acid CONVERTS_TO Prostaglandins`] is found in two non full-text core documents [`PMID6314583 and PMID6821892`]. Despite this, because this path also utilizes the [`Prostaglandins ISA Epoprostenol`] step, this path would also not be found using only non full-text versions of the core documents, but is present in ours because bibliographic associations included the additional references.

4.1.4 Eicosapentaenoic Acid to Raynaud Disease Path via Platelet Aggregation

We can see a more obvious interaction using the MeSH concept *Platelet Aggregation* directly in our fourth graph result (Figure 38). The predicate [`Eicosapentaenoic Acid DISRUPTS Platelet Aggregation`] appears in the abstract of one core document [`PMID6303363`] where it was said that Eicosapentaenoic acid "has been reported to be a potent antagonist of platelet aggregation and also to reduce the incidence of cardiovascular disorders" (Terano et al. 1983).

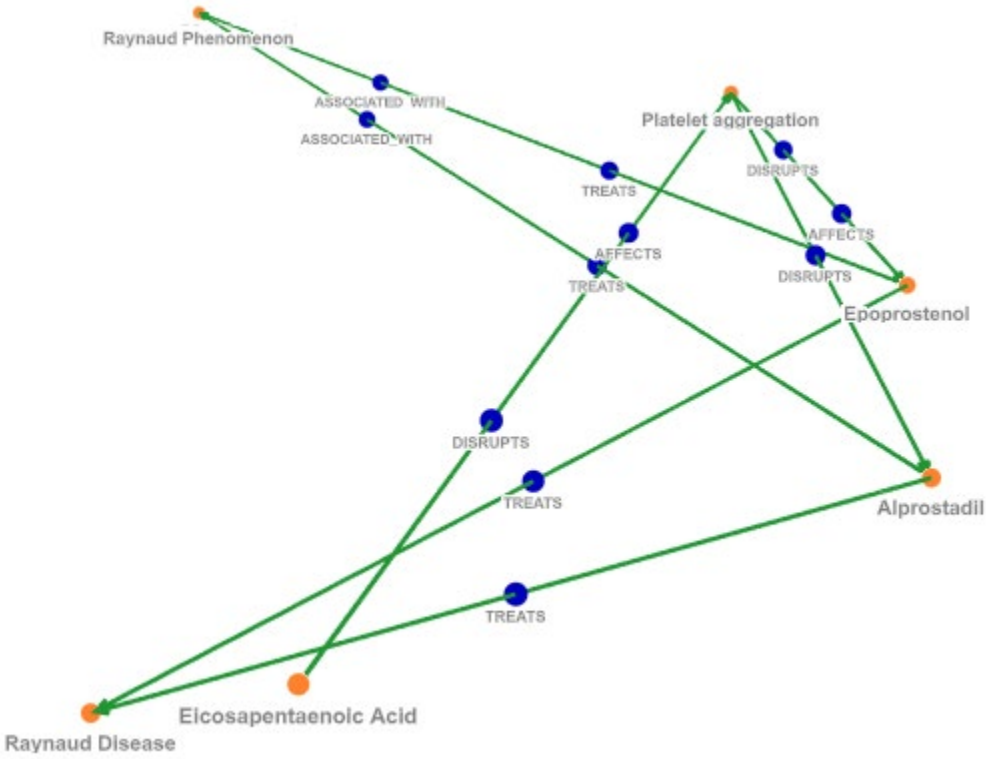


Figure 38. Platelet Aggregation Subgraph

From there, there are associations between *Platelet Aggregation* and both *Epoprostenol* and *Alprostadil.* Although the path between *Epoprostenol* and *Raynaud's Disease* was previously discussed, the predicates [`Epoprostenol DISRUPTS Platelet Aggregation`] and [`Epoprostenol AFFECTS Platelet Aggregation`] are new. These

are generated from one core paper and over ten cited articles. The title of the 1985 core paper [PMID2408588], “Inhibition of platelet aggregation by a new stable prostacyclin introduced in therapy of patients with progressive scleroderma”, produces the first predicate (Keller, J., Kaltenecker, A., Schricker, K. T., Neidhardt, B., & Hornstein, O. P. (1985).

The second [Alprostadil DISRUPTS Platelet Aggregation] does not appear in the non full-text version of any core document. Core documents [PMID6301111 and PMID6302714] cite four different papers [PMID190267, PMID364545, PMID7003784, and PMID7427564] which generate this predicate, and one of which [PMID7427564] is hit on the search terms. As an example, Minkes et al. [PMID190267] state in their abstract “Furthermore, dBcAMP and PGE1 both inhibit platelet aggregation” (Minkes 1977) while Whittle et al. [PMID364545] state “The activity of prostacyclin (PGI2), PGE1 or PGD2 as inhibitors of platelet aggregation” in theirs (Whittle 1978). As Prostaglandin E1 (PGE1), is also known as *Alprostadil*, these yield the predicate in question.

The final predicate that has not been discussed, [Alprostadil TREATS Raynaud’s Disease], appears in two core references [PMID3977414, PMID7259326]. The 1981 paper [PMID7259326] by Martin et al. concludes that *Alprostadil* “may therefore be suitable treatment for Raynaud's phenomenon and the vascular insufficiency of systemic sclerosis and other connective tissue diseases” (Martin et al. 1981)

4.1.5 Blood Viscosity

The second intermediate *Blood Viscosity* was found by Cameron et al. (2015) via the following predicates: [Eicosapentaenoic Acid DISRUPTS Blood Viscosity],

[Fish Oils AFFECTS Blood Viscosity], [Ketanserin DISRUPTS Blood Viscosity], and [Ketanserin TREATS Raynaud Phenomenon]. This shows up in our 15th result (Figure 39).

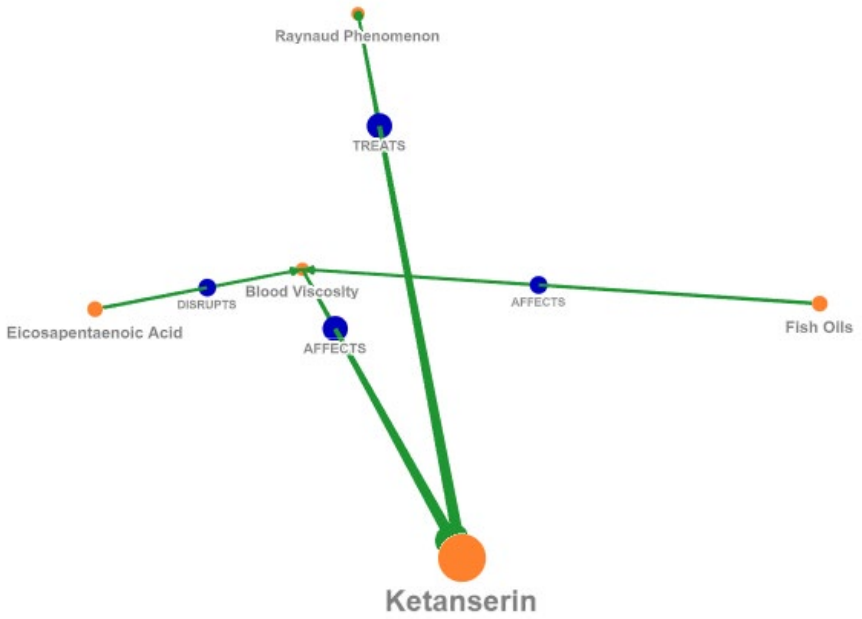


Figure 39. Blood Viscosity Path

The predicates are mostly found directly within the titles and abstracts of core documents. The predicate [Eicosapentaenoic Acid DISRUPTS Blood Viscosity] is from [PMID6303363] where Terano et al. state that "We recently reported that EPA also reduces whole blood viscosity" (Terano 1983). Similarly, the predicate [Fish Oils AFFECTS Blood Viscosity] is contained within the title of Woodcock et al.'s paper [PMID6320945]: "Beneficial effect of fish oil on blood viscosity in peripheral vascular disease" (Woodcock 1984). [Ketanserin TREATS Raynaud Phenomenon] is lastly found in multiple core documents [PMID6812750, PMID6365102, PMID6432198].

The one exception, [Ketanserin DISRUPTS Blood Viscosity] is not found within any of the non full-text documents. Instead, only a weaker relation [Ketanserin AFFECTS Blood Viscosity] exists within Walker et al.'s paper [PMID2412054] where they state "Ketanserin given intravenously for seven days to patients with very severe leg ischaemia, significantly improves whole blood viscosity..." (Walker 1985).

For this path, there is no additional benefit from utilizing bibliographic citations and the only change from not using full-text is the weaker relation between *Ketanserin* and *Blood Viscosity*.

4.1.6 Vascular Reactivity

Cameron et al. reported that they were unable to retrieve the intermediate concept of Vascular Reactivity directly. They later observed that SemRep erroneously parsed the phrase *Vascular Reactivity* (C1660757) into two tokens: *Vascular* (Blood Vessels-C0005847) and *Reactivity* (Reactive-C0205332) (Cameron 2016). Only after adjusting the predicates manually were they able to achieve the results in Figure 40.

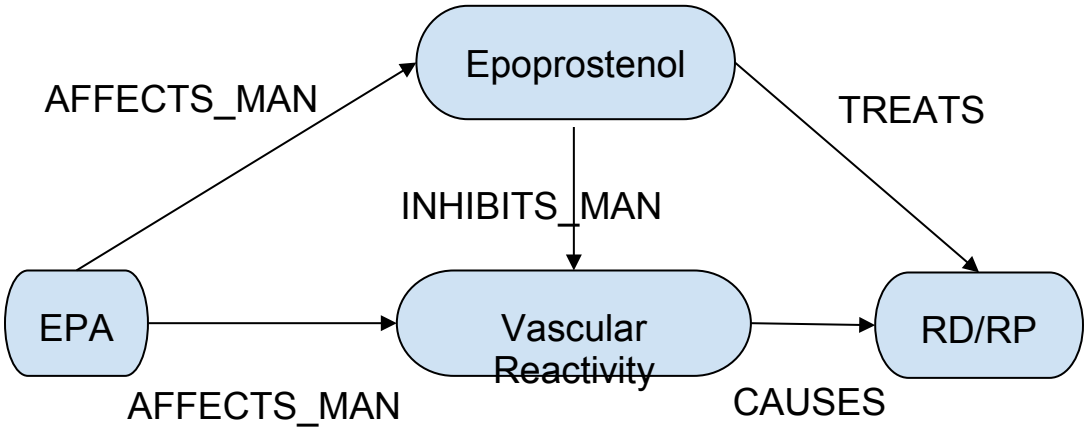


Figure 40. Vascular Reactivity Path from Cameron et al.

In their results, multiple predicates had to be manually adjusted, indicated with the suffix notation of "_MAN" in the graph. Here, these are from [Eicosapentaenoic Acid AFFECTS_MAN Vascular constriction], [Epoprostenol INHIBITS_MAN Vascular constriction] and [Epoprostenol TREATS Raynaud Syndrome].

However, our results (Figure 41) generated a path without any modification to the predicates or utilizing full-text. A particular predicate, [Epoprostenol ISA Vasodilator Agents] can be found within two documents [PMID7003784 and PMID114606]. The first states that "Prostaglandins E1 (PGE1) and I2 (prostacyclin, PGI2) are potent vasodilators and inhibitors of platelet aggregation ", while the second states "As PGI2 is the most potent cerebral vasodilator drug tested". Both of these are cited by one of the core documents [PMID6301111].

The second predicate [Vasodilator Agents TREATS Raynaud Disease] is also only included via citation. The core document by Belch et al. [PMID3883365] cites an earlier paper by Belch et al. [PMID7025341] where they measured the effects of "prostacyclin (PGI2) a potent antiplatelet and vasodilator agent in 5 female patients with Raynaud's syndrome" with positive results (Belch et al. 1981). Thus, the entire path is generated due to the inclusion of bibliographic citations.

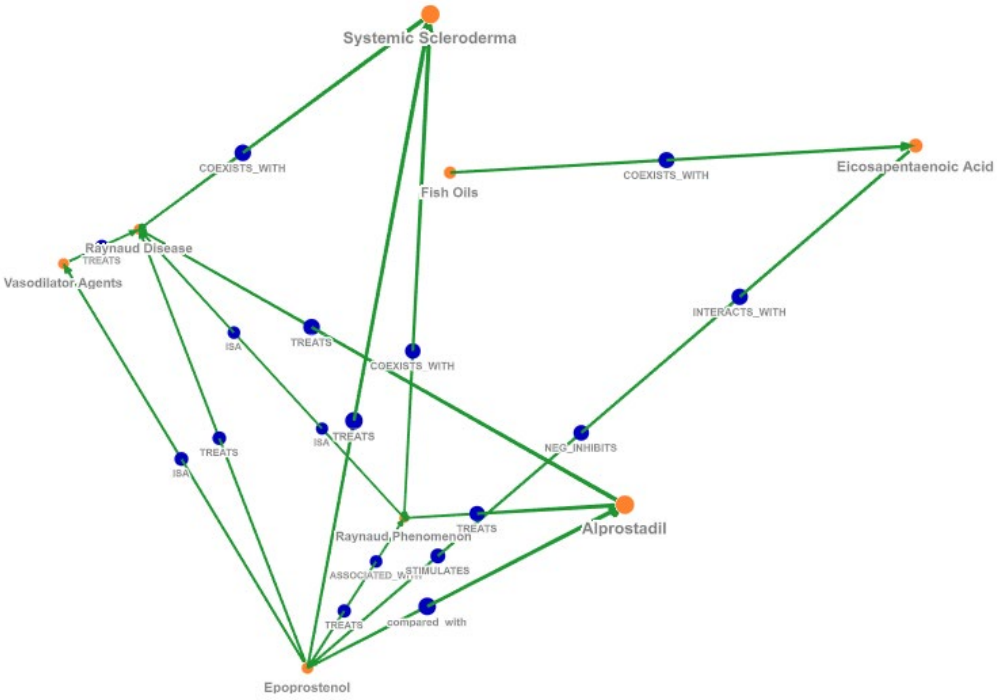


Figure 41. Vascular Reactivity Path

4.1.7 Retrospective Study: Fish Oils and Raynaud Disease - Conclusions

In work later performed by Cameron as part of his dissertation, he attempted to rediscover Swanson's *Raynaud Syndrome–Dietary Fish Oils* hypothesis without the use of full-text. Through a similar predicate analysis he noted that "these results collectively suggest that titles and abstracts alone, might NOT be sufficient" (Cameron et al. 2016). As such, this matches the results of this project in that several of his resulting paths were only available due to the usage of predicates generated from full-text. This project, however, shows that it is possible to reproduce the results of Swanson's *Raynaud Syndrome–Dietary Fish Oils* hypothesis (using a similar corpus) without the use of full-text by modifying Cameron et al.'s approach by incorporating bibliographically related documents.

In Appendix A, Table A-3 details the precision and recall at each ranked result when using citations. Table A-4 does the same when not using citations but otherwise follows the same algorithm. A graph of the results can be seen here in Figure 42.

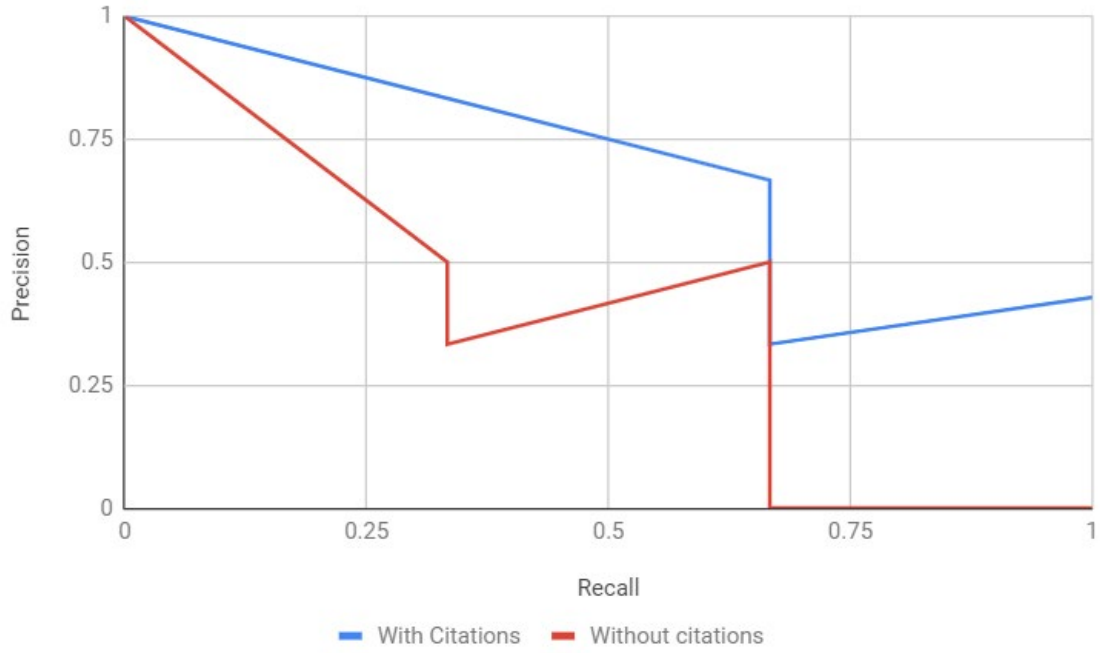


Figure 42. Precision Recall Curve for Fish Oils – Raynaud Disease Study

The shape of this graph departs from the typical sawtooth shape of precision-recall curves. When considering standard document relevancy, a document result is unique and can be considered relevant or not. Here, each result is a subgraph which may contain more than one possible path, and each path could contain several concepts. For example, the very first result using citations was shown in Figure 41. In this graph paths may go through *Epoprostenol*, *Alprostadil*, *Vasodilator Agents*, and *Systemic Scleroderma*, the precise meaning of which may vary: *Epoprostenol* and *Alprostadil* generally imply an association with Platelet Aggregation, *Vasodilator Agents* refers to Vascular Reactivity, and *Systemic Scleroderma*, although related to Raynaud's Disease isn't a direct

connection. This implies that the first search result has a recall of 0.667 having two of the three intermediate topics, and a precision of 0.667 for having two relevant and one non-relevant topics.

A second factor that is different from standard document information retrieval approaches is that a subgraph result may technically contain new information but no new relevancy information. Consider the first (Figure 41) and the second (Figure 37) returned results. The second result has a new concept, *Prostaglandins*, when compared to the first result, but semantically this is still referring to the topic of Platelet Aggregation. It provides additional information into the interconnectedness between the concepts but at the same time does not demonstrate increased recall.

An alternative is to examine the F_1 measure as a function of the top k results (Figure 43). The search without citations only returned a total of eight results, but up to that point using citational information produced a better F_1 score.

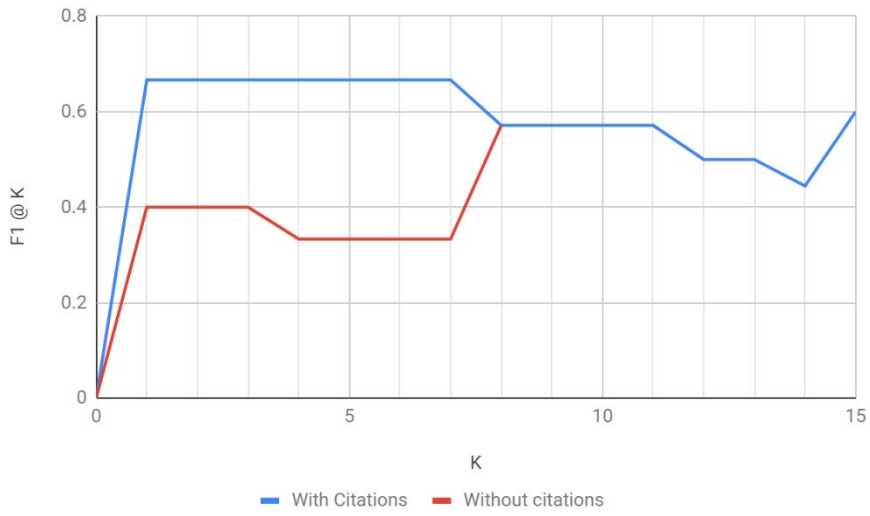


Figure 43. F_1 @ K measure for Fish Oils – Raynaud Disease Study

Beyond the behavior of the graphs themselves, the most noticeable difference is that without using citations, one of the three intermediate concepts (*Vascular Reactivity*) was unable to be recovered.

4.2 Retrospective Study: Magnesium and Migraine Disorders

In his 1987 paper, "Migraine and Magnesium: Eleven Neglected Connections" Swanson described his discovery of how Magnesium deficiency might exacerbate Migraines via eleven different interrelated mechanisms (Swanson 1988). These are listed below in Table 14. We attempted to reproduce these results via the approach described previously.

Table 14. Swanson's Eleven Intermediates for Magnesium-Migraine Disorders

Calcium Channel Blockers
Vascular Mechanisms
Prostaglandins
Epilepsy
Serotonin
Inflammation
Spreading Cortical Depression
Hypoxia
Platelet Aggregation
Stress/Type A personality
Substance P

Our corpus consisted of 131 of the original MEDLINE documents cited by Swanson (Table B-1) along with the other MEDLINE documents they cited, totaling approximately 2,900 references. Some documents were missing abstracts (and thusly missing SemRep abstract-generated predicates), for these abstracts/introductions were located and added manually along with any generated predicates. A listing of the predicates along and their corresponding sentences are recorded by document identifier within Table B-2.

Unless specified otherwise, each search was performed with a source concept of *Magnesium* (C0024467) and a destination of *Migraine Disorders* (C0149931). Even today the exact nature of migraine hasn't been completely solved and there is debate if it is primarily a vascular (pertaining to blood vessels), neurological (pertaining to neurons), or a mixture of the two. We examined each of the eleven intermediates that Swanson discovered to determine if our system has recovered the discovery. In addition, we will discuss a twelfth intermediate concept, Diabetes, not identified by either Swanson or Cameron.

4.2.1 Calcium Channel Blockers

It stands that if the source of pain is caused by vasoconstriction, then a vasodilator should provide relief. Peroutka explained that "calcium channel antagonists are a recently developed class of vasodilators that prevent the influx of calcium into vascular smooth muscle. The unique pharmacologic effects of these agents provide a theoretical basis for their use in the treatment of migraine" (Peroutka 1983).

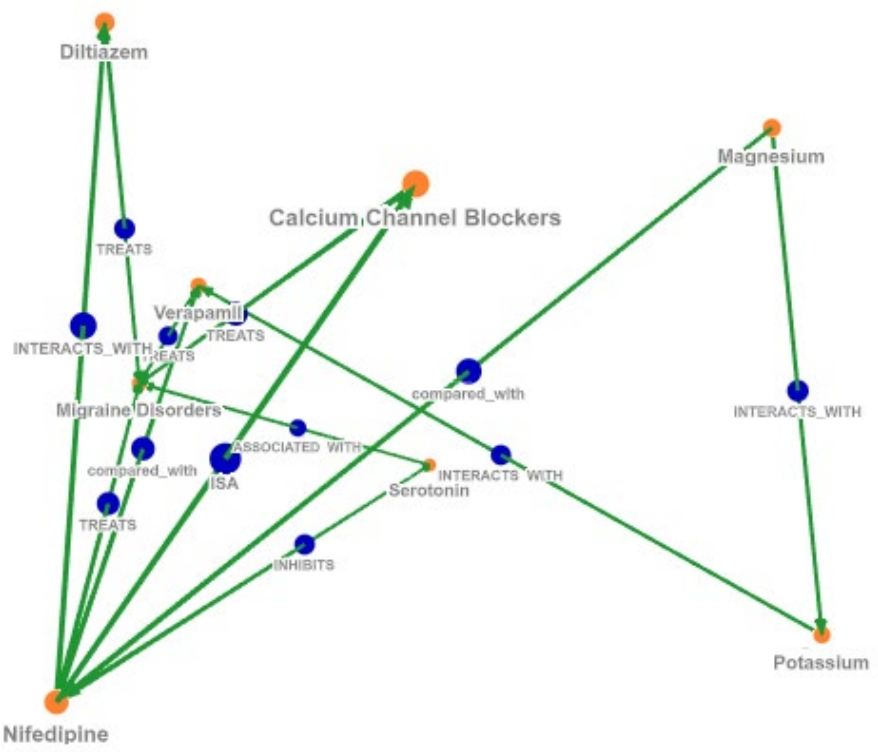


Figure 44. Calcium Channel Blockers

Our second search result using citations are shown in Figure 44. Known calcium channel blockers, Verapamil and Nifedipine, both treat Migraine as suggested by multiple articles. The article by Markley, Cheronis, and Piepho [PMID6539877] entitled "Verapamil in prophylactic therapy of migraine" states "Verapamil significantly reduced both headache frequency and duration with few side effects" (Markley, Cheronis, and Piepho 1984). The article by Meyer and Hardenberg [PMID6358126] indicates that "Nifedipine and verapamil provided equivalent relief for cluster but produced more side effects, and were less effective, than nimodipine in control of migraine" (Meyer and Hardenberg 1983). Similar statements to the use of Verapamil and Nifedipine in the use of migraines can be found in other documents. [PMID2425960, PMID6339937, PMID2425960] All of these produce at least one of the two predicates

[C0042523|Verapamil - TREATS - C0149931|Migraine Disorders] and [C0028066|Nifedipine - TREATS - C0149931|Migraine Disorders].

The presence of Magnesium acts like a Calcium antagonist, having been compared to both Verapamil and Nifedipine. The predicates connecting *Magnesium(C0024467)* to *Verapamil(C0042523)* and *Nifedipine(C0028066)* can be found in two documents [PMID7297597, PMID3458981]. In the paper by Sjögren A. and Edvinsson L. [PMID3458981] it was indicated that "the order of potency for eliciting relaxation was: nifedipine greater than verapamil greater than magnesium" (Sjögren and Edvinsson 1986). While in the second reference by Turlapaty, Weiner, and Altura [PMID7297597] it is noted that a deficiency of magnesium induces vascular constriction: "Previous studies on isolated blood vessels indicate that [a] withdrawal of magnesium ($[Mg2+]_0$) induces calcium-dependent contractile responses" (Turlapaty, Weiner, and Altura 1981).

In Andersson's 1986 paper [PMID2424267] he calls out several calcium channel blockers in his statement: "The cardiovascular effects of different calcium channel blockers (CCB), exemplified by nifedipine, verapamil and diltiazem, are not identical" (Andersson 1986). Other papers also illustrate this relationship [PMID2981405, PMID6202853, PMID6708731, PMID6869560, PMID7195070] (Figure 45).

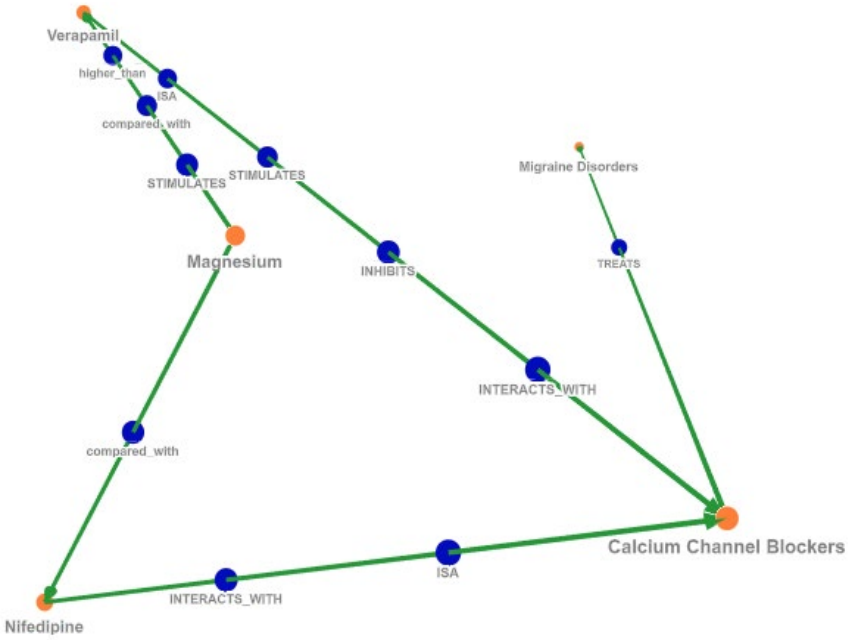


Figure 45. Calcium Channel Blockers Subgraph

Multiple documents associate *Calcium Channel Blockers* with *Migraine Disorders*, including the paper by Peroutka mentioned earlier [PMID6358127, PMID3521194, PMID6715160]. Taken together these explicitly show that both *Nifedipine* and *Verapamil* are *Calcium Channel Blockers* and that the blockers as a whole are used as treatment for migraines.

4.2.2 Vascular Mechanisms

Multiple subgraphs were examined connecting *Magnesium* and *Migraine Disorders* via vascular mechanisms. In the following sections we examine *Dilatation*, *Blood Vessels*, *Contraction*, and *Vasospasm* to understand the underlying connections.

4.2.2.1 Dilatation

The 1977 paper by Hachinski et al. [PMID597797] attempted to examine regional cerebral blood flow (rCBF) and dilation during migraines. One theory of migraine at the time conjectured that the primary mechanism consists of a series of events initiated by vasospasm of the cerebral arteries followed by a reactive vasodilation. The distension of the vessels was assumed to be the source of the headache pain. Although their findings did not support the speculation, it did not discount distention of the intracranial arteries during migraine headache (Hachinski et al. 1977). The predicate seen in the fourth search result (Figure 46), [C0012359|Pathological Dilatation - COEXISTS_WITH - C0149931|Migraine Disorders], was extracted from Hachinski's paper, signifying one of the underlying possibilities for migraines.

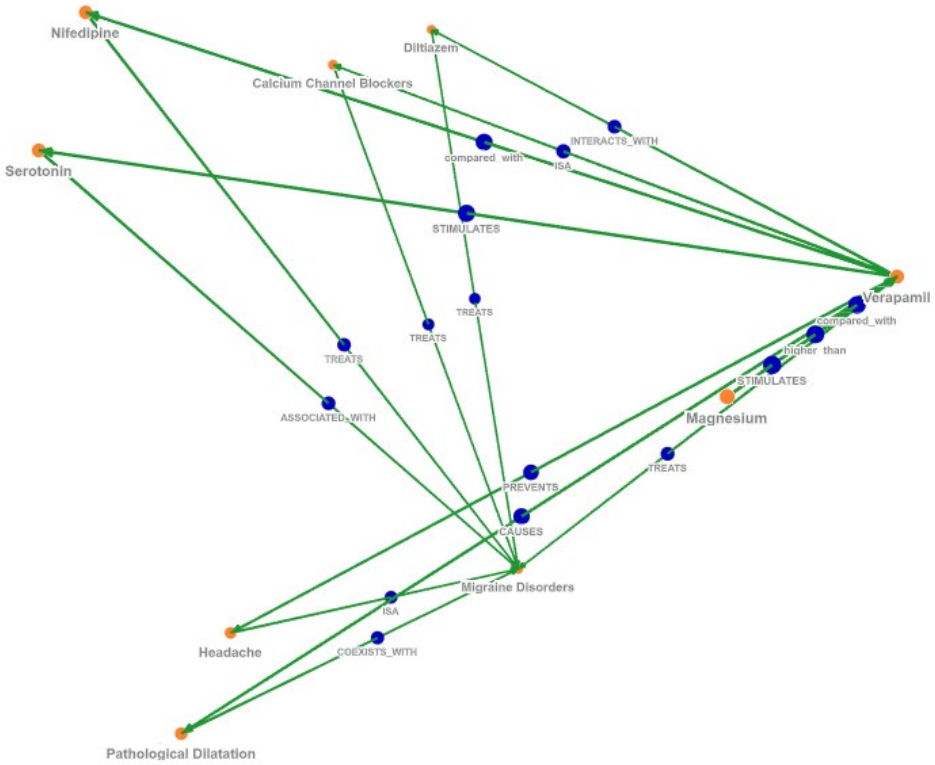


Figure 46. Vascular Mechanisms (Dilatation) Subgraph

The interaction along calcium channels has already been discussed, both in terms of Verapamil and Magnesium acting as a vasodilator, or in the case of Magnesium deficiency, acting as a vasoconstrictor.

4.2.2.2 Vascular Mechanisms - Summary

Our resulting subgraphs have shown vascular connections between *Magnesium* and *Migraine Disorders* using a path length of three using our constrained corpus. If we were to extend our corpus to include the additional documents that Cameron utilized within his study then several additional predicates would have appeared and allowed a connection at a path length of two. The 1972 article [PMID4260015] "Magnesium metabolism from the viewpoint of cardioangiology. II. Findings in magnesium metabolism in vascular diseases" by Wustenberg et al. generates the predicate [C0024467|Magnesium - ASSOCIATED_WITH - C0042373|Vascular Diseases]. Although *ASSOCIATED_WITH* is not as strong of a direct implication as *INHIBITS*, implied by Swanson, its existence would have been sufficient reason to investigate paths of length three. Additionally, the predicate [C0149931|Migraine Disorders - ISA - C0042373|Vascular Diseases] would have been generated from Domzal's 1975 article [PMID1153064], wherein he states "EEG changes suggest--according to the author--that migraine is a primary cerebral and only secondarily a vascular disorder" (Domzal 1975).

4.2.3 Prostaglandins

As is common with many biochemical systems, a particular chemical can have one effect in larger concentrations, while a deficiency will provoke the opposite response

as a means of regulation. As Horrobin's experimental results showed this was the case for prostaglandins - indicating that higher concentrations can cause vascular dilation, while the opposite was true in low concentrations (Horrobin 1977). Various experiments have shown that the administration of Prostaglandins could cause migraine-like symptoms, such as in Carlson's study where Prostaglandin E1 was infused intravenously into eight healthy male subjects with no prior history of migraine - all experienced symptoms with a sufficient dosage (Carlson, Ekelund, and Orö 1968).

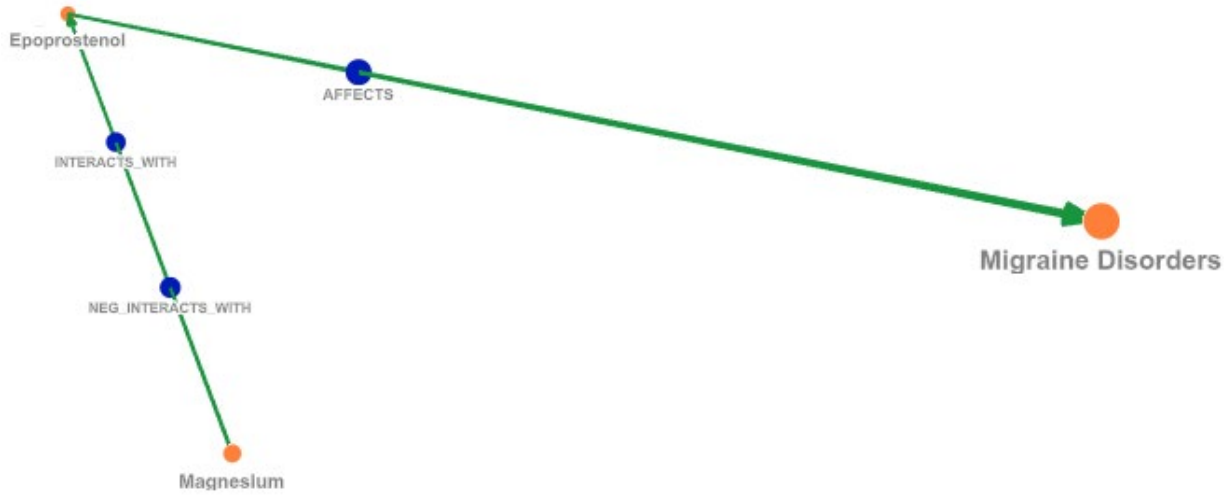


Figure 47. Prostaglandins Subgraph – Path Length of 2

The title of their article, "The effect of infused prostacyclin in migraine and cluster headache", by Peatfield et al. [PMID7026501] implies some relationship between prostacyclin (Epoprostenol), a type of Prostaglandin, and migraines. They noted that "We have assessed the role of prostacyclin by infusing it into eight patients with migraine and eight with cluster headache. Most of the subjects developed a dull throbbing headache during the infusion" (Peatfield, Gawel, and Rose 1981). The title itself generates two different predicates: [C0033567|Epoprostenol - AFFECTS - C0009088|Cluster

Headache] and [C0033567|Epoprostenol - AFFECTS - C0149931|Migraine Disorders].

Meanwhile, Prostacyclin was noted by Briel, Lippert, and Zahradnik [PMID3898850] to be a "potent naturally occurring vasodilator and inhibitor of platelet aggregation", then continue to state that "Magnesium is known to exert an inhibitory effect on coagulation and platelet function", thus prompting them to investigate the effects of "magnesium sulfate on vascular prostacyclin synthesis and platelet prostacyclin interaction" (Briel, Lippert, and Zahradnik 1985). This last statement having generated the predicate [C0024467|Magnesium - INTERACTS_WITH - C0033567|Epoprostenol] in Figure 47.

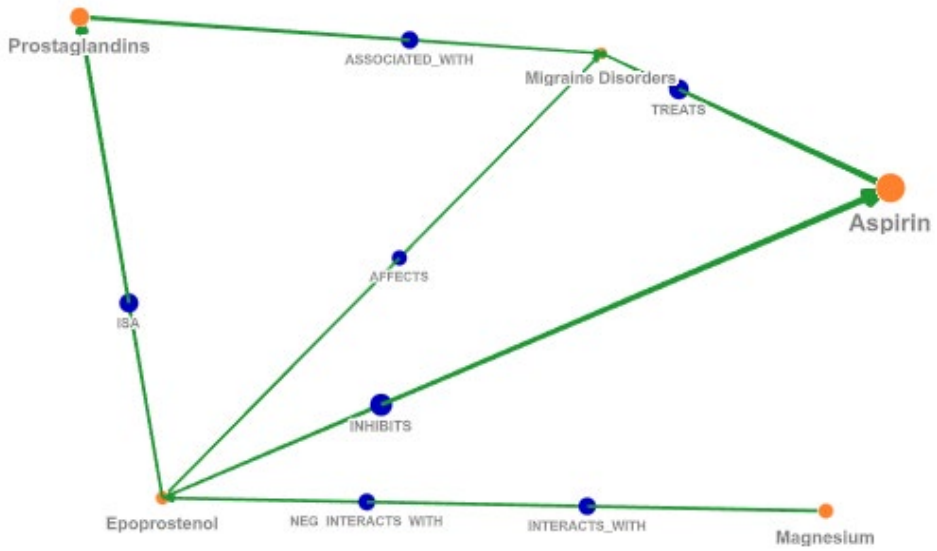


Figure 48. Prostaglandins Subgraph – Path Length of 3

Using a path length of three we can see some of the additional details that we would expect (Figure 48).

The predicate [C0033567|Epoprostenol - ISA - C0033554|Prostaglandins] is derived from the sentence "Prostacyclin appears to be the most active of the vasodilating prostaglandins with inflammatory and hyperalgesic properties" in the article by Peatfield et al. [PMID7026501] (Peatfield et al. 1981). Two documents [PMID4016946, PMID89390] also generate the predicate [C0033554|Prostaglandins - ASSOCIATED_WITH - C0149931|Migraine Disorders]. Both articles make a claim that "Prostaglandins (PG), particularly PGE, may be linked to the pathophysiology of migraine", namely because "PGEs cause vasodilation and hyperalgesia, both typical reactions of inflammation". This view that vascular headache is an "inflammatory reaction" explains the "local role of PGs and the effectiveness of PG-inhibitors in the treatment of migraine" (Parantainen, Vapaatalo, and Hokkanen 1985).

It has been shown that *Magnesium Deficiency* can increase prostaglandin production. This was noted by Soma et al. in their 1988 paper [PMID3238000] wherein their findings suggested "that the biosynthesis of eicosanoids, mainly PGI2, is stimulated in Mg deficiency, and this may provide protection against intracellular Mg depletion and Ca accumulation, so as to counteract to the constricted and hyperreactive state of the vasculature in such a condition" (Soma et al. 1988.)

4.2.4 Epilepsy

Migraines have been a mystery for decades due to the mixture of vascular and neurogenic symptoms. Swanson noted that "a connection between migraine and epilepsy has been suspected for most of the past century" (Swanson 1988). In 1969, Basser wrote his paper [PMID4978139] entitled "The Relation of Migraine and Epilepsy" which

illustrated their study into a potential relationship between the two: “It is the purpose of this paper to examine the question of the relation of migraine and epilepsy by an analysis of 1,800 cases of migraine seen at the Northcott Neurological Centre since 1951 with particular reference to those cases in which migraine and epilepsy were manifested in the same individual” (Basser 1969). Our first search result (Figure 49) shows the predicate [C0014544|Epilepsy - COEXISTS_WITH - C0149931|Migraine Disorders] derived from this.

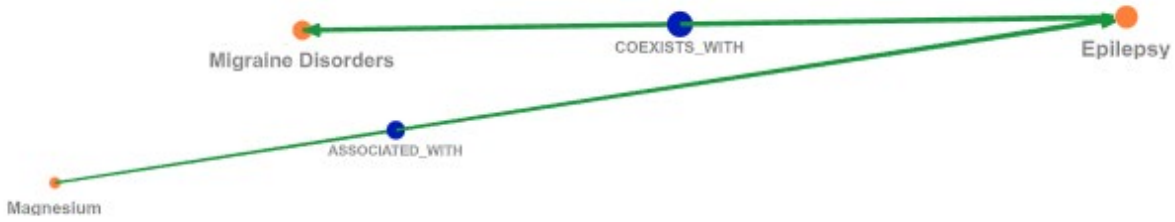


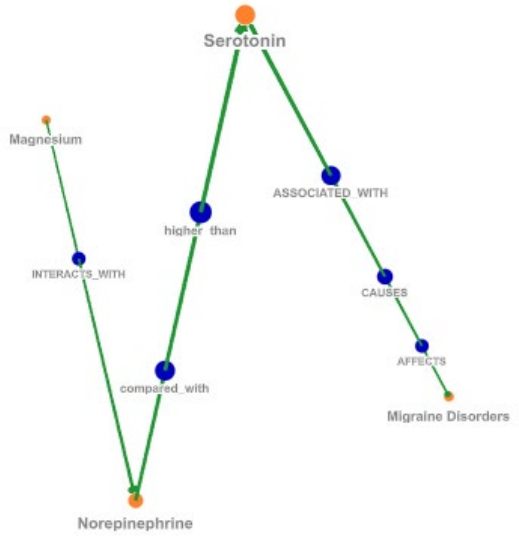
Figure 49. Epilepsy Subgraph – Path Length of 2

Canelas et al. wrote a paper in 1965 [PMID14338128] entitled “Disorders of Magnesium Metabolism in Epilepsy”, from which the predicate [C0024467|Magnesium - ASSOCIATED_WITH - C0014544|Epilepsy] was derived. As an interesting note, in their introductory paragraph they state that “The possibility that convulsions may occur in Mg deficiency led some investigators to study the metabolism of this metal in the epilepsies, and a trend to low blood concentrations was usually found” (Canelas, De Assis, and De Jorge 1965). A second article, [PMID108361], also generated this predicate.

4.2.5 Serotonin

Serotonin is present in both blood platelets and blood serum that acts as a neurotransmitter and causes vascular constriction. Swanson noted that Hilton and Cumings discovered "increased platelet aggregability and sensitivity to serotonin release in headache-free migraine patients", and that many researchers have also suggested that platelets and serotonin have some association with migraines (Swanson 1988). Several studies within our corpus have associated Serotonin to migraines [PMID6275679, PMID5658323, PMID5297855, PMID1254466, PMID1146499]. Each producing a predicate to connect the two concepts (Figure 50).

"Serotonin is released from platelets during a migraine attack, thus causing a transient increase in plasma serotonin which, through its vasoconstrictive action, is then thought to play a role in the development of the headache phase of the attack. Plasma serotonin is rapidly metabolized, leading to a drop in blood levels of serotonin as a consequence of a migraine episode" (Swanson 1988). It has been further suggested that a serotonin releasing factor is present within the blood of migraine patients. Free fatty acids (Figure 51) have been suggested for this (Anthony 1982).



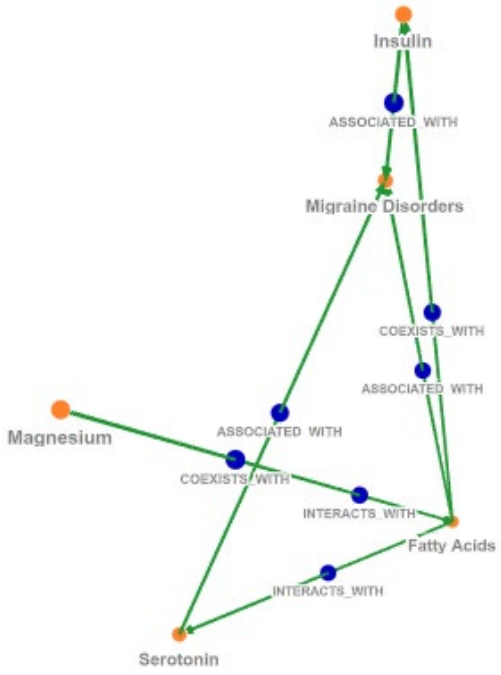


Figure 50. Mg - Migraine search result 14 *Figure 51. Mg – Migraine search result 5*

Our corpus contained six references that connected *Serotonin* to vasoconstriction [PMID6948814, PMID6538269, PMID4814375, PMID379918, PMID7060640, PMID1262918]. One example [PMID6948814] of which states that "In isolated human pial arteries (diameter 0.4-0.5 mm), contractions were produced by potassium, noradrenaline, serotonin, and prostaglandin F2 alpha" (Brandt, Andersson, Edvinsson, and Ljunggren 1981). Given that serotonin is a known vasoconstrictor this is no surprise. We have also previously noted the vasodilator effects of *Magnesium* and *Calcium Channel Blockers*. The predicate [C0036751|Serotonin - STIMULATES - C0042523|Verapamil] is derived from [PMID7150872] in the sentence "The calcium channel blocking agent, verapamil (10(-6)M), inhibited completely contractile responses to KCl; contractile responses elicited by angiotensin II and 5-HT were attenuated by verapamil" (Altura and

Turlapaty 1982). This matches our expectations (Figure B-5). The same is true for the predicate [C0028066|Nifedipine - INHIBITS - C0036751|Serotonin] derived from [PMID450213] and is seen in Figure B-3.

Our results (Figure 50 and Figure 51) have shown interconnections between *Magnesium* and *Migraine Disorders* via *Serotonin* using a path length of three within our constrained corpus. Cameron states that the 1987 article [PMID3629724] by Pertseva et al. produces the predicate [Magnesium - INTERACTS_WITH - Serotonin] (Cameron 2014). In this case, by extending our corpus to include the additional document that Cameron utilized within his study then we gain an additional predicate that would have appeared and allowed a connection at a path length of two. However, upon inspection, the abstract of the article states "A stimulating effect of serotonin, guanine nucleotides and sodium fluoride is found as well as a dependence of the catalytic activity of adenylate cyclase on magnesium and manganese ions" (Pertseva and Soltitskaia 1987). This produces the predicate [C0001492|Adenylate Cyclase - INTERACTS_WITH - C0024467|Magnesium], which is different than expected.

4.2.6 Inflammation

In his paper Swanson posits that "pain is a classic sign of inflammation, and migraine has been described as a sterile inflammatory disease of cranial blood vessels" (Swanson 1988). Acute inflammation is generally characterized by vasodilation, increased permeability, and increased blood flow. In 1979 Moskowitz et al. [PMID90971] proposed a mechanism for pain in headache based upon the abnormal release of Substance P from trigeminal nerve terminal terminals. The trigeminal nerve is the largest

of the cranial nerves and provides the primary pathway for the transmission of head pain. Moskowitz also hypothesized that pain was the result of an induced release of vasodilators such as Substance P from the nerve fibers (Moskowitz, Romero, Reinhard, Melamed, and Pettibone 1979).

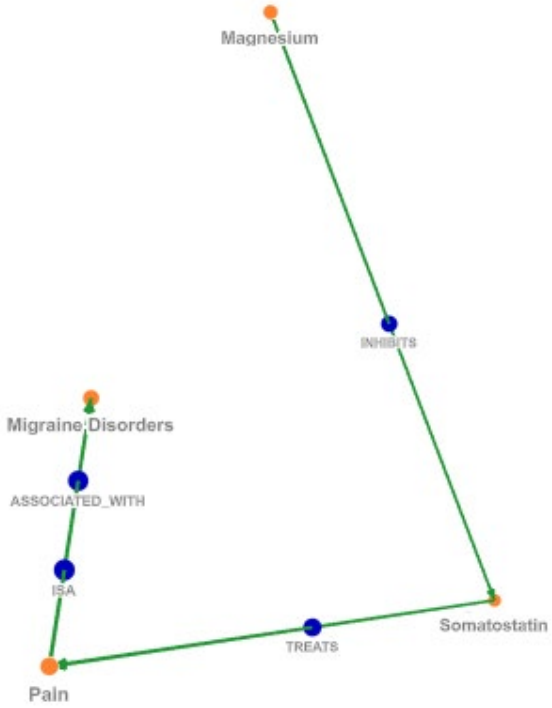


Figure 52. Inflammation Subgraph

The relationship between *Pain* and *Migraine Disorders* is fairly common as one would expect. The titles of several articles indicate as such: [PMID3585352, PMID6830162, PMID7310427, PMID20784580], each generating the predicate [C0030193|Pain - ASSOCIATED_WITH - C0149931|Migraine Disorders]. However, given the complexity of migraine, the explanation is not as straightforward.

Sicuteri, Renzi, and Geppetti wrote an article in 1986 [PMID2433912] where they construct the predicate [C0037659|Somatostatin - TREATS - C0030193|Pain] for

headaches and further implicate the involvement of *Substance P*. “Opiates and somatostatin inhibit the release of substance P from primary sensory neurones and relieve both pain and autonomic symptoms of cluster headache attack” (Sicuteri, Renzi, and Geppetti 1986).

To complete the association as shown in Figure 52, we have to determine the relationship between *Magnesium* and *Somatostatin*. An article within our corpus sheds some light onto this and provides the necessary predicate. Curry and Bennett’s article [PMID6141690] “Magnesium requirement for somatostatin inhibition of insulin secretion” suggests from the title alone that *Magnesium* is necessary for *Somatostatin* to function properly - at least in certain capacities. The predicate relationship, INHIBITS, in this case is misleading but upon examination fits into our expectations. As we’ve seen so far it is the case of Magnesium deficiency which is in relation to Migraine conditions. Here, a lack of *Magnesium* causes *Somatostatin* to not function properly.

4.2.7 Spreading Cortical Depression

Spreading Cortical Depression (SCD), or Spreading Depression of Leao is described as a self-propagating wave of depolarization across the cerebral cortex. Experimentally it could be induced by means of a stimulus applied to the cerebral cortex, and results showed that it was accompanied by vasodilation of the arteries. SCD has also been found to be induced by hypoxic conditions and has been implicated in migraines. (Swanson 1988). Additional research into spreading depression has shown two different types, one based upon glutamate and the other on potassium. The introduction of Magnesium ions were able to block the potassium based SCD [PMID739264,

PMID6150068] (Van Harreveld 1978, 1984), while a deficiency induced spreading depression in some situations [PMID3763042, PMID3801897] (Walther, Lambert, Jones, Heinemann, and Hamon 1986; Anderson, Lewis, Swartzwelder, and Wilson 1986).

Unfortunately, our system was unable to produce a subgraph that could relate *Magnesium* and *Migraine Disorders* via *Spreading Cortical Depression* using a path length of 3 or less. It is expected that clinical verbiage used within abstracts did not use a consistent terminology to be able to produce the necessary predicates, which is consistent with the results of Cameron et al.

4.2.8 Brain Hypoxia

Hypoxia is defined as a deficiency in the amount of oxygen reaching a particular section of tissue. In 1974, Bucking and Baumgartner suggested [PMID4842482] that migraine and spreading depression might be initiated by cerebral hypoxia. They found that experimentally induced hypoxia caused a burst of cerebral electrical activity that resembled spreading depression (Bücking and Baumgartner 1974). It was later suggested that the mechanisms of both SCD and brain hypoxia/anoxia are related in the ionic permeability of brain cells (Hansen and Zeuthen 1981).

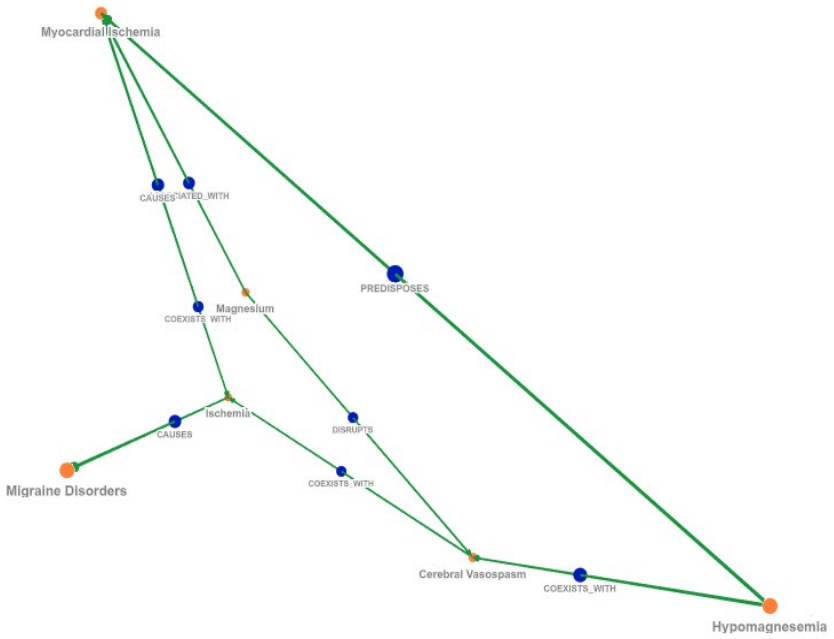


Figure 53. Hypoxia Subgraph

As can be seen in Figure 53, the predicate [`C0024467|Magnesium - ASSOCIATED_WITH - C0151744|Myocardial Ischemia`] is generated by another article by Altura and Altura [`PMID6149922`]. They note that "Mg++ ions are essential for regulation of Na+ and K+ transport across cell membranes" and "under conditions where cellular Mg++ is depleted (e.g. hypoxia, ischaemia, anoxia)" (Altura and Altura 1984). Additionally of note, they state "A reduction in the extracellular Mg++ concentration can produce hypertension, coronary vasospasm and potentiation of vasoconstrictor agents by allowing excess entry of Ca++; concomitantly, the potency of vasodilator agents is reduced" (Altura and Altura 1984). The predicate [`C0151723|Hypomagnesemia - PREDISPOSES - C0151744|Myocardial Ischemia`] comes from similar statements in [`PMID527222`].

Both Altura BM's 1979 article [PMID390330] on ischemic heart disease and the 1976 reference by Hearse et al. [PMID939020] produce the associations between *Ischemia* and *Myocardial Ischemia*. Of lesser interest, but related biochemically we see *Cerebral Vasospasm*. Previously we have seen a connection between *Magnesium* and *Cerebral Vasospasm*. For Figure 53, all three predicates [C0024467|Magnesium - DISRUPTS - C0265110|Cerebral Vasospasm], [C0265110|Cerebral Vasospasm - COEXISTS_WITH - C0151723|Hypomagnesemia], and [C0265110|Cerebral Vasospasm - COEXISTS_WITH - C0022116|Ischemia] come from [PMID6680619] by Altura and Altura.

The final link, [C0022116|Ischemia - CAUSES - C0149931|Migraine Disorders], can be found in two documents. The first by Skinhoj E and Paulson OB [PMID5803690], comes from the following statement: "According to the classical theory the first (prodromal) phase in migraine is caused by ischemia within the internal carotid system and the second phase (headache) by a vasodilation, especially within the external carotid system" (Skinhoj and Paulson 1969). The second from Welch et al. [PMID184066] where they state that "Since biochemical abnormalities reported herein were common to occlusive CVD and migraine headache, it seems probable that they are due to ischemia associated with both conditions and possibly related to the resultant disorder of cerebral energy metabolism" (Welch et al. 1976).

4.2.9 Platelet Aggregation

The hypothesis that migraine is a platelet disorder was proposed by Hanington citing evidence in the difference in the behavior of platelets in migraine patients

(Hanington, Jones, Amess, and Wachowicz 1981; Hannington 1986). Other research has since disputed this (Steiner, Joseph, and Rose 1985; Joseph and Welch 1987), instead suggesting that differences in platelet behavior is a consequence to other changes such as stress, the release of catecholamines, FFAs, serotonin, and prostaglandins (Swanson 1988).

Previously we have seen that Prostacyclin (Epoprostenol) inhibits platelet aggregation and that Magnesium can enhance this effect when present, or inhibit it during the case of Magnesium deficiency (Figure 48).

A more direct connection to the concept *Platelet Aggregation* shows up at length 4 (Figure 54), but there are many paths of little value generated at this length. In his paper Cameron mentions a 1987 article by Briel et al. [PMID2440758] that confirms the predicate [Magnesium - INHIBITS - Platelet Aggregation]. However, this paper does not generate the predicate with SemRep.

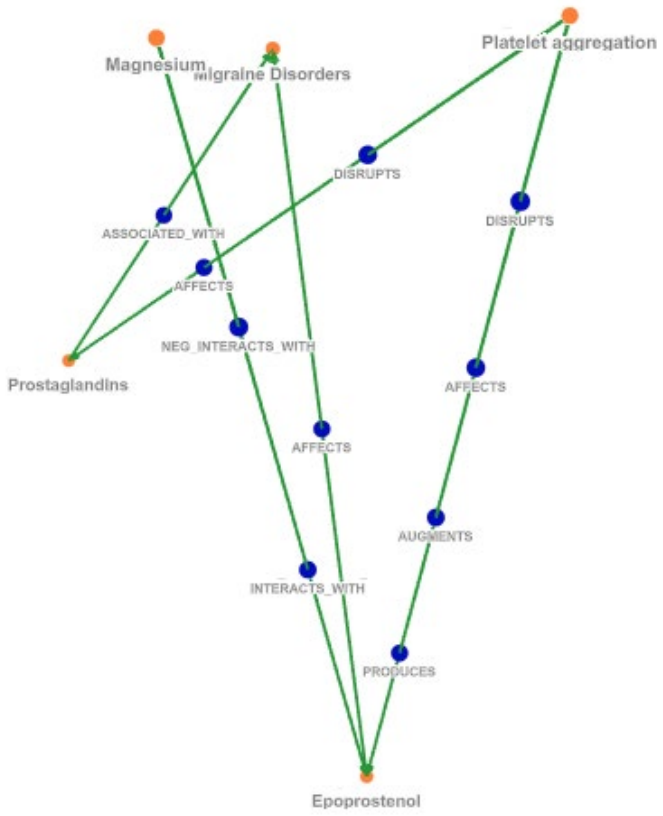


Figure 54. Platelet Aggregation Subgraph – Path Length of 4

The other paths of note in Figure 54 connect *Prostaglandins* and *Platelet Aggregation*. We can find the predicate [C0033567|Epoprostenol - DISRUPTS - C0032176|Platelet aggregation] from "Prostacyclin (PGI2), a strong vasodilator of cerebral vessels and potent inhibitor of platelet aggregation" inside [PMID7048862] (Quintana, Konda, Ishibashi, Yoshimoto, and Suzuki 1982) and other documents [PMID7014260, PMID6992233, PMID6117893, PMID6110816, PMID6117893]. The other predicate [C0033554|Prostaglandins - DISRUPTS - C0032176|Platelet aggregation] is derived from both [PMID372242] and [PMID361756].

4.2.10 Stress / Type A Personality

Swanson noted that the "descriptions of migraine-associated personality in the older literature bear a close resemblance to descriptions of what is now called the type A personality" (Swanson 1988). Several studies have reported a correlation between those of type A personality and frequency of headaches.

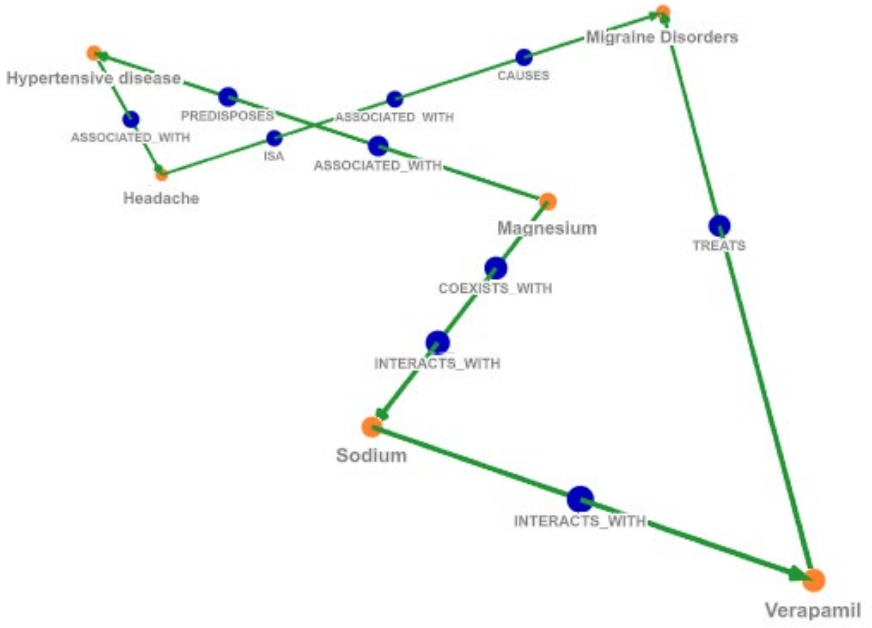


Figure 55. Stress Subgraph

Our study shows a meaningful path connecting *Magnesium* and *Migraine Disorders* that related to *Stress*/*Type A personality* by means of *Hypertensive disease*. In the 1954 article by Friedman et al. [PMID13214278] entitled "Migraine and tension headaches; a clinical study of two thousand cases" the authors state "Among this group are headaches associated with hypertension, arteriosclerosis, infection, brain tumor, hematoma, allergy, and cranial trauma" which generates the predicate [C0018681|Headache - ASSOCIATED_WITH - C0020538|Hypertensive disease] in Figure 55 (Friedman, Von Storch, and Merritt 1954).

Swanson explained that “stress causes an increases in blood levels of free fatty acids, which in turns induces loss of blood magnesium”, or in other words, “type A behavior in effect entails a virtual deficiency of Magnesium” (Swanson 1988). This is backed up by McCarron [PMID6847018] who states that “Many studies suggest that reduced consumption of calcium or magnesium is associated with an increased risk of developing hypertension and cardiovascular disease” (McCarron 1983) which generates the predicate [C0024467|Magnesium - PREDISPOSES - C0020538|Hypertensive disease].

4.2.11 Substance P

During our prior analysis of a connection involving inflammation/pain, the neuropeptide “Substance P*”* was referred to as another possible factor in the genesis of migraines. Our study did not find a meaningful path connecting *Magnesium* and *Migraine Disorders* that related to *Substance P*. However, an exploratory analysis did find a connection utilizing the concept *Cluster Headache*.

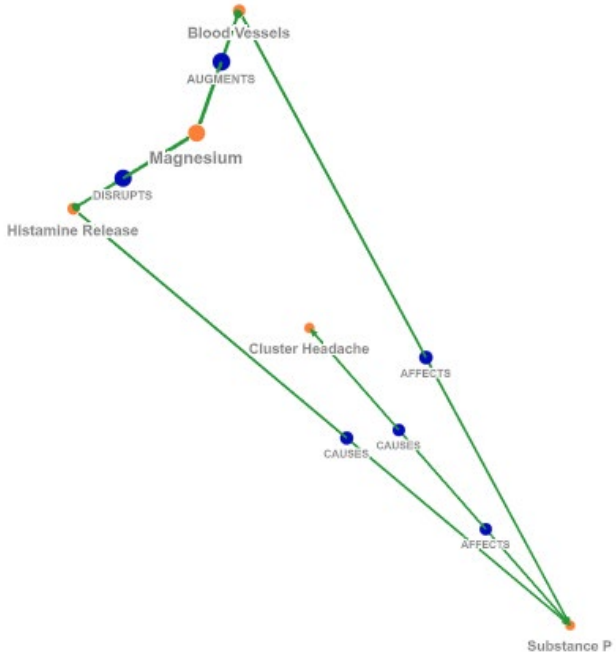


Figure 56. Substance P Subgraph – Path Length of 3

Two papers within our corpus related *Substance P* to *Cluster Headache*. The first, was a 1983 paper by Sicuteri et al. [PMID6193886] where their title, "Substance P and endogenous opioids: how and where they could play a role in cluster headache" generates the graph predicate. They state that "Substance P appears to be involved in the transmission of pain signals from the periphery to the spinal cord and brain stem" and that "Substance P containing neurons are responsible for the neurogenic vasodilation" (Sicuteri, Rainò, and Geppetti 1983). The second paper, "Substance P in the human iris: possible involvement in echothiophate-induced miosis in cluster headache" by Fanciullacci et al. [PMID2452018], also connects the two concepts by title alone (Fanciullacci et al. 1988).

The predicate [C0038585|Substance P - CAUSES - C0019595|Histamine Release] are produced from the sentence "Substance P (SP) induces histamine release

from isolated rat peritoneal mast cells at concentrations of 0.1-10 muM" written in an article by Fewtrell et al. [PMID6184468] (Fewtrell et al. 1982). In the same avenue, the connection between *Substance P* and *Blood Vessels* comes from sentence "Indirect evidence has abundantly been presented to support the view that substance P (SP) is involved in the vasodilatation following activation of fine calibre pain fibres", by Rosell et al. [PMID6171999] (Rosell et al. 1981). Both of these connections fit our earlier reasoning that inflammation, pain, and Substance P are interrelated.

The one new connection in Figure 56 is [C0024467|Magnesium - DISRUPTS - C0019595|Histamine Release]. This predicate also comes from Fewtrell et al. [PMID6184468] where they state that "Extracellular calcium (0.1-1 mM), magnesium (1-10 mM) and cobalt (0.01-0.1 mM) all inhibit SP-induced histamine release when added before the peptide" (Fewtrell et al. 1982). It therefore stands that if the presence of Magnesium inhibits SP-induced histamine, then a magnesium deficiency would cause help provoke headaches in this context.

4.2.12 Diabetes

In addition to the aforementioned eleven connections (Table 14) there was one unexpected concept that showed up in our initial graph (Figure 57) of length two paths - *Insulin*. The predicate [C0024467|Magnesium - COEXISTS_WITH - C0021641|Insulin] is derived from Flink et al. [PMID572225] where the authors note that "FFA fell by a maximum of 65% or 0.44 meq/liter and Mg fell by a maximum of 0.31 meq/liter during the glucose-insulin infusions" (Flink, Shane, Scobbo, Blehschmidt, and McDowell 1979). The other predicate [C0021641|Insulin - ASSOCIATED_WITH -

C0149931|Migraine Disorders] came from [PMID65562], a paper by Hsu et al. (Hsu et al. 1977). In our result set for paths of length three this relationship appears in the third answer, shown in Figure 58.

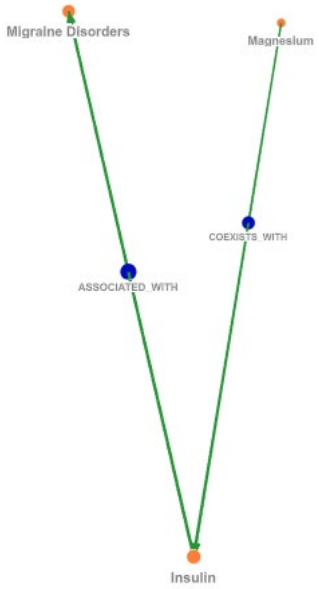


Figure 57. Insulin Subgraph

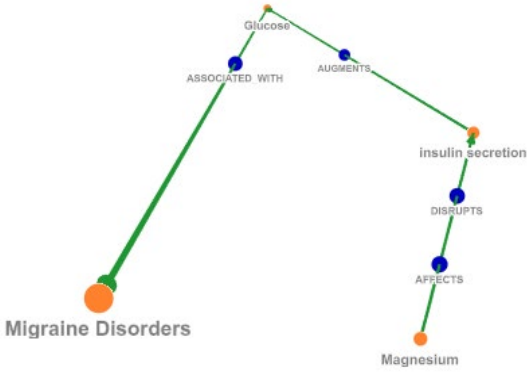


Figure 58. Insulin (Glucose) Subgraph

Performing a search between {*Magnesium, Hypomagnesemia*} and *Migraine Disorders* yields Figure 59. Of primary interest is the predicate [C0151723|Hypomagnesemia - COEXISTS_WITH - C0011849|Diabetes Mellitus] which comes from four documents [PMID7068108, PMID6802656, PMID5682248, PMID527222] where each group of authors found a connection. Fujii et al. [PMID7068108] state that "Hypomagnesemia has been reported to occur in diabetes mellitus in the course of recovery from ketoacidoses, as well as during insulin maintenance therapy" (Fujii, Takemura, Wada, Akai, and Okuda 1982). McNair et al. [PMID6802656] found that "net tubular reabsorption of magnesium is decreased in diabetic patients in presence of

hyperglycaemia, leading to hypermagnesiuria and hypomagnesaemia" (McNair, Christensen, Christiansen, Madsbad, and Transbøl 1982). Jackson and Meier, in their 1968 paper [PMID5682248] "Routine serum magnesium analysis. Correlation with clinical state in 5,100 patients", show that "Diabetes mellitus was the most common condition associated with hypomagnesemia" (Jackson and Meier 1968). If the conditions of Diabetes as a disease make magnesium deficiency prevalent, then it is also likely that migraines are also common for individuals with Diabetes.

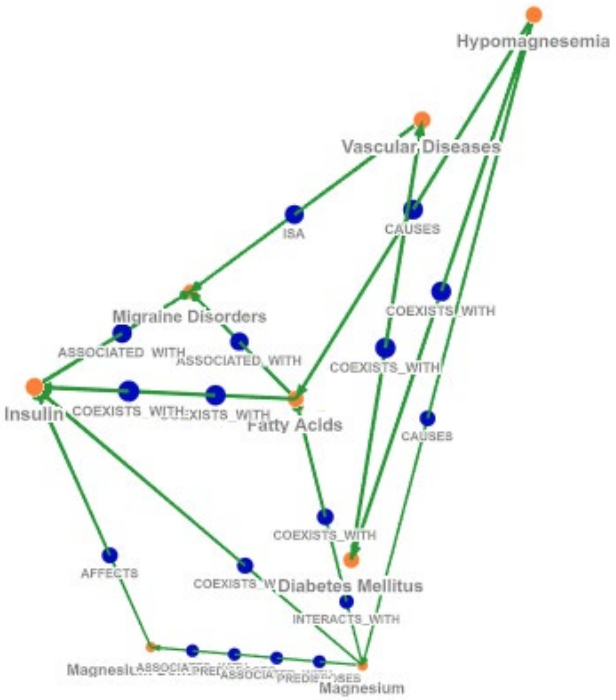


Figure 59. Diabetes Subgraph

In the paper "Increased platelet aggregation in early diabetes mellitus" by Sagel et al. [PMID1138583], they note a "tendency toward vascular disease in diabetes mellitus" and conclude that "platelet aggregation may be increased early in diabetes mellitus" (Sagel, Colwell, Crook, and Laimins 1975). Additionally. Rayssiguier notes in his 1984

paper [PMID6399344] studying the role of magnesium in arteriosclerosis that "magnesium and potassium depletion have also been reported in diabetes and the vascular implications of this should be considered" (Rayssiguier 1984). More recent studies have started to investigate the connections between Diabetes and headaches/migraines [Aamodt, Stovner, Midthjell, Hagen, and Zwart 2007; Berge et al. 2013; Fagherazzi et al. 2019; Haghighi et al. 2015; López-de-Andrés et al. 2018].

It should be noted that the connection [C0042373|Vascular Diseases - COEXISTS_WITH - C0011849|Diabetes Mellitus] which completes the path is only found by including cited documents. This document [PMID1138583] does not have any of the search MeSH terms, and is only included via the citation from a paper by Joseph and Welch [PMID3308769], cited by Swanson which details a study on platelets and migraines (Swanson 1988).

4.2.13 Retrospective Study: Magnesium and Migraine Disorders - Conclusions

The goal of this retrospective study and comparison with Cameron's results is twofold: 1) A further test of reproducibility - that is, in general, similar results can be achieved, and 2) answer the question if the inclusion of citational information improves upon the prior results. The results are summarized in Table 15. For the eleven intermediate concepts, all of the seven that Cameron found were found in our reduced corpus with path lengths less than or equal to three. For the four concepts (*Hypoxia*, *Stress*, *Substance P*, *Spreading Depression*) noticed by Swanson that Cameron did not find, we were able to discover three. Substance *P* was discovered via discovery browsing of related concepts, so it is possible they could have been found within Cameron's system under a different search. Both *Stress/Type A Personality* and *Hypoxia* were found within

our default search. This implies that overall the results were similar and differences are likely due to the differences in datasets due to updates to documents, MeSH terms, and a different version of the SemRep tool. Our results include a twelfth concept, *Diabetes*, undiscovered by Swanson and Cameron, where some predicates could only be uncovered via including cited documents.

Table 15. Comparison of rediscoveries with those of Cameron (Magnesium)

Intermediate	Fleig	Cameron
Calcium Channel Blockers	Found at 2	Found at 2
Prostaglandin	Found at 2	Found at 2
Epilepsy	Found at 2	Found via discovery browsing.
Serotonin	Found at 3	Found at 2
Vascular	Found at 3	Found at 2
Inflammation	Found at 3	Found via discovery browsing.
Hypoxia	Found at 3	Did not find
Platelet Aggregation	Found at 3	Found via discovery browsing?
Stress (Type A personality)	Found at 3.	Did not find
Substance P	Found (at 3) via discovery browsing.	Did not find
Spreading Cortical Depression	Not Found	Did not find
Diabetes	Found at 2	Did not find

In Appendix B, Table B-3 details the precision and recall at each ranked result when using citations. Table B-4 does the same when not using citations but otherwise follows the same algorithm. A graph of the Precision-Recall curve can be seen here in Figure 60, and the F_1@K measure in Figure 61.

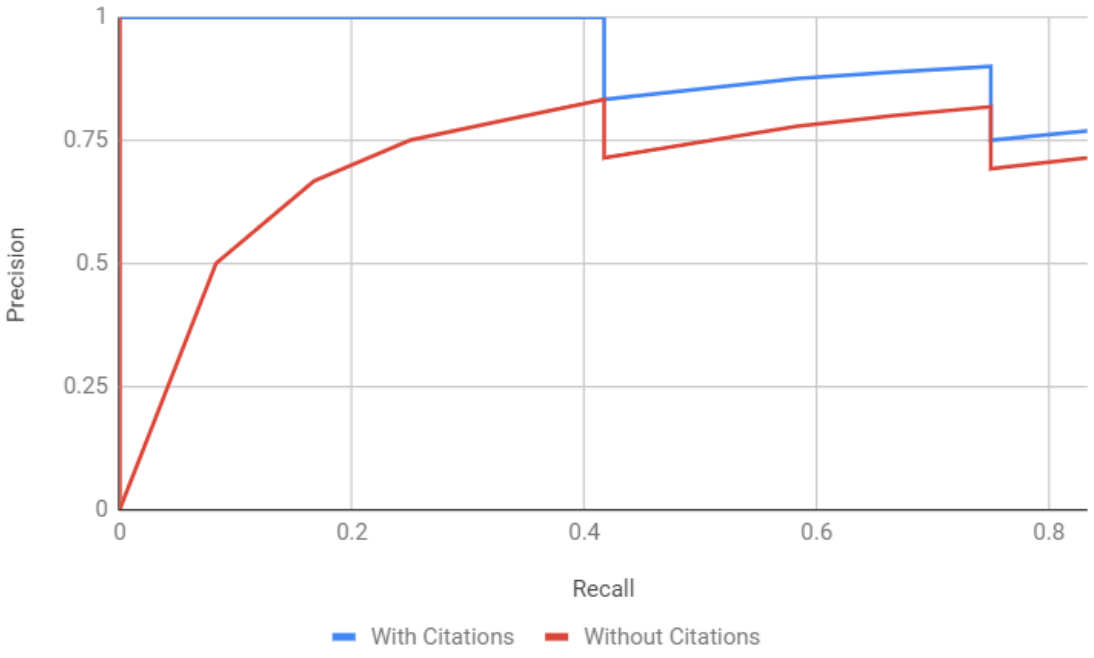


Figure 60. Precision – Recall Curve for Magnesium – Migraine Disorder Study

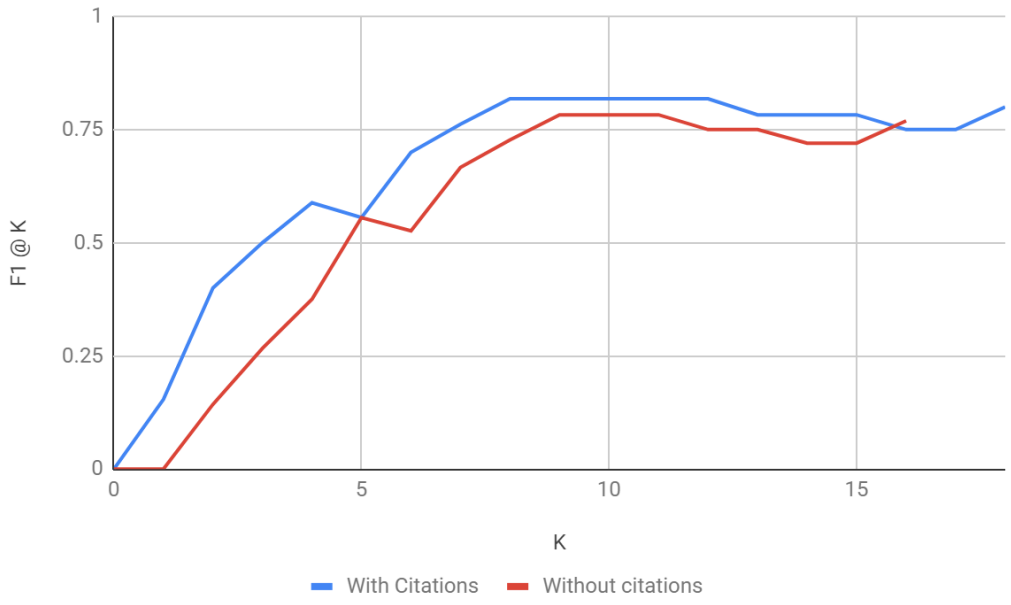


Figure 61. F_1 @ K measure for Magnesium – Migraine Disorder Study

Comparison results show that using citations in the initial search does not uncover one of the expected intermediates that could not be found without their use as there was sufficient justification for several of the terms to account for the discovery. But, the inclusion of citations does cause the relevant results to appear higher in the ordered result set, yield additional results with exploratory searches, and illuminate one new intermediate concept. Analysis of the data indicates this stems from two factors. The first is caused by additional predicates that although yield no new intermediates in this case, may enhance the clustering. The second is caused by a higher IDF threshold.

The IDF threshold is calculated independently for each search, based upon the unique concepts that are returned. Without citations the query result set contains 962 unique UMLS concepts with a mean of 8.11 and a calculated cutoff at 3.945. When using citations the query result set contains 1943 unique UMLS concepts with a mean of 8.49 and calculated cutoff at 4.566. Several of the non-relevant concepts appearing within the non-citation search fall below this higher value. This trend continues for all five of the studies performed (Table 16).

Table 16. Number of unique concepts and average IDF by study

Number of Unique Concepts	Avg IDF	Std Dev	Origin Study
449	7.872367037	2.164025258	Fish Oils (W/o Citations)
962	8.119543301	2.13191342	Magnesium (W/o Citations)
967	8.177050274	2.137530552	Fish Oils (With Citations)
1276	8.242275259	2.123573465	Arginine (W/o Citations)
1686	8.492098715	2.107634478	Testosterone (W/o Citations)
1943	8.489334596	2.011232172	Magnesium (With Citations)
2163	8.508816686	2.072720142	Arginine (With Citations)
2260	8.628326503	2.091449369	Parkinson's (W/o Citations)
2745	8.700485948	2.064198644	Testosterone (With Citations)
3077	8.73736814	2.06278992	Parkinson's (With Citations)

This behavior is to be expected given the nature of language. Zipf's law states that, when given a corpus comprised of natural language, the frequency of any word is inversely proportional to its rank in the frequency table for that corpus. In other words, that the most frequent word will occur about twice as often as the second most frequent word, three times that of the next and so on and so forth (Zipf 1949). This distribution follows that of a power law probability distribution. Previously we stated that IDF values for each unique concept were calculated based upon the 85.7 million predicates generated from Semrep. Plotting the document frequencies for each concept in rank order we find that they behave as expected by Zipf's law (Figure 62).

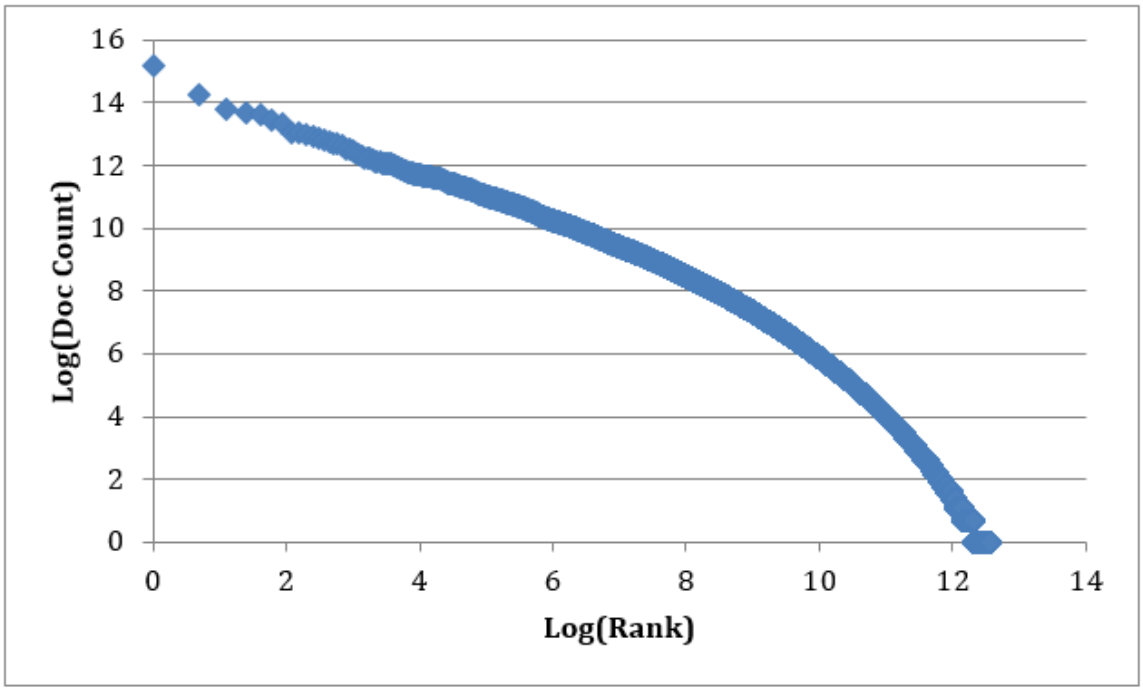


Figure 62. Log – Log Plot of Concepts Illustrating Zipf's Law

The probability density for the Zipf distribution has been defined as Equation 4.1, where ζ(s) is the Riemann zeta function (Eq 4.2), k is the rank position, and s characterizes the distribution.

$$f(k|s) = \frac{k^{-s}}{\zeta(s)} \tag{4.1}$$

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}, \; s > 1 \tag{4.2}$$

The shape of this probability curve is known as a Pareto distribution and can be seen in Figure 63.

Figure 63. Example Pareto Distribution

When considering the set of unique concepts within a document set, such as our search results without using citations, that set will contain some distribution of concepts out of the entire corpus following Zipf's law. That is to say we expect, on average, more concepts with a lower IDF score. By expanding the search to include citationally related documents a greater number of concepts are returned, thereby increasing the average IDF score. Following the probability distribution selecting random sets shows us this behavior and explaining the additional differences between results (Figure 64).

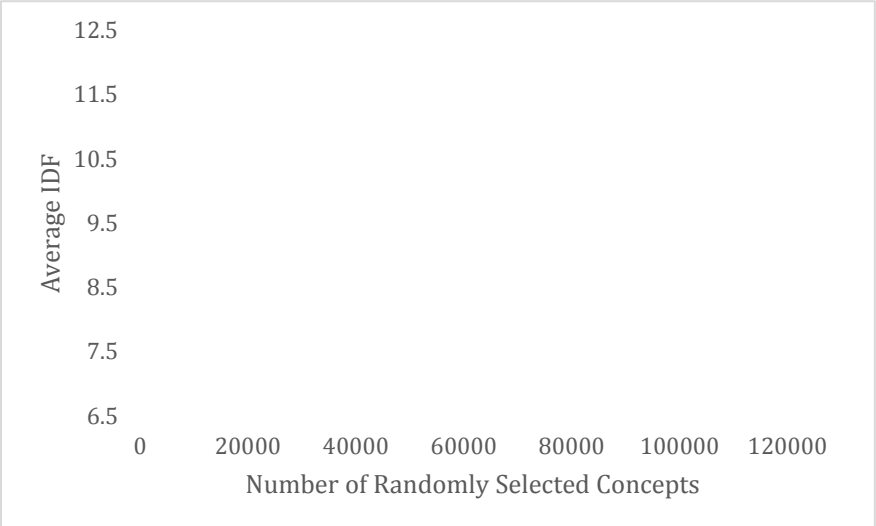


Figure 64. Average IDF for Random Selection of Unique Concepts

4.3 Retrospective Study: Somatomedin and Arginine

In his 1990 paper, "Somatomedin C and Arginine: Implicit Connections between Mutually Isolated Literatures" Swanson described four connections between the protein Somatomedin C, also termed Insulin-Like Growth Factor 1 IGF), and the dietary amino acid Arginine. These were: stimulus of Growth Hormones, the promotion of Wound Healing, recovery from malnutrition, and the general improvement of Body Mass/Weight (Swanson 1990). We attempted to reproduce these results via the approach described previously.

Our corpus consisted of 150 of the original MEDLINE documents cited by Swanson (Table C-1) along with the other MEDLINE documents they cited, totaling 2,125 references. Some documents were missing abstracts (and thusly missing SemRep abstract-generated predicates), for these abstracts/introductions were located and added manually along with any generated predicates. A listing of the pertinent predicates along and their corresponding sentences are recorded by document identifier within Table C-2.

Unless specified otherwise, each search was performed with source concepts of Somatomedins (C0037657) and Insulin-Like Growth Factor I (C0021665), along with a destination of Arginine (C0003765). As an example, Figure 65 shows a complete graph of all paths of length two returned. Of the four intermediates we immediately see three of them, along with a new intermediate, *Diabetes*.

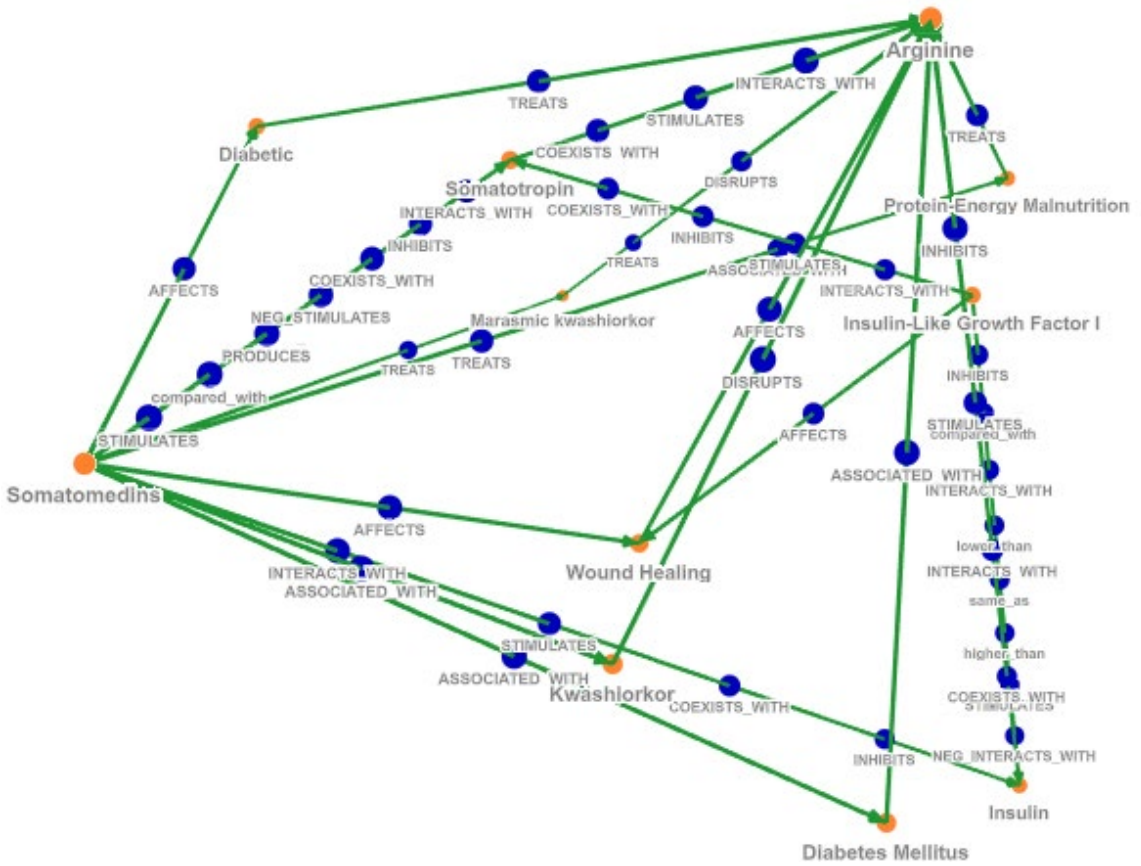


Figure 65. Somatomedin – Arginine paths of length 2

4.3.1 Growth Hormone

The first intermediate, Growth Hormone (GH), is also known as somatotropin or human growth hormone (HGH). It stimulates both growth and the regeneration and reproduction of cells. Arginine, Somatomedin, and IGF all have direct associations to somatotropin as can be seen in Figure 66 where our results with a path length of two are shown.

The article entitled "Effect of arginine on serum levels of insulin and growth hormone in obese subjects" [PMID6027286] by Copinschi, Wegienka, Hane, and Forsham (1967) suggests [C0037663|Somatotropin - INTERACTS_WITH -

C0003765|Arginine] by title alone. Three other articles also provide similar associations [PMID4201417, PMID5571779, PMID3683183].

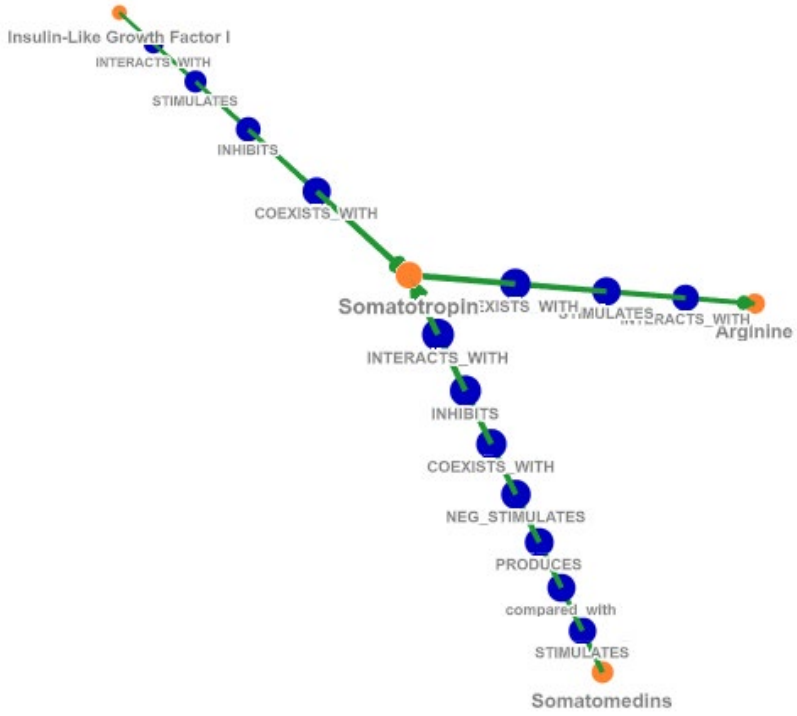


Figure 66. Growth Hormone Subgraph

Multiple sources provide the relationships between Somatotropin and IGF1. Manson, Smith, and Wilmore wrote in their article "Growth hormone stimulates protein synthesis during hypocaloric parenteral nutrition" [PMID2899993] that "GH was associated with an increase in insulin and insulin-like growth factor-I concentrations" (Manson, Smith, and Wilmore 1988). This sentence produces the predicate [C0037663|Somatotropin - STIMULATES - C0021665|Insulin-Like Growth Factor I]. This is similarly echoed in other references: "Levels of circulating IGF I and IGF II are affected by growth hormone" [PMID3322823] (Hammerman 1987), "By contrast, both GH and PL stimulate IGF-I synthesis" [PMID6338399] (Adams, Nissley,

Handwerger, and Rechler 1983), and "There was also a significantly enhanced increase of SM-C/IGF I in the presence of GH" [PMID6385593] (Stracke, Schulz, Moeller, Rossol, and Schatz 1984).

The explicit relationship between Somatomedins to GH/Somatotropin is clear in the article "Specific binding of a somatomedin-like polypeptide in rat serum depends on growth hormone" [PMID967246] by Moses, Nissley, Cohen, and Rechler, where they state that "Somatomedins are growth hormone-dependent polypeptides" (Moses, Nissley, Cohen, and Rechler 1976). In addition to this, in 1980 Hintz et al. wrote in their article "Interaction of somatomedin-C with an antibody directed against the synthetic C-peptide region of insulin-like growth factor-I" [PMID6153391] that "Insulin-like growth factor-I is a human plasma peptide with strong structural homology to human prosinsulin. This peptide has been classified as a somatomedin on the basis of its biological actions and growth hormone dependence" (Hintz, Liu, Marshall, and Chang 1980). Additional references also produce the predicates relating *Somatomedin* and *Somatotropin*: [PMID4506104, PMID3803997, PMID6639868].

4.3.2 Malnutrition

The second intermediate, *Malnutrition*, is discovered through *Kwashiorkor* which is a type of severe form of protein malnutrition and *Protein-Energy Malnutrition*. The subgraph illustrating these relationships is shown in Figure 67.

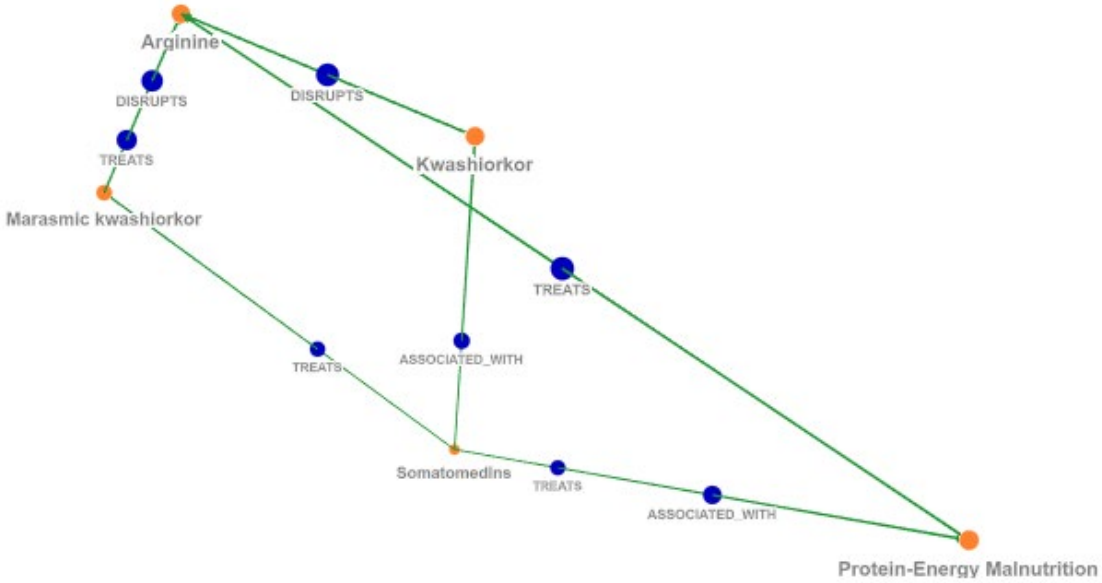


Figure 67. Malnutrition Subgraph

Several predicates relating *Arginine* and *Kwashiorkor, Marasmic Kwashiorkor,* and *Protein-Energy Malnutrition* were generated from the 1986 article entitled "Serum insulin-like growth factors I and II concentrations and growth hormone and insulin responses to arginine infusion in children with protein-energy malnutrition before and after nutritional rehabilitation" by Soliman et al. [PMID3099250]. They stated that "GH responses to arginine were depressed in the three malnourished groups and improved significantly in marasmic-kwashiorkor and marasmic children after nutritional rehabilitation" (Soliman et al. 1986). They then further state that "insulin responses to arginine were impaired in kwashiorkor, and marasmic-kwashiorkor children and improved significantly after refeeding".

In "Effects of dietary composition on somatomedin activity in growing rats" [PMID107298], Reeves et al. studied how diet affects somatomedin activity. The predicate [C0037657|Somatomedins - ASSOCIATED_WITH - C0022806|Kwashiorkor] was

produced by their conclusion that "these data suggest that dietary composition has a direct effect on plasma somatomedin activity and that the severe growth retardation associated with protein malnutrition may be related to its additional effect on serum somatomedin" (Reeves et al. 1979). Three other references produce like predicates: [PMID96201, PMID6792631, PMID467343].

Finally, two other papers [PMID412936, PMID105670] examined Somatomedins in relation to malnutrition. As an example, in the first paper, Hintz et al. stated that their "study was undertaken to define the interrelationships of somatomedin, growth hormone, and an inhibitor of SM in protein-calorie malnutrition" (Hintz, Suskind, Amatayakul, Thanangkul, and Olson 1978).

4.3.3 Wound Healing

Our results show that *Wound Healing* is directly connected to each of *Arginine*, *Somatomedins*, and *Insulin-Like Growth Factor I*. Each of these is illustrated in Figure 68. The predicate [C0003765|Arginine - AFFECTS - C0043240|Wound Healing] is derived from the 1985 article by Barbul et al [PMID3923266] where their conclusions "indicate that high arginine levels in IVH solutions improve wound healing and thymic immune function following injury" (Barbul et al. 1985).

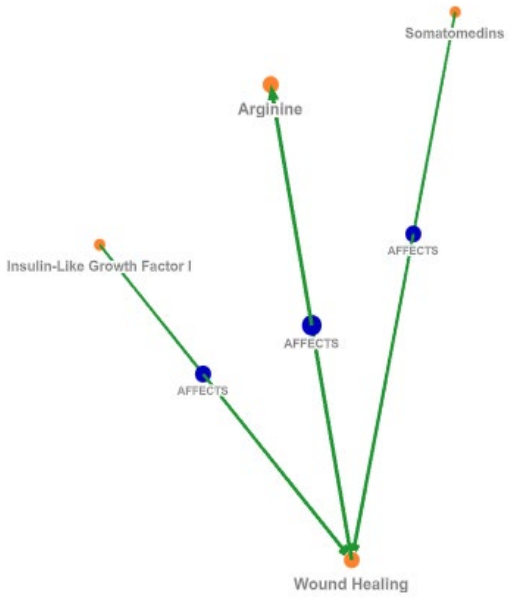


Figure 68. Wound Healing Subgraph

Spencer, Skover, and Hunt investigated the relationship between *Wound Healing* and *Somatomedins* from which the predicate [C0037657|Somatomedins - AFFECTS - C0043240|Wound Healing] is derived. In their 1988 article "Somatomedins: do they play a pivotal role in wound healing?" [PMID3380847] they state that "An understanding of the role of somatomedins in normal wound healing might be able to help us understand abnormalities of the repair process such as keloid formation" (Spencer, Skover, and Hunt 1988).

Although these two predicates are sufficient to prove the intermediate *Wound Healing*, the additional predicate [C0021665|Insulin-Like Growth Factor I - AFFECTS - C0043240|Wound Healing] is also present within the subgraph. The originating article "Regenerating endothelial cells express insulin-like growth factor-I immunoreactivity after arterial injury" [PMID3690631] by Hansson, Jennische, and

Skottner concludes that “IGF-I is likely to be involved in the repair of the intima in injured arteries” (Hansson, Jennische, and Skottner 1987).

4.3.4 Diabetes

In addition to the aforementioned three connections there were a few unexpected intermediate concepts that showed up: these include *Insulin*, *Diabetic*, *and Diabetes Mellitus* (Figure 69). Although these were not called out as intermediates in Swanson’s original paper (Swanson 1990) it should be noted that Cameron et al. also uncovered *Insulin*.

In 1980 Binoux, Hossenlopp, Lassarre, and Seurin studied the production of Somatomedin [PMID6986740]. Their findings suggested that “insulin plays an important role in SM generation, as it itself is capable of both stimulating the release of SM and amplifying the stimulatory effect of GH” (Binoux, Hossenlopp, Lassarre, and Seurin 1980). This conclusion produces the predicate [C0021641|Insulin - INTERACTS_WITH - C0037657|Somatomedins]. This interplay is also backed up by the predicate [C0021641|Insulin - COEXISTS_WITH - C0037657|Somatomedins] from Kogawa et al. [PMID6751804] where “the addition of human growth hormone and/or insulin to the perfusates caused a significant increase in somatomedin after 120 minutes of recirculation” (Kogawa, Takano, Hizuka, Asakawa, and Shizume 1982). Other sources also show relationships between the two concepts: [PMID6368579, PMID1261514, PMID132381]

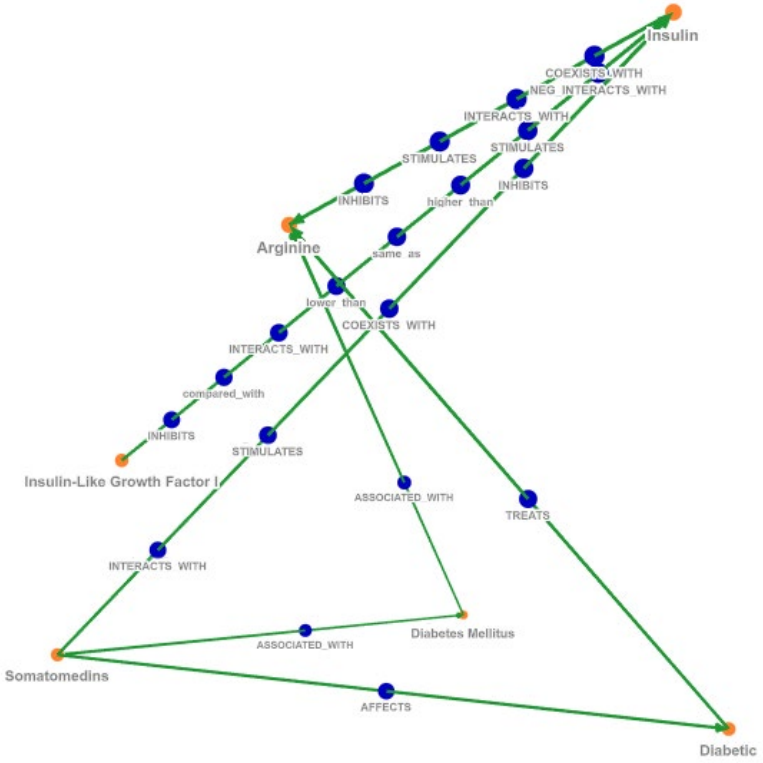


Figure 69. Diabetes Subgraph

The predicate relating *Arginine* and *Insulin*, [`C0003765|Arginine - COEXISTS_WITH - C0021641|Insulin`], comes from "Sex-determined variation in serum insulin and growth hormone response to amino acid stimulation" [`PMID5917038`], a study by Merimee, Burgess, and Rabinowitz. They state that "during the course of experience with subjects who were given infusions of arginine, we observed that female subjects showed a greater rise in serum insulin and in serum HGH than did males of a comparable age" (Merimee, Burgess, and Rabinowitz 1966). Three other studies [PMID6027286, PMID4903729, PMID6754563] also generate predicates linking *Arginine* and *Insulin*.

The relations between either *Somatomedins* or *Arginine* and *Diabetic/Diabetes Mellitus* are generated from a few sources. The observation "of a growth hormone-

resistant decrease in somatomedin activity associated with conditions of insulinopenia suggest that somatomedin and growth might be related to metabolic control in diabetes mellitus" from Phillips and Orawski [PMID142677] produces [C0037657|Somatomedins - ASSOCIATED_WITH - C0011849|Diabetes Mellitus] (Phillips and Orawski 1977). The next, [C0037657|Somatomedins - AFFECTS - C0241863|Diabetic], is likewise generated by the observation in [PMID2413420] that "diabetic children, treated conventionally, have normal circulating IGF levels, but both growth rate and serum IGF concentration may increase dramatically when diabetic control is optimized" (Hill and Milner 1985). Finally, the 1979 study "Prolactin nonresponsiveness to arginine in diabetes mellitus" examined the interaction of Arginine within diabetic patients (Le Roith, Shapiro, Jabotinsky, and Spitz 1979).

4.3.5 Body Weight (Growth)

Performing a closed discovery yields a subgraph result that includes *Growth* (Figure 70). Multiple studies generate predicates associating *Arginine* and *Growth*: "Dietary arginine may play a critical role in growth of normal as well as neoplastic tissue" [PMID430251] (Milner and Stepanovich 1979), "The diet supplemented with arginine and glycine improved growth before and after trauma" [PMID430225] (Pui and Fisher 1979), "The rôle of arginine in growth" [PMID20276173] (Borman et al. 1946), and "Effect of arginine deficiency on growth and intermediary metabolism in rats" [PMID4430939] (Milner, Wakeling, and Visek 1974).

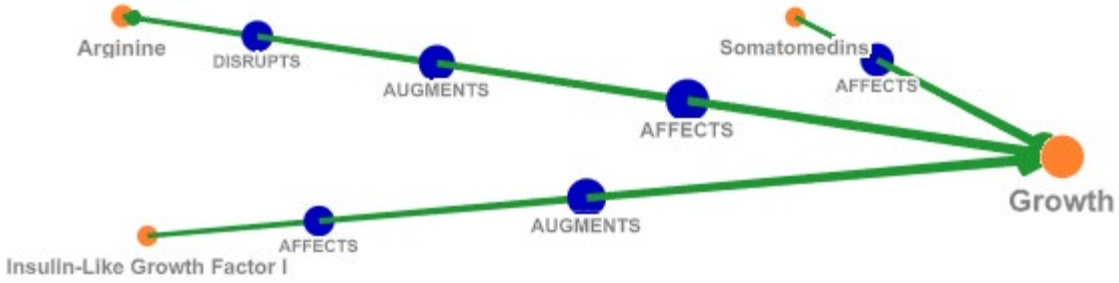


Figure 70. Growth (Body Weight) Subgraph

The inter-relationship between *Growth* and *Somatotropin* has been known for over half a century as evident by the well-known study by Salmon and Daughaday where they hypothesized the existence of a mediating factor (Salmon and Daughaday 1957). Multiple documents generate predicates relating the two. The predicate [C0037663|Somatotropin - AUGMENTS - C0018270|Growth] comes from [PMID6243390] where the authors acknowledge the earlier work: "It is now over 20 years since Salmon and Daughaday hypothesized that stimulation of growth by growth hormone (somatotropin) was mediated by a circulating 'sulfation factor'" (Phillips, Vassilopoulou-Sellin, and Reichard 1979). Similar quotes can be found in other references: "Somatomedin (SM) is a growth hormone (GH)-dependent peptide, circulating in normal serum, which directly causes skeletal growth" [PMID1128255] (Daughaday, Phillips, and Herington 1975), "The significance of pituitary growth hormone (GH) for the regulation of skeletal growth is uniformly recognized" [PMID4538721] (Hall 1972), and "In the whole animal a major function of GH is to stimulate the longitudinal growth of the skeleton, and GH acts by controlling the

production of a second series of hormones, the somatomedins" [PMID1128683] (Francis and Hill 1975).

The other predicates shown in Figure 70 have already been discussed in our examination of previous intermediates. These include the association of *Somatropin* to both *Somatomedin* and *Insulin-Like Growth Factor I*.

4.3.6 Retrospective Study: Somatomedin and Arginine - Conclusions

The goal of this retrospective study is the same as our previous ones and the results are summarized in Table 17. We were able to successfully uncover all five of the expected intermediate concepts. In comparison, Cameron's study was not able to find *Wound Healing*.

Table 17. Comparison of rediscoveries with those of Cameron (Somatomedins)

Intermediate	Fleig	Cameron
Growth Hormone	Found at 2	Found at 2
Body Weight	Found at 2	Found at 2
Malnutrition	Found at 2	Found via discovery browsing.
Wound Healing	Found at 2	Not Found
Diabetes	Found at 2	Found at 2

In Appendix C, Table C-3 details the precision and recall at each ranked result when using citations. Table C-4 does the same when not using citations but otherwise follows the same algorithm. A graph of the Precision-Recall can be seen here in Figure 71.

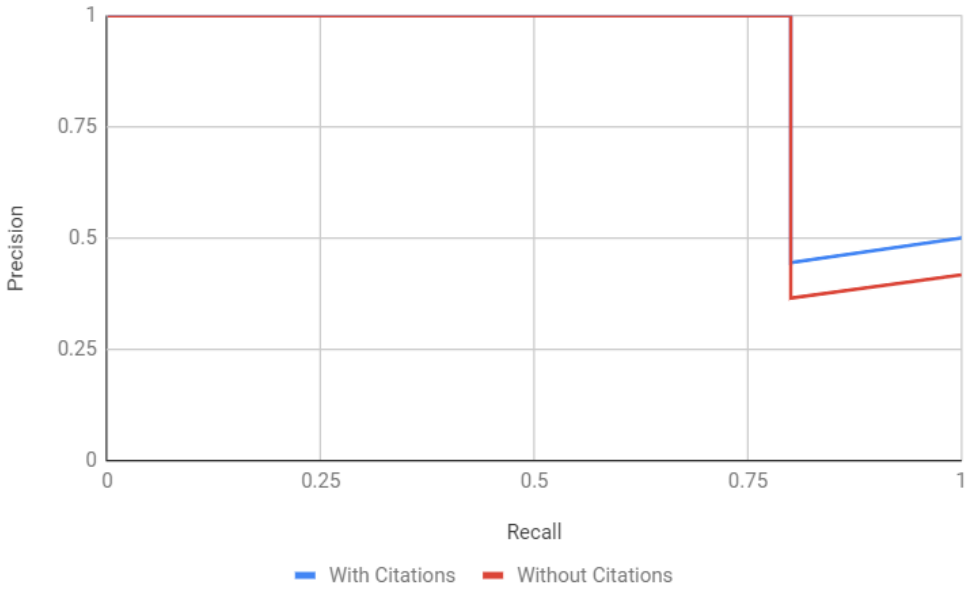


Figure 71. Precision – Recall Curve for Somatomedins – Arginine Study

As we were able to uncover all of the expected intermediates with just a path length of two, the slight improvement in our Precision-Recall curve when using citations is purely due to the higher IDF threshold which we discussed previously. Figure 72 shows the F1@K graphs for each approach and their expected similarity.

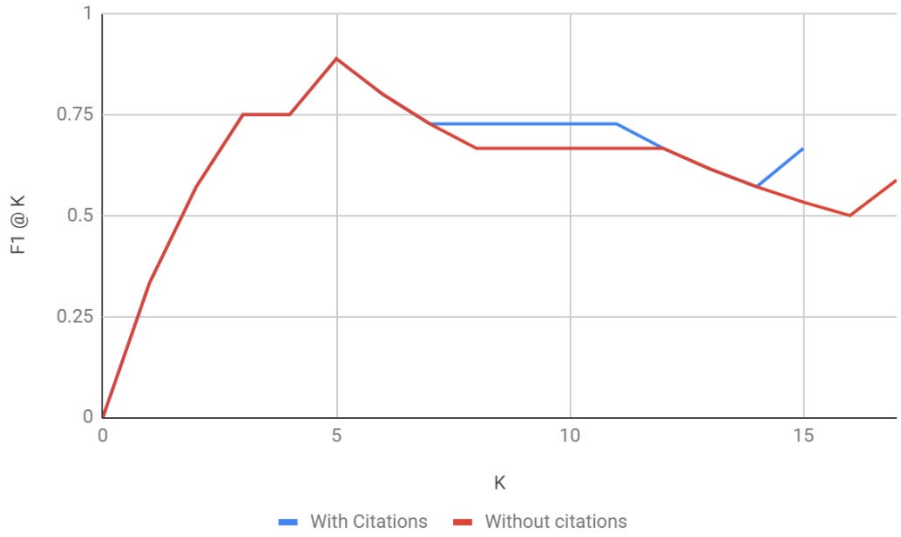


Figure 72. F_1 @ K measure for Somatomedins – Arginine Study

4.4 Retrospective Study: Testosterone and Sleep

In their 2012 paper, "A Closed Literature-Based Discovery Technique Finds a Mechanistic Link Between Hypogonadism and Diminished Sleep Quality in Aging Men" Miller et al. described their insights into the associations between *Hyogonadism* and diminishing sleep quality in men. It had been known that the process of aging in healthy adults causes a reduction in sleep quality due to decreased REM sleep, more frequent awakenings, and total sleep duration (Espiritu 2008). It had also been previously suggested that hormonal changes could be responsible for a shorter sleep duration (Trenell, Marshall, and Rogers 2007), though no mechanism was known.

Miller et al. stated that a consequence of hypoganism is high levels of Cortisol, and therefore theorized that as men age and testosterone levels decrease their cortisol levels increase due to the onset of hypogonadism. (Miller et al. 2012). As such we expect to find the intermediate *Cortisol*/*Hydrocortisone*. We attempted to reproduce these results via the approach described previously.

Our corpus consisted of 52 of the original MEDLINE documents cited by Miller et al. (Table D-1) along with the other MEDLINE documents they cited, totaling 2,039 references. A listing of the pertinent predicates along and their corresponding sentences are recorded by document identifier within Table D-2. Our search was performed with the source concept *Testosterone* (C0039601) and the destination of *Sleep* (C0037313, C0037322, C0234451), finding results at path lengths of one and two.

Multiple papers call out the direct association between *Testosterone* and *Sleep* directly. Andersen and Tufik's 2008 paper [PMID18519168] states it directly in their title "The effects of testosterone on sleep and sleep-disordered breathing in men: its bidirectional interaction with erectile function" (Andersen and Tufik 2008). In the 1997

paper [PMID9329339] by Leibenluft et al. they claim that their results indicated that "testosterone has relatively specific and discrete effects on sleep and hormonal rhythms in men" (Leibenluft et al. 1997). Lastly Axelsson et al. wrote in their 2005 paper [PMID15914523] that "testosterone increased during sleep and fell during waking, whereas circadian effects seemed marginal" (Axelsson, Ingre, Åkerstedt, and Holmbäck 2005). Our first search result illustrates this (Figure 73).

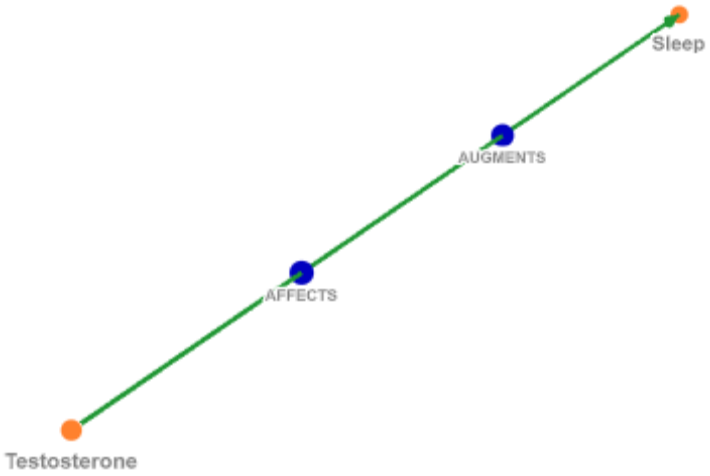


Figure 73. Direction relationship between Testosterone and Sleep

Several results also appear which continue to emphasize that a relationship exists, but does not provide a deeper semantic meaning. These include *Gonadal Steroid Hormones* (Figure D-3), *Hormones* (Figure D-5), *Aging* (Figure D-7), *Growth Hormone Secretion* (Figure D-8), *Circadian Rhythms* (Figure D-9), and *Estrogens* (Figure D-13).

4.4.1 Hydrocortisone (Cortisol)

Hydrocortisone does show up in our ninth search result (Figure 74), the association having been generated from multiple source references. *Cortisol* is a

glucocorticoid secreted by the adrenal cortex in response the physical and psychological stress. Glucocorticoids are involved in the metabolism of carbohydrates, proteins, and fats and have anti-inflammatory properties. The release of *Cortisol* affects the metabolism by attempting to help maintain blood glucose levels during physical exercise. *Testosterone* is an anabolic hormone important in the growth and maintenances of muscle, bone, and red blood cells. Similar to *Cortisol*, *Testosterone* levels increase in response to exercise and stress.

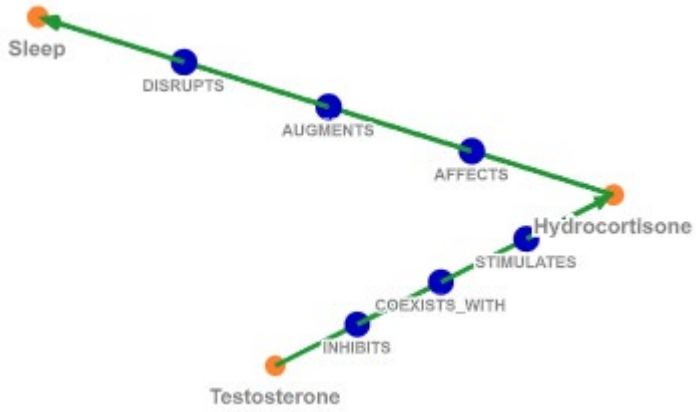


Figure 74. Relationship between Testosterone and Sleep via Hydrocortisone

Previous research has established that under particular circumstances *Cortisol* and *Testosterone* exhibit a negative relationship. In the paper [PMID15724043] "Fetal plasma testosterone correlates positively with cortisol" Gitau et al. noted that "unlike the norm in the adult, where testosterone production is often inhibited by cortisol, in the fetus there is a positive link between the two" (Gitau, Adams, Fisk, and Glover 2005). Doerr and Pirke concluded that the "administration of cortisol leading to plasma levels as seen under treatment with ACTH suppresses testosterone by abolishing or flattening the nocturnal

rise" in their paper [PMID956348] "Cortisol-induced suppression of plasma testosterone in normal adult males" (Doerr and Pirke 1976). Two additional papers also discuss the relationship: [PMID15886244, PMID15841103].

The paper [PMID10841212] "Sleep impairments in healthy seniors: roles of stress, cortisol, and interleukin-1 beta" by Prinz, Bailey, and Woods noted that "healthy older women and men with higher levels of free cortisol under a mild stress condition had impaired sleep" (Prinz, Bailey, and Woods 2000). The researchers Kern, Dodt, Born, and Fehm also corroborates this [PMID8548511] in the results where "changes in sleep-dependent secretion of GH and cortisol correlated significantly with an age-dependent decrease in slow wave sleep" (Kern, Dodt, Born, and Fehm 1996). Five other papers also discuss the relationship between *Cortisol* and *Sleep*: [PMID3790626, PMID3661052, PMID2541159, PMID8077308, PMID6822642].

4.4.2 Alzheimer's Disease

One of our results indicated a potential connection via the intermediate *Alzheimer's Disease* (Figure 75). Upon inspection this path exists due to two cited papers included within the corpus. The first [PMID16344336], entitled "Effects of testosterone on cognition and mood in male patients with mild Alzheimer disease and healthy elderly men", by Lu et al. concluded that "testosterone replacement therapy improved overall quality of life in patients with AD" but had "minimal effects on cognition" (Lu et al. 2006).

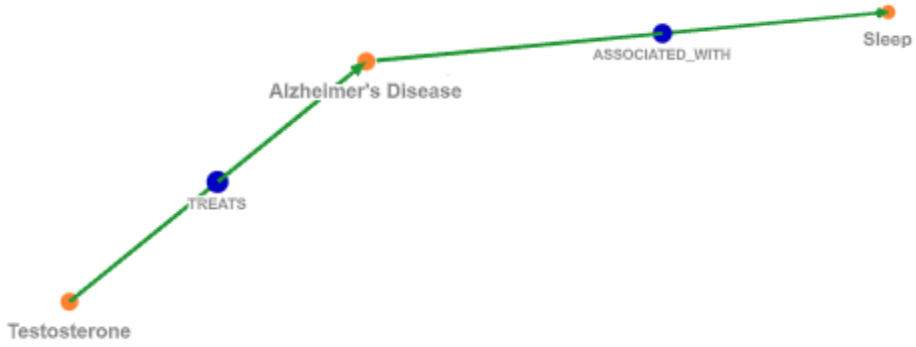


Figure 75. Relationship between Testosterone and Sleep via Alzheimer's disease

The second was a paper [PMID12531038] by McCurry et al. titled "Treatment of sleep disturbance in Alzheimer's disease" only suggested that "sleep problems in AD are multi-factorial, and influenced by a variety of demographic, physical, psychiatric and situational factors" (McCurry, Reynolds, Ancoli-Israel, Teri, and Vitiello 2000). The predicates in question were generated from very general relationships and thus are inconclusive, but does leave open the possibility that sleep problems in some Alzheimer's patients could be due to decreased Testosterone.

4.4.3 Interleukin-1 Beta

Interleukin-1 beta is a cytokine protein and is a pro-inflammatory cytokine involved in the body's inflammatory response to infections or conditions associated with an inflammatory state. From [PMID15240608], *Testosterone* has been shown to have "immune-modulating properties" and evidence suggests that it "may suppress the expression of the pro-inflammatory cytokines" (Malkin et al. 2004). Previously we noted the relationship between reduced testosterone and greater amounts of cortisol, a known glucocorticoid. This is further evidenced by a paper by Chang and Opp [PMID

10956236] entitled “IL-1 is a mediator of increases in slow-wave sleep induced by CRH receptor blockade”. Therein they describe that “IL-1 promotes sleep, and glucocorticoids inhibit IL-1 synthesis” (Chang and Opp 2000).

Figure 76. Relationship between Testosterone and Sleep via IL-1 beta

Several other papers provide support for the role of IL-1 in sleep. In [PMID 9781819] Taishi et al. state that “much evidence implicates interleukin-1beta (IL-1beta) in sleep regulation” (Taishi et al. 1998). In the same year Fang, Wang, and Krueger [PMID9530230] wrote that “Interleukin-1 beta (IL-1 beta) is a well characterized sleep regulatory substance” (Fang, Wang, and Krueger 1998). A later study [PMID12000022]. in 2001 by some of the same individuals continued to support the earlier work (Krueger, Obál, Fang, Kubota, and Taishi 2001).

The original paper by Miller et al. did not state IL-1 (or any other cytokine) as a potential intermediate. However, given the papers just mentioned, it seems likely that these proteins could be potentially involved within the biochemical interactions involving sleep and the role of testosterone (Figure 76).

4.4.4 Parkinson’s Disease

Within our results, one answer showed *Parkinson’s Disease* as an intermediate (Figure D-6). This path is due entirely to ten papers which studied the effects of

Parkinson's Disease with either *Testosterone* or *Sleep*. These are documents included in our local corpus that were cited from core papers in our final study involving the connections between Parkinson's Disease and Crohn's Disease. It should be noted that, as such, we would not have this result from just the corpus constructed from Miller et al.'s work. The paper by Okun et al. [PMID16682542] indicates that "testosterone deficiency has been reported in patients with Parkinson disease (PD), Alzheimer disease, and Huntington disease" (Okun et al. 2006). And numerous papers illustrate the relationship between sleep disorders and PD: [PMID2259351, PMID18591114, PMID17942122, PMID15389999, PMID12621636, PMID12438461, PMID11798367, PMID10752568].

4.4.5 Retrospective Study: Testosterone and Sleep - Conclusions

We were able to successfully uncover the single expected intermediate concept, *Cortisol/Hydrocortisone*. As we were able to recover the concept with a path length of two, as expected, Cameron's study was also able to recover the concept. Additionally, because we found the solutions at path lengths of two and less we expect there to be little difference between the citational and non-citational approaches. In fact, the resulting subgraphs were identical between the two trials, yielding the same P-R (Figure 77) and F_1@K (Figure 78) curves.

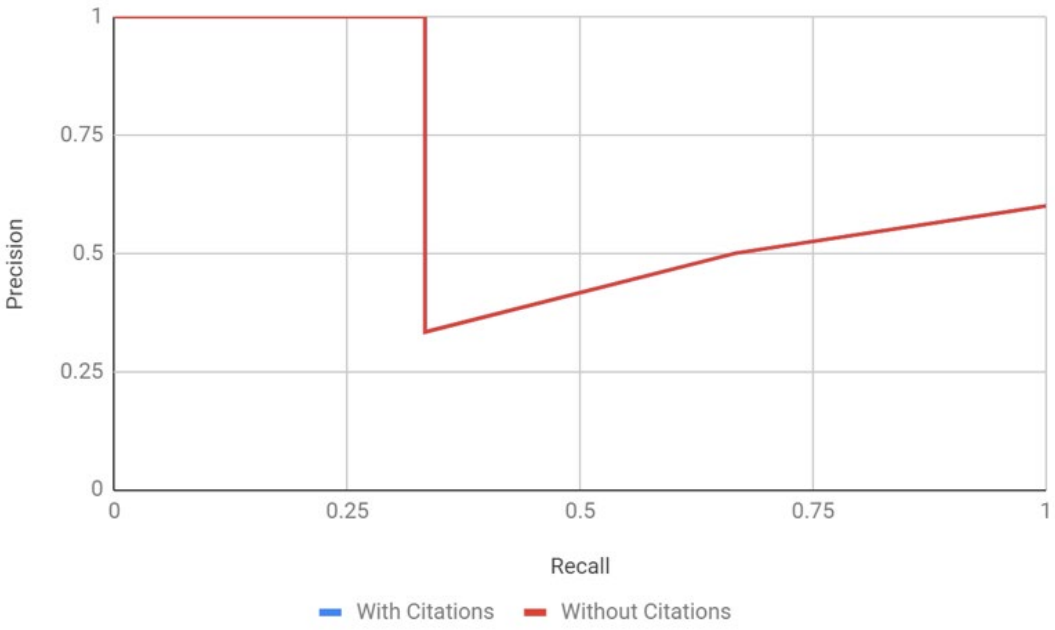


Figure 77. Precision – Recall Curve for Testosterone – Sleep Study

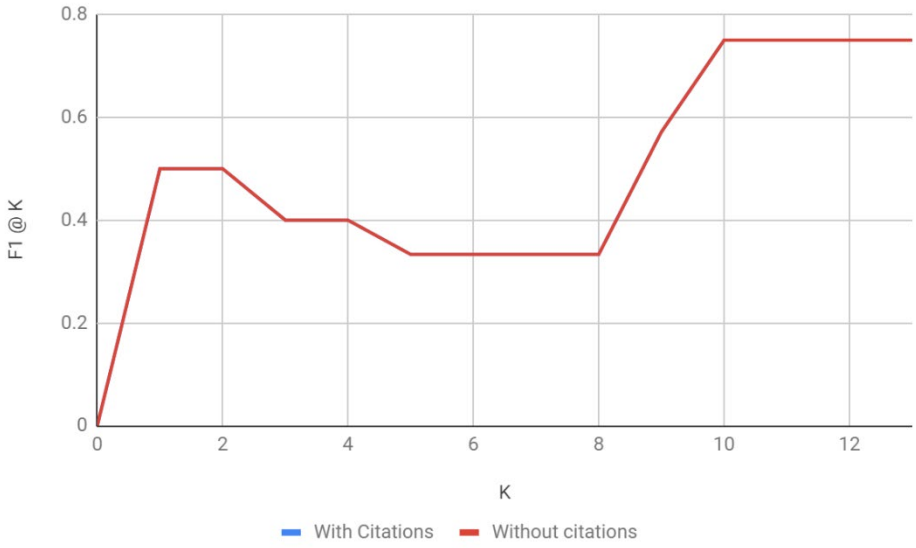


Figure 78. F_1 @ K measure for Testosterone – Sleep Study

4.5 Retrospective Study: Parkinson's Disease and Crohn's Disease

Parkinson's disease (PD) is a neurodegenerative disorder that affects dopamine-producing neurons within the brain. In turn, dopamine activates the nerve cells that coordinate normal muscle activity. In patients with PD, there is a reduction in the supply of dopamine which impairs their ability to control movement.

Crohn's disease (CD) is a type of chronic inflammatory bowel disease (IBD) that affects the lining of the digestive tract. CD is also considered an autoimmune disease as the immune system attacks the gastrointestinal tract causing abdominal pain, diarrhea, vomiting, and other complications.

In 2014 Kostoff published a study titled "Literature-related discovery: common factors for Parkinson's Disease and Crohn's Disease" where in the interconnections between these two diseases were explored via LBD. Kostoff concluded that there were three major themes that unified the PD and CD literatures: these were Genetics, Neuroimmunology, and Cell Death. Additionally, he noted that these three themes are not independent from each other. In one example Kostoff states "there are genetic determinants of the inflammatory response" which are immunological in nature, and that "naturally occurring genetic variants in important inflammatory mediators such as TNF-alpha appear to alter inflammatory responses" (Kostoff 2014).

Reproducing Kostoff's study, our corpus consisted of 32 of the original MEDLINE documents cited by Miller et al. (Table E-1) along with the other MEDLINE documents they cited, totaling 1,974 references. A listing of the pertinent predicates along and their corresponding sentences are recorded by document identifier within Table E-2. Our search was performed with the source concept *Parkinson Disease* (C0030567)

and the destination of *Crohn's Disease* (C0010346), finding results at path lengths of two and three.

4.5.1 Immunology

Inflammation is a normal response for the body's immune system when tissues are injured by trauma, toxins, bacteria, or other causes. Chemicals such as histamine are released by the damaged cells which lead into the surrounding area causing swelling. This is a signal to the body's white blood cells to help protect the area from foreign substances.

Performing a closed search with a max path length of two we find that four of the eight results match this case: *Inflammation* (Figure E-2), *Chronic Inflammation* (Figure E-5), *Immune Response* (Figure E-6), and *Anti-Inflammatory Agents* (Figure E-8). Past research has well documented this generalized connection. Whitton's paper "Inflammation as a causative factor in the aetiology of Parkinson's disease" [PMID17339843] considers the "possible use of anti-inflammatory drugs in PD" (Whitton 2007). Additionally, "accumulating evidence has suggested that inflammation in the brain participates in the pathogenesis of Parkinson's disease" (Li, Wang, Pei, Liu, and Hong 2005). Similar statements connect PD to *Inflammation* in [PMID19221310, PMID15869932, PMID15109580, and PMID17156147].

Crohn's disease is, by definition, an inflammatory bowel disease "characterized by transmural inflammation" (Vallance et al. 2005) and Burke et al. noted [PMID17156147] that the "pathogenesis of inflammation in CD has been extensively investigated" (Burke et al. 2007). Additional statements that connect *Inflammation* to CD

can be found in [PMID12474223, PMID15058528, PMID11052175, and PMID11743591].

Another result (Figure E-7), *Cytokine*, connects *PD* to *CD* and yields more direct information. Cytokines are a broad category of proteins important to cell signaling. More directly for our study, they can be involved with immune responses and inflammation. Wahner et al. noted in 2007 [PMID17984451] that "markers of neuroinflammation, including activated microglia and increased levels of circulating proinflammatory cytokines, have been observed in the brains and CSF of patients with Parkinson disease" (Wahner, Bronstein, Bordelon, & Ritz 2007). Mogi et al. also noticed the connection [PMID10400088]: "the levels of proinflammatory cytokines such as tumor necrosis factor (TNF)-alpha are increased in the striatum and cerebrospinal fluid from patients with Parkinson's disease (PD)" (Mogi et al. 1999). Similar observations were made in [PMID9550432] connecting cytokines in the pathogenesis of Crohn's Disease.

Performing a closed search with a path length of three, and limiting paths that go through one of these intermediates (Figure E-10), we can see additional details to these associations. The first result of the search (Figure 79) shows a number of new potential intermediate terms of interest: *Peroxisome Proliferator-Activated Receptors* (PPAR), *Inflammatory Bowel Diseases* (IBD), *Interleukin-1 beta*, *Ulcerative Colitis* (UC), *Tumor Necrosis Factor-alpha* (TNF alpha), and the *TNF gene*.

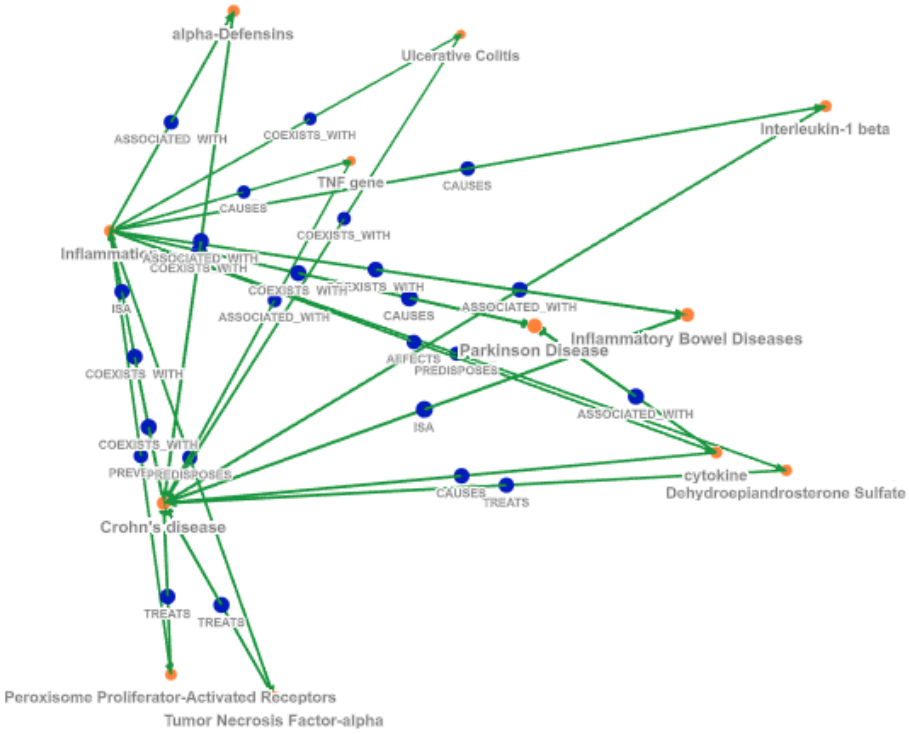


Figure 79. Relationship between PD and CD (Inflammation)

Although these concepts seem like a good fit, and many are associated with *Inflammation* and *Crohn's Disease*, none are directly linked to *Parkinson's disease* in the graph. Thankfully, however, the subsequent results provide more detail.

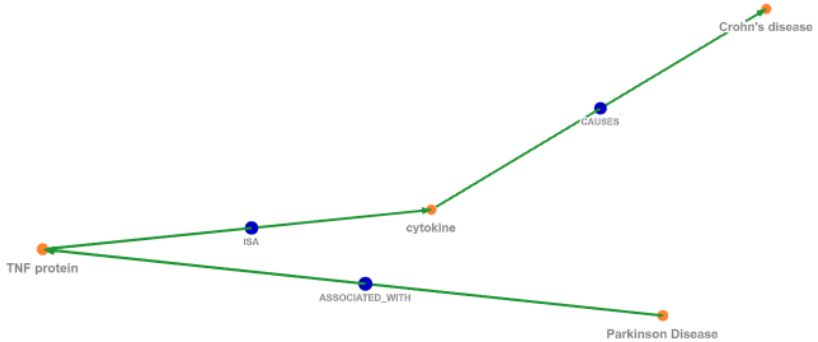


Figure 80. Relationship between PD and CD – TNF protein

While *Tumor Necrosis Factor-alpha* and *TNF gene* both appeared in Figure 79, we find *TNF* protein (Figure 80) in the second search result. The predicate [`C1448177|TNF protein - ASSOCIATED_WITH - C0030567|Parkinson Disease`] comes from [PMID12205053], a paper by Sriram. Therein they studied the potential role for TNF-alpha in PD, stating that "the proinflammatory cytokine TNF-alpha is an obligatory component of dopaminergic neurodegeneration" and that "enhanced expression of the proinflammatory cytokine, tumor necrosis factor (TNF)-alpha, has been found in association with glial cells in the substantia nigra of patients with PD" (Sriram 2002). Similar predicates come from other documents [PMID14724828, PMID17483119].

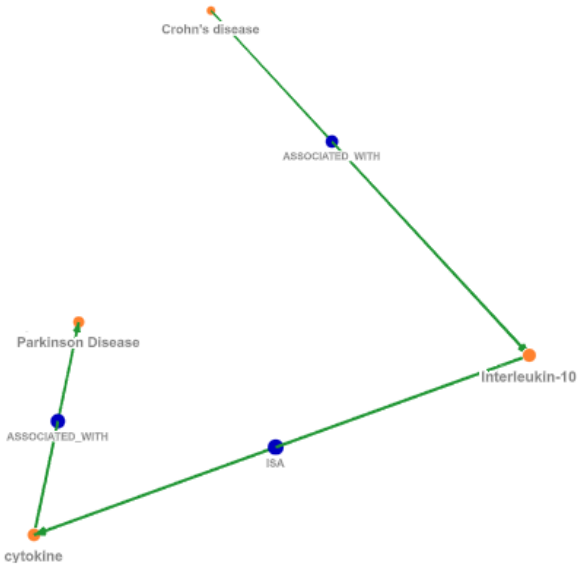


Figure 81. Relationship between PD and CD – IL-10

Another related result from the search (Figure 81) shows the intermediate *Interleukin-10*. The predicate [`Interleukin-10 - ISA - cytokine`] was extracted from a 2005 paper [PMID15879017] (Chevrier et al. 2005). The subsequent association

with *Crohn's Disease* stems from [PMID16461743] where they conclude "the NOD2fs mutation results in a loss-of-function phenotype in human myeloid DC and imply decreased immune regulation by IL-10 as a possible mechanism for this mutation in CD" (Kramer, Netea, De Jong, Kullberg, and Adema 2006). Of note, their paper associates this immune regulating cytokine with *Crohn's Disease*, but it is also within the context of a genetic variation.

4.5.2 Cell Death

The last two results of the *Inflammation* search (Figure E-19 and Figure E-21) show a relation with the concept of *Cell Death* via *Immunology* – specifically *Cytokine* or *Inflammation*. The first of the two, shown here as Figure 82, indicated that *Parkinson's Disease* is related to *Apoptosis*. Apoptosis is a regulated form of pre-programmed cell death which can occur either because it senses some amount of stress (internal) or because it of external signals from other cells.

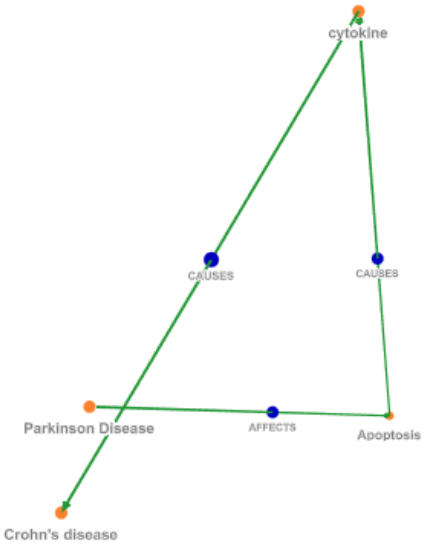


Figure 82. Relationship between PD and CD – Apoptosis

In 2005, Maraganore et al. performed a genetic study of *Parkinson's Disease*, in their paper [PMID16252231] "High-resolution whole-genome association study of Parkinson disease", where we extracted [`C0030567|Parkinson Disease - AFFECTS - C0162638|Apoptosis`]. They identified eleven single nucleotide polymorphisms (variation in a single nucleotide that occurs at a specific position) associated with PD and discovered that "the protein encoded by this gene plays an important role in neurogenesis and in neuronal apoptosis, which is consistent with existing hypotheses regarding PD pathogenesis" (Maraganore et al. 2005). *TNF-alpha* is a cytokine known to be a large mediator of apoptosis, and three additional papers in our corpus then further relate *Cytokine* to *Apoptosis*: [PMID16493072, PMID11922776, PMID9422513].

Figure 83. Relationship between PD and CD – Cell Death (Inflam)

Figure 83 shows Cell Death explicitly as a connection. This is derived from two references. The first, [`C0030567|Parkinson Disease - AFFECTS - C0007587|Cell`

Death] is from a paper [PMID17908040] by Tweedie, Sambamurti, and Greig who detail that "TNF-alpha has been demonstrated to play a major role in central nervous system (CNS) neuroinflammation-mediated cell death in AD, PD and amyotrophic lateral sclerosis (ALS) as well as several other CNS complications" (Tweedie, Sambamurti, and Greig 2007). The same predicate is also extracted from a second paper [PMID8196673].

The predicates connecting *Cell Death* to *Inflammation*/*Anti-inflammatory Agents* is supported by two studies [PMID15109580, PMID17984451]. One of which indicates that "antiinflammatory agents inhibit dopaminergic cell death in animal models of PD, and there is one epidemiological report that their use significantly diminishes the risk of PD" (McGeer and McGeer 2004).

Further evidence of Cell Death being a common factor lies hidden in one of results from the original search (Figure E-3), shown in Figure 84 below. When the connections are examined, the listed concept Anti-Anxiety Agents (C0040616) is found to be a misnomer. The predicate [C0040616|Anti-Anxiety Agents – TREATS – C0030567|Parkinson Disease] is extracted from Olanow's 2006 paper [PMID16717254], stemming from the sentence "this raises the possibility that anti-apoptotic agents might be neuroprotective in PD" (Olanow 2006).

Figure 84. Relationship between PD and CD – Cell Death (apoptosis)

A similar problem occurred in two documents making the connection with Crohn's Disease: [`C0040616|Anti-Anxiety Agents - TREATS - C0010346|Crohn's disease`]. Hanauer summarized [PMID15580149] that "the past decade has brought forth a series of novel biologic agents targeting tumor necrosis factor (TNF) for the treatment of Crohn's disease" and his review discussed "the results of controlled clinical trials of anti-TNF agents for Crohn's disease" (Hanauer 2004). In this case, "anti-TNF" was interpreted by SemRep as Anti-Anxiety Agents by mistake, and we've already noted how TNF is associated with apoptosis. The second article, [PMID17470824] by Sandborn also encountered the same issue with SemRep where "anti-TNF" was picked up as Anti-Anxiety Agents (Sandborn 2007).

With a better understanding of the connections found in Figure 84, this shows a more direct connection between Parkinson's Disease and Crohn's Disease via the higher order concept Cell Death.

4.5.3 Genetics

Previously we referenced a paper by Maraganore et al. [PMID16252231] in regards to the intermediate *Apoptosis*, wherein the authors stated they "identified 11 *Single Nucleotide Polymorphism*s that were associated with *Parkinson's Disease*" (Maraganore et al. 2005), generating the predicate [`C0752046|Single Nucleotide Polymorphism - ASSOCIATED_WITH - C0030567|Parkinson Disease`]. This same concept also appeared as one of the original search results (Figure 85). A large scale international study [PMID17052658] the next year by Elbaz et al. further corroborated this by identifying "13 single-nucleotide polymorphisms (SNPs) significantly associated with Parkinson's disease" (Elbaz et al. 2006).

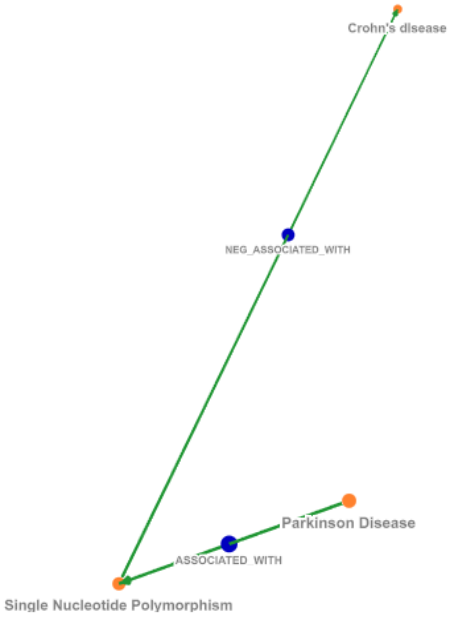


Figure 85. Relationship between PD and CD – Single Nucleotide Polymorphism

The predicate [`C0752046|Single Nucleotide Polymorphism - NEG_ASSOCIATED_WITH - C0010346|Crohn's disease`] which connects Crohn's Disease is extracted from [PMID15685540] where the authors wrote that "the G2677T

SNP was not associated with UC or CD" (Ho et al. 2005). This, unfortunately, is a weak association. Like before, we can dig for additional meaning by attempting a deeper search. A different search performed (Figure E-23) focused around Single Nucleotide Polymorphism between PD and CD, we find several genetic intermediates (Figure 86).

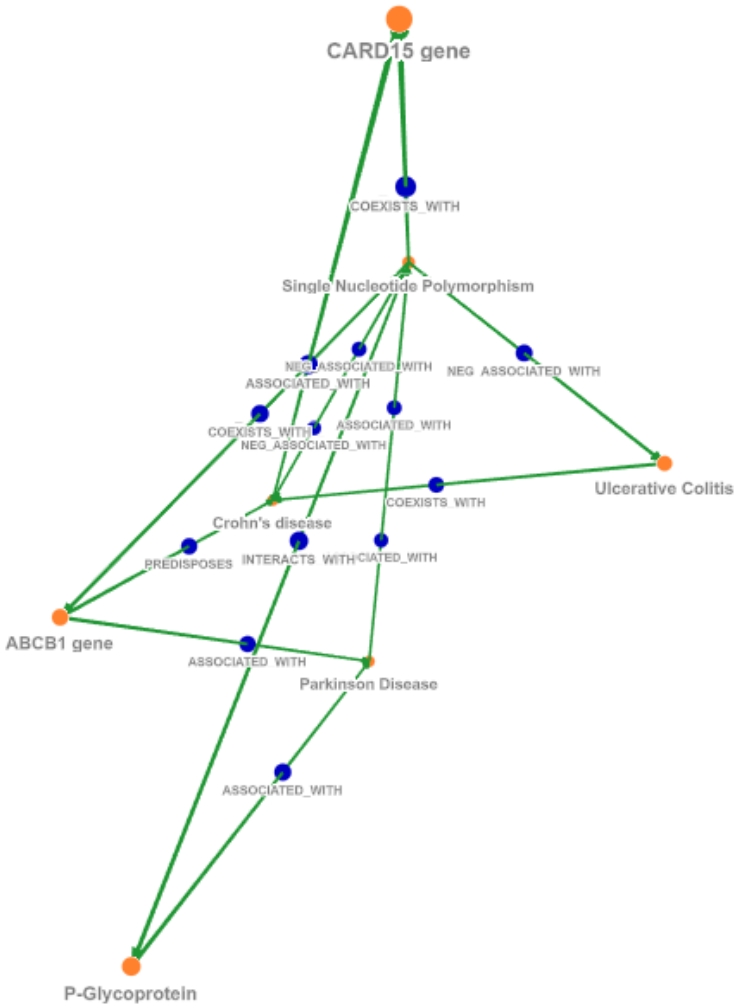


Figure 86. Relationship between PD and CD, PL3 – Single Nucleotide Polymorphism

From the figure we find that *ABCB1 gene* connects to all three: *Parkinson's Disease*, *Crohn's Disease,* and *Single Nucleotide Polymorphism*. This is not a surprise as the last result of our original search connected PD and CD via *ABCB1 gene* (Figure E-9). Additionally *P-Glycoprotein*, also known as Multidrug Resistance Protein 1 (MDR1), is

encoded by the *ABCB1 gene*. Ishikawa et al. reported in [PMID15256718] that "genetic variations of the human ABCB1 (P-glycoprotein/MDR1) gene have been most extensively studied" and "more than fifty single nucleotide polymorphisms (SNPs) and insertion/deletion polymorphisms in the ABCB1 gene have been reported" (Ishikawa et al. 2004).

Other research has further connected the gene to each of the two diseases. In their paper "Polymorphisms in human MDR1 (P-glycoprotein): recent advances and clinical relevance" [PMID14749689] Marzolini, Paus, Buclin, and Kim stated "An increasing number of studies have also implicated certain commonly occurring SNPs in MDR1 in problems including altered drug levels and host susceptibility to diseases such as Parkinson's disease, inflammatory bowel disease, refractory seizures, and CD4 cell recovery during human immunodeficiency virus therapy" (Marzolini, Paus, Buclin, and Kim 2004). Their work points an associating with both PD and IBD, but not directly with CD. This is connected in our graph by the paper "MDR1 gene: susceptibility in Spanish Crohn's disease and ulcerative colitis patients" [PMID16374256]. The authors concluded that "considering our results and those from others, the MDR1 gene behaves as a common risk factor for both CD and UC" (Urcelay et al. 2006)

4.5.4 Retrospective Study: PD and CD – Conclusions

After consideration, the results of our search with a maximum path length of two (Figure E-1), cover all three expected high level concepts: *Immunology, Cell Death, and Genetics*. The search results are again identical between the citational and non-citational approaches, yielding the same P-R (Figure 87) and F_1@K (Figure 88) curves.

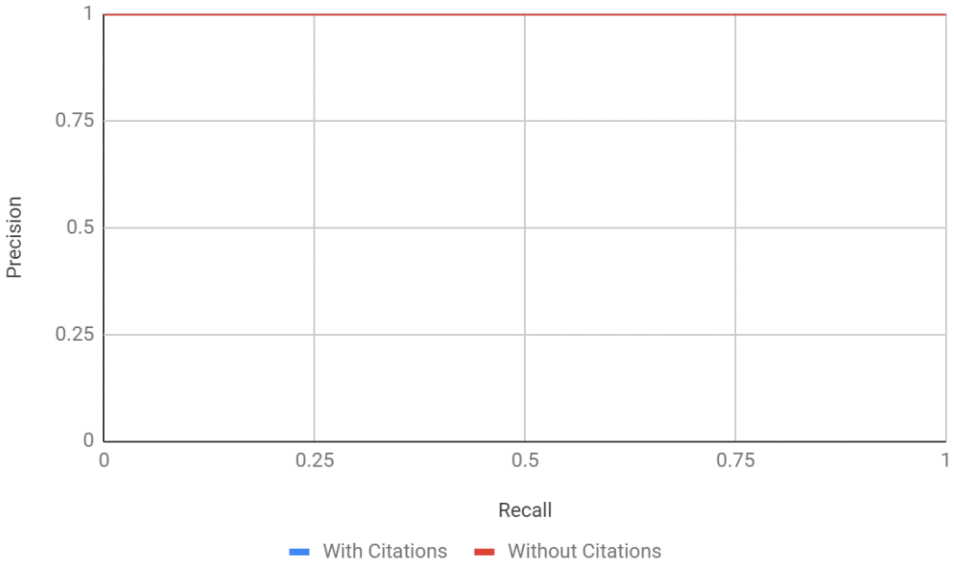


Figure 87. Precision – Recall Curve for Parkinson – Crohn Study

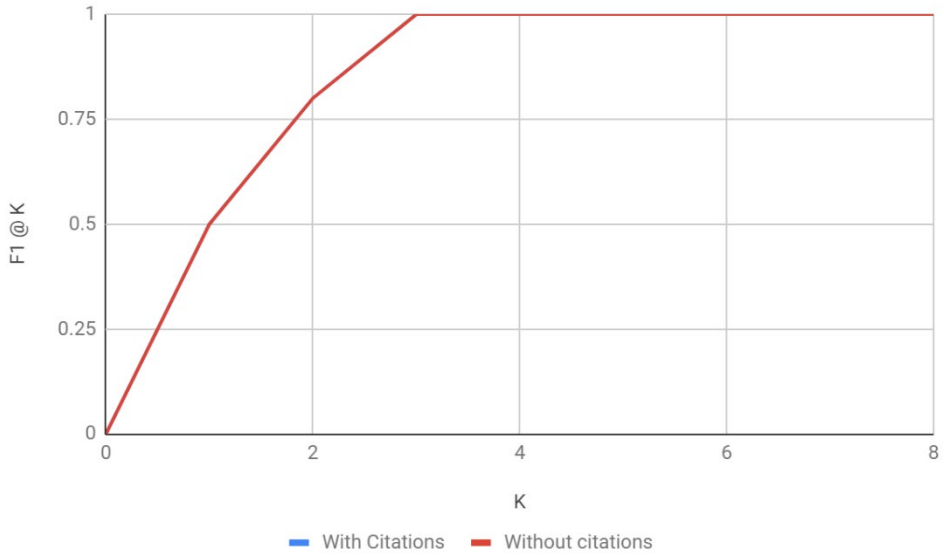


Figure 88. F_1 @ K measure for Parkinson – Crohn Study

However, the details of those connections were not able to explain those associations and, as such, prompted several secondary searches to probe deeper. Of the concepts that appeared within the subsequent searches a number only appeared due to the use of citations.

With the *Inflammation* search (path length three) there were five separate concepts that only appeared when using citations given our corpus. Comparing Figure E-11 and Figure E-12, we find a difference for the concepts *Interleukin-1 beta, TNF gene, and Peroxisome Proliferator-Activated Receptors* (PPAR). *Interleukin-1 beta* is another cytokine, *TNF gene* is responsible for *Tumor Necrosis Factor-alpha* which we discussed previously, and *Peroxisome Proliferator-Activated Receptors* are a group of proteins that help regulate the expression of genes. Another cytokine, *Interleukin-10,* is yet another example that appears when using citations (Figure E-15 and Figure E-16). Lastly, the intermediate *Apoptosis* (Figure E-19) does not appear without citations.

With the *Single Nucleotide Polymorphism* the term *P-Glycoprotein* does not appear without citations as well. This can be seen in the comparison of Figure E-24 and Figure E-25. Although the *ABCB1* gene was already present in the graph, the presence of the corresponding *P-Glycoprotein* helps complete the overall semantic picture.

4.6 Summary

This chapter presented the results of five retrospective experiments performed to determine the validity of a semantic-based LBD approach that makes use of bibliographic citations. In two cases (e.g. the study of *Somatomedin* and *Arginine* and the study of *Testosterone* and *Sleep)*, strong relationships were found by examining predicates of path

length two or less. In these situations we have shown that including citations does not improve the results. However, in the other three studies the inclusion of citations uncovered crucial or additional related concepts. In two of the three cases (*Fish Oils* and *Magnesium* studies) a portion of the intermediates were only uncovered with the use of citational information. In the last case, the *Parkinson's Disease – Crohn's Disease* study, although all of the expected concepts appeared within the graph at a path length of two, the supporting abstracts did not provide sufficient evidence to back the claim. This prompted additional searches with a path length of three wherein multiple concepts appeared due to using predicates included via citations. We conclude from these five studies that the use of information from citationally related documents improves the qualitative results as shown by F_1 and Precision-Recall curves.

Chapter 5

Summary

5.1 Foundation of Work

Cameron et al. (2015) devised a closed-discovery context-driven LBD approach based upon the earlier graph-based semantic summarization methods. Their system automatically generates one or more graphs representing the semantic associations between two user-provided concepts. These graphs are then used to find semantically related pathways from one concept to another. They performed multiple retrospective studies, recovering many key intermediates while utilizing full-text.

In work later performed by Cameron as part of his dissertation, he attempted to rediscover Swanson's *Raynaud Syndrome–Dietary Fish Oils* hypothesis without the use of full-text. Through a similar predicate analysis he noted that "these results collectively suggest that titles and abstracts alone might NOT be sufficient" (Cameron et al. 2016). The results of this study have shown that some intermediates can additionally be uncovered by including cited documents when working with only titles and abstracts.

5.2 Summary of Results

We have performed five retrospective experiments performed to determine the validity of a semantic-based LBD approach that makes use of bibliographic citations. In two cases, such as the Somatomedin and Testosterone studies, strong relationships were found by examining predicates constructing a path with length two or less. In these

situations we have shown that including citations does not improve the results. However, in the other three studies the inclusion of citations uncovered crucial or additional related concepts.

The first study of these successfully reproduced Swanson's original *Fish Oils – Raynaud Disease* hypothesis without the use of full-text by modifying the approach by Cameron et al. to incorporate bibliographically related documents. A detailed examination shows that the intermediate *Vascular Reactivity* is not present in the core references and only appears by using full-text or by including cited documents.

The *Magnesium – Migraine Disorders* discovery by Swanson was the second study. For the eleven expected intermediate concepts, all of the seven that Cameron found were found in our reduced corpus. For the other four concepts (*Hypoxia, Stress, Substance P, Spreading Depression*) noticed by Swanson that Cameron did not find, we were able to discover three - both *Stress/Type A Personality* and *Hypoxia* were found within our original search, and *Substance P* was discovered via discovery browsing of related concepts. Additionally, our results include a twelfth concept, *Diabetes*, undiscovered by Swanson and Cameron, only included via cited documents.

The third study to yield strong results was a reproduction of Kostoff's research on the common factors of *Parkinson's disease* and *Crohn's disease*. Here, all three high level concepts (*Immunology, Cell Death, and Genetics*), appeared in some fashion without the use of cited documents, but the details of those connections were not able to explain those associations. Thus, several secondary searches were used to probe deeper. Of the concepts that appeared within the subsequent searches a number only appeared

due to the use of citations. Between the three intermediate topics there were an additional six separate concepts that only appeared when using citations given our corpus.

In conclusion, our findings conclude that the inclusion of citational relationships can be used to improve the results of LBD research. Quantitative and qualitative improvements are made by uncovering intermediates that would not be found otherwise. This not only can improve the resulting Precision-Recall curves for the results, but can provide additional meaning in cases where a supporting statement is too generic, such as within our *Parkinson's disease – Crohn's disease* study.

5.3 Future Work

This research's contribution is the illustration of the information contained within citational references as it applies to LBD. While evidence suggests that it supports the validity of this approach, this is also the biggest limitation of this work. This is relevant in that although titles and abstracts for most documents tend to be easy to acquire due to modern digitalization efforts, citational information from one document to others is not. Further research leveraging these associations is possible, either in datasets where citations are available, or in a more complex manner as greater amounts of this information is recorded.

Also, due to the nature of our five retrospective studies and their focus on bibliographic citations, each study was performed using a relatively small corpus. As noted previously, this corpus was constructed from the references listed in each of the original studies along with the papers that they themselves cited. As such, the entire corpus contained only thousands of papers. Thus, despite the fact that concept IDF values

were calculated using all of the MEDLINE papers, predicate graphs were constructed only using the reduced corpus. An investigation should be made into the efficacy of using bibliographic citations at full scale.

Additionally there are systemic limitations that should be addressed. First, and foremost, is the reliance upon the generation of semantic predications from a NLP tool. Both as noted in this paper and in the work by Cameron et al. (2016), several predicates were found to be erroneous due to an incorrect concept. This is only one instance of the potential difficulties. Further difficulties arrive when contracting semantic predications are generated from text, either due to nuances in the way the text was composed, results may be based upon apriori assumptions or situations, or changing/conflicting scientific evidence.

In a similar fashion a more specific improvement can also be made regarding the association of similar concepts. It was noticed several times during our exploratory studies that there are separate concepts that are very similar (e.g. *Magnesium*, *Magnesium Deficiency*, and *Hypomagnesemia).* When predicates are generated from documents different concepts might be used, making it challenging to find a connecting path. For example, perhaps there is a path connecting *Magnesium Deficiency* to *Migraine Disorders* but this won't be found for a search based on *Magnesium*. An expert user that is well versed in the subject matter could be aware of such pitfalls and make searches accordingly but an average user would miss out on potential connections. Modifications could be made to treat concepts that are sufficiently similar but would need to do so within a given context (referring back to section 2.5.2 in regards to LSA).

Appendix A

Fish Oils – Raynaud Syndrome Study

Table A-1. Fish Oils – Raynaud Syndrome Core Corpus Medline IDs

53042	4002440	6229551	6340424	6775675
58309	4015748	6298902	6352267	6788326
277163	4082084	6301111	6365102	6812750
596950	6088955	6302714	6376801	6827988
2408588	6097237	6303363	6424425	6890719
2412054	6098049	6307322	6432198	7031981
2412071	6098051	6314583	6449756	7037038
2997286	6100033	6318123	6540787	7161779
3157318	6128596	6320840	6613908	7259326
3883365	6130329	6320945	6636033	7295999
3888229	6134122	6321621	6639916	7364865
3977414	6136879	6330926	6707529	7470209
3990714	6209510	6339921	6718836	14289442

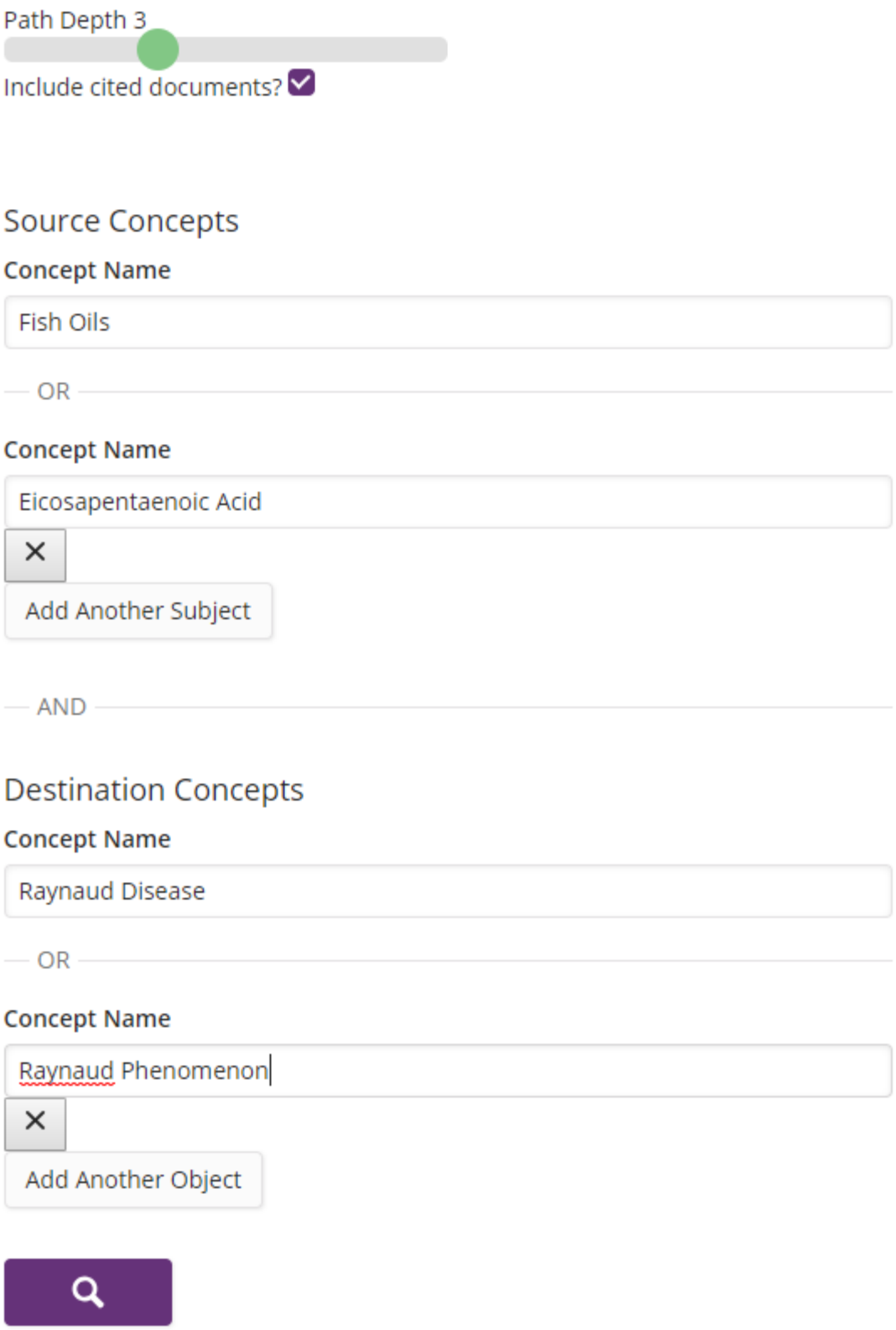


Figure A-1. Fish Oils – Raynaud Disease Search Parameters

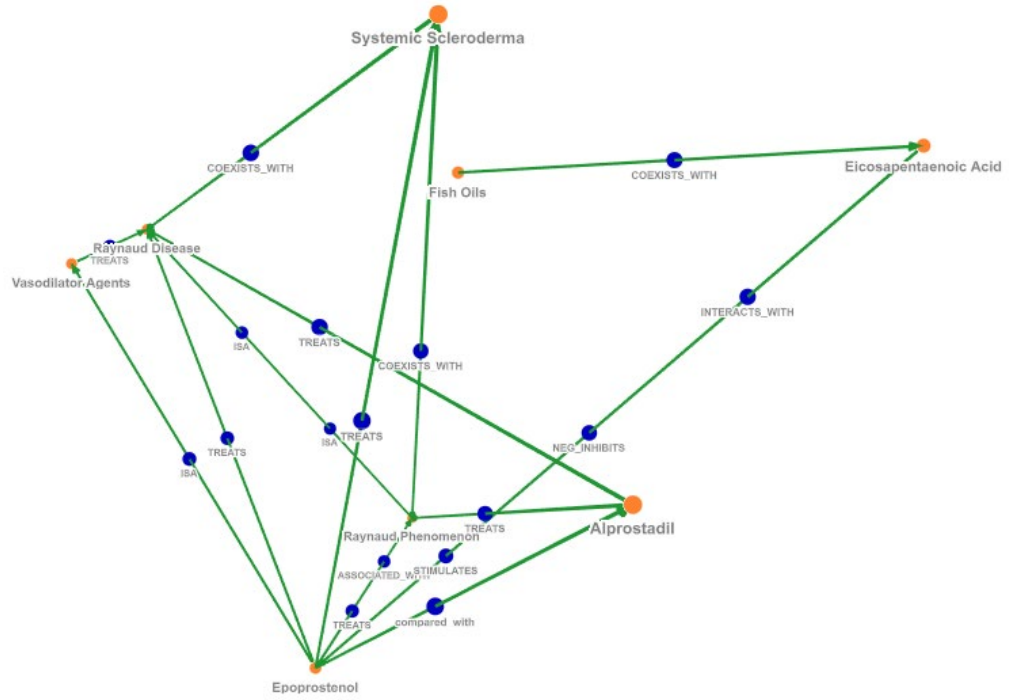


Figure A-2. Fish Oils – Raynaud Disease result 1 (with citations)

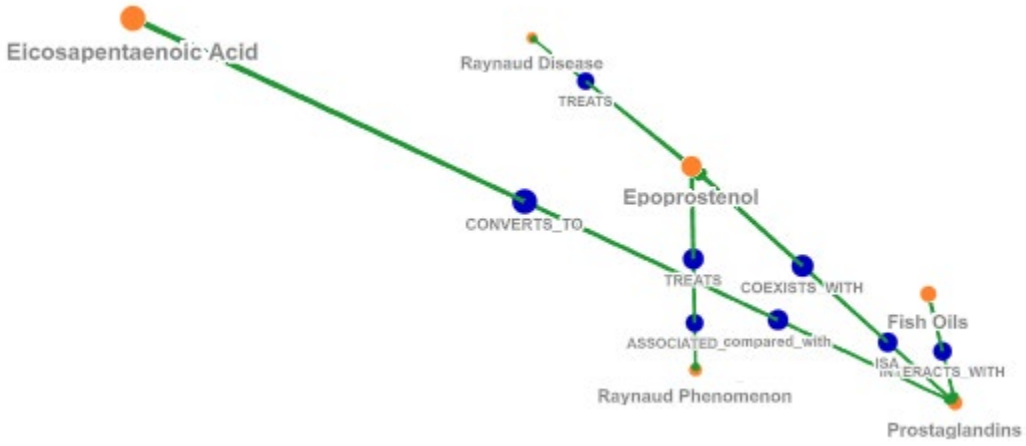


Figure A-3. Fish Oils – Raynaud Disease result 2 (with citations)

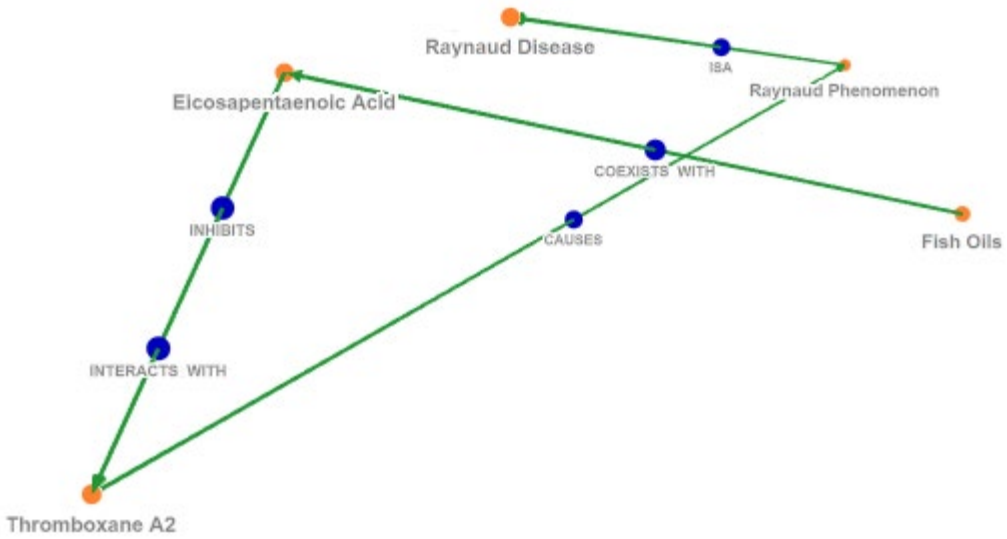


Figure A-4. Fish Oils – Raynaud Disease result 3 (with citations)

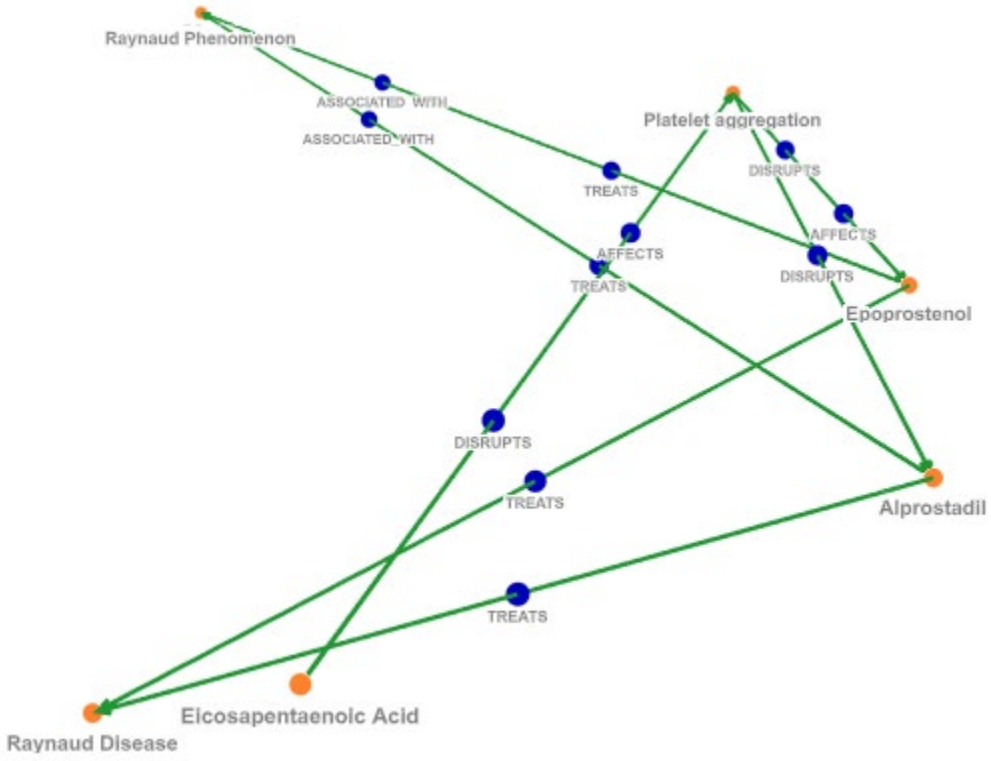


Figure A-5. Fish Oils – Raynaud Disease result 4 (with citations)

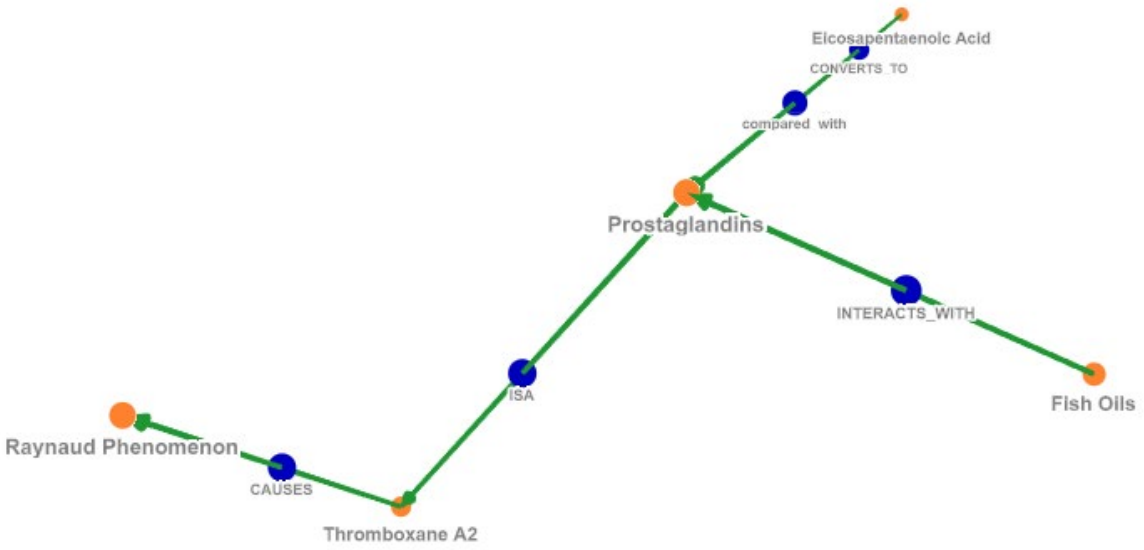


Figure A-6. Fish Oils – Raynaud Disease result 5 (with citations)

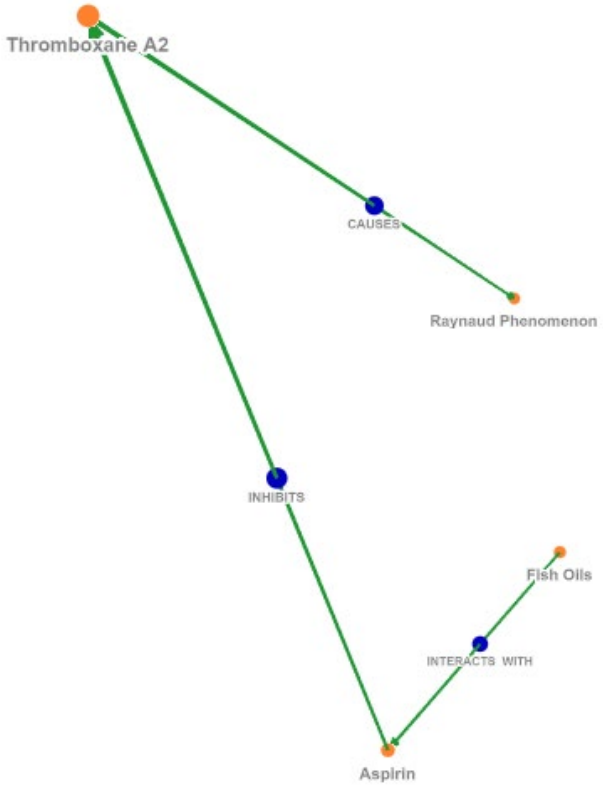


Figure A-7. Fish Oils – Raynaud Disease result 6 (with citations)

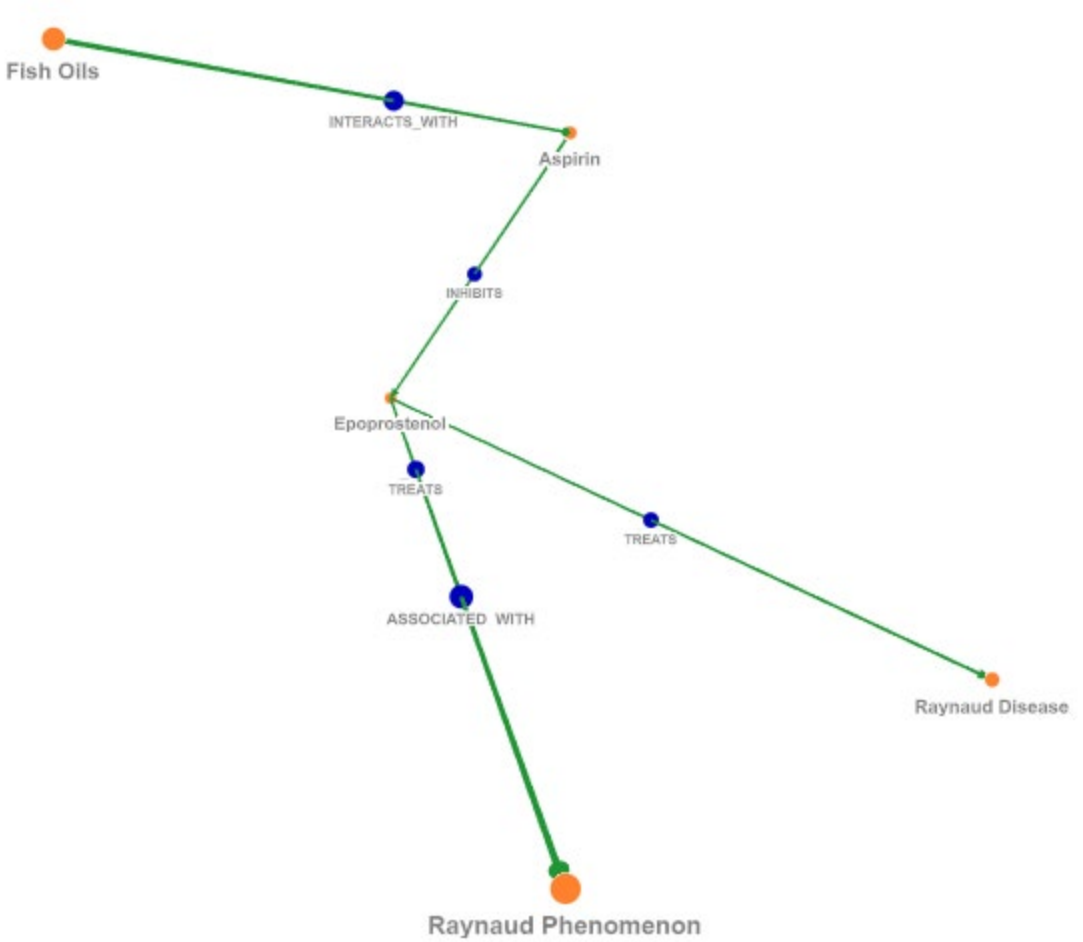


Figure A-8. Fish Oils – Raynaud Disease result 7 (with citations)

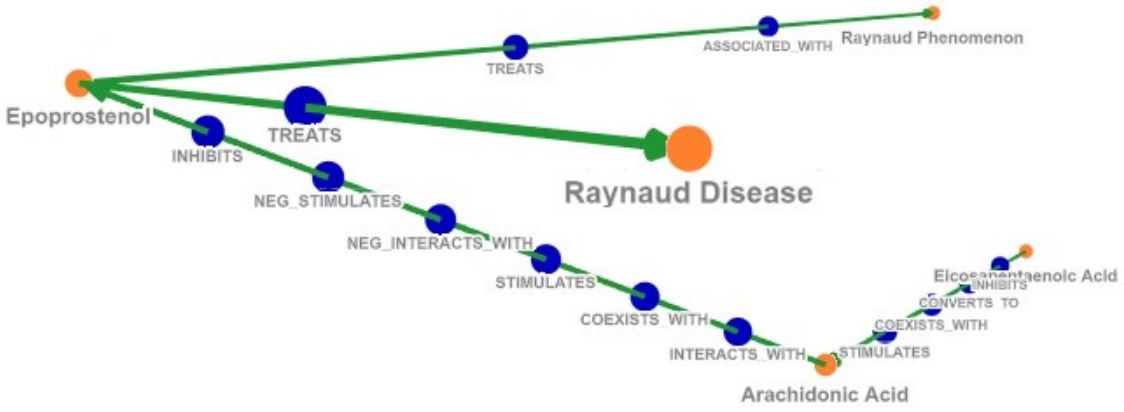


Figure A-9. Fish Oils – Raynaud Disease result 8 (with citations)

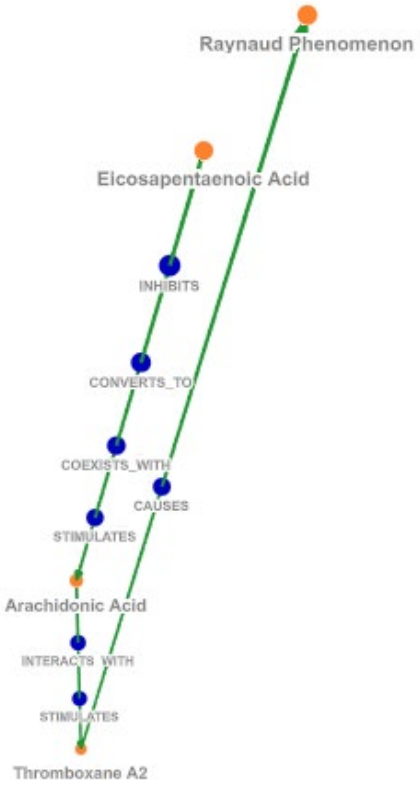


Figure A-10. Fish Oils – Raynaud Disease result 9 (with citations)

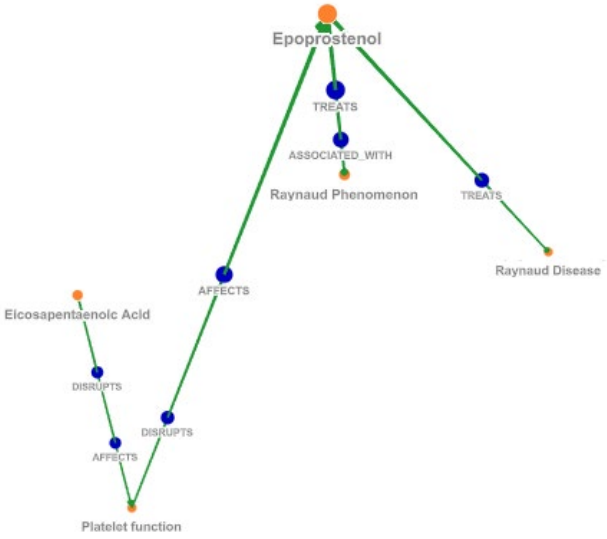


Figure A-11. Fish Oils – Raynaud Disease result 10 (with citations)

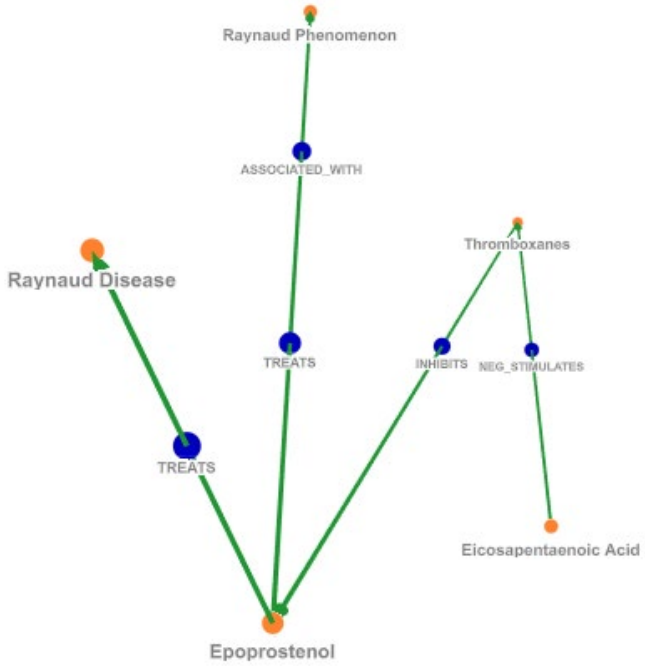


Figure A-12. Fish Oils – Raynaud Disease result 11 (with citations)

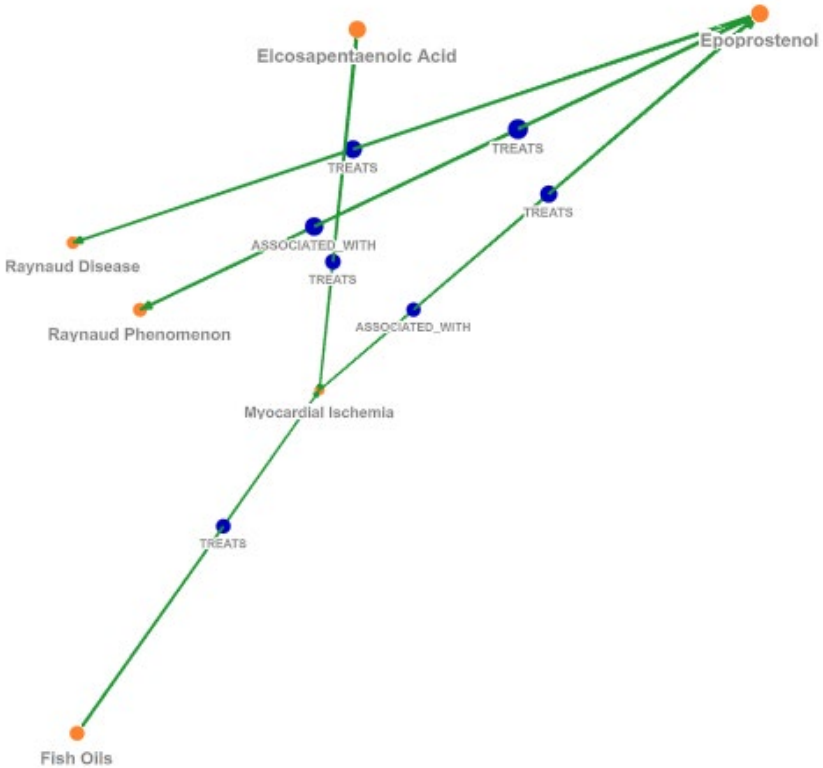


Figure A-13. Fish Oils – Raynaud Disease result 12 (with citations)

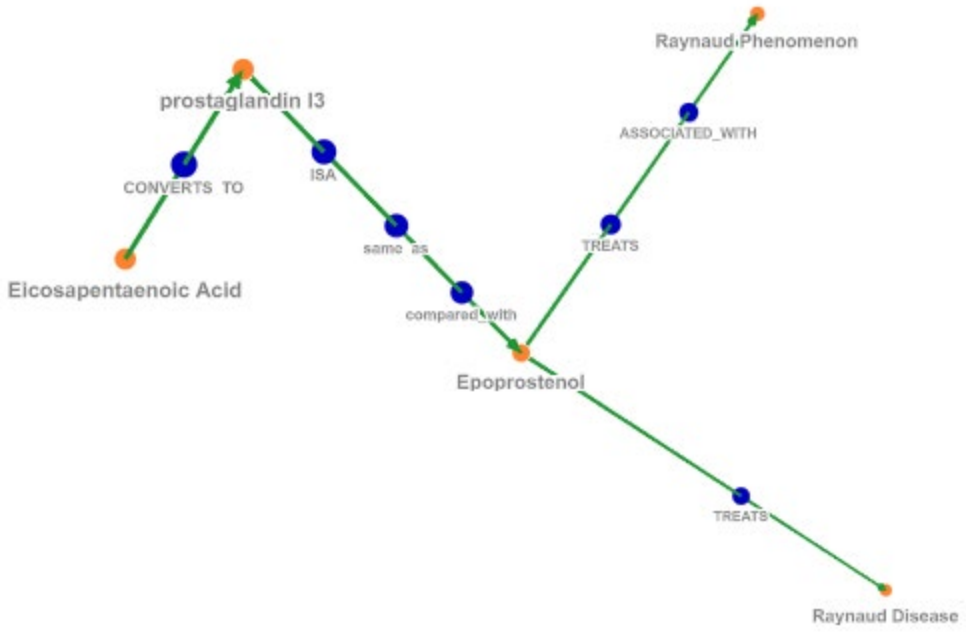


Figure A-14. Fish Oils – Raynaud Disease result 13 (with citations)

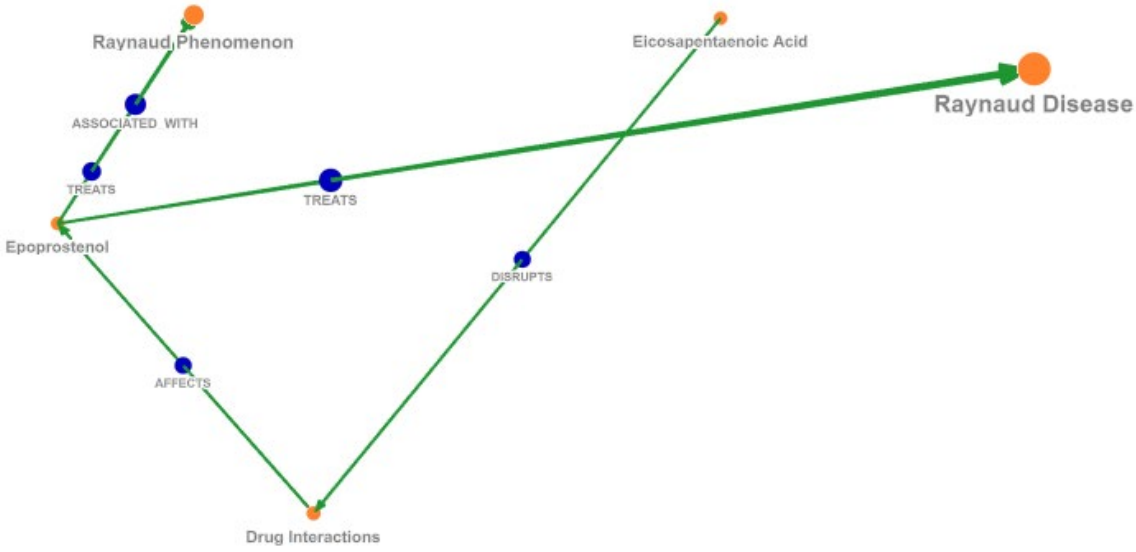


Figure A-15. Fish Oils – Raynaud Disease result 14 (with citations)

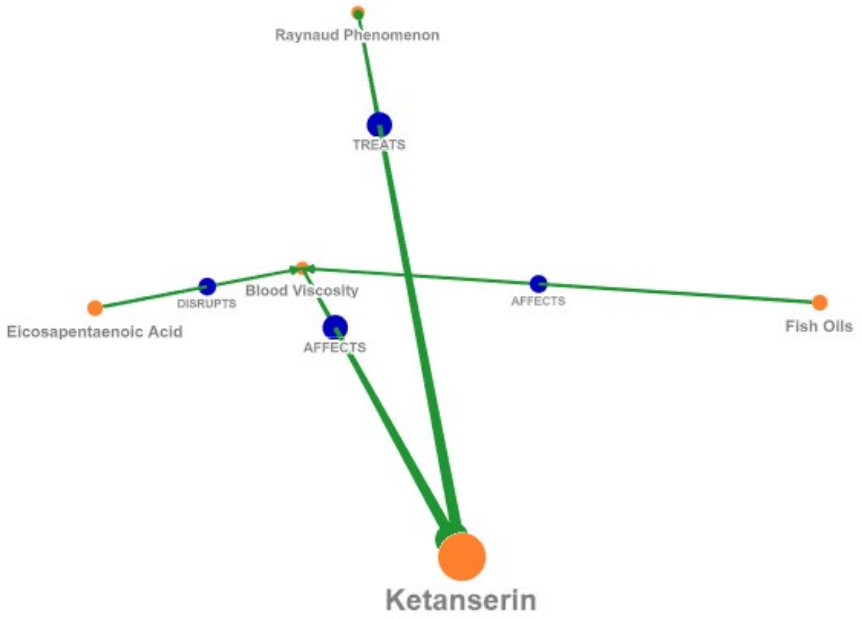


Figure A-16. Fish Oils – Raynaud Disease result 15 (with citations)

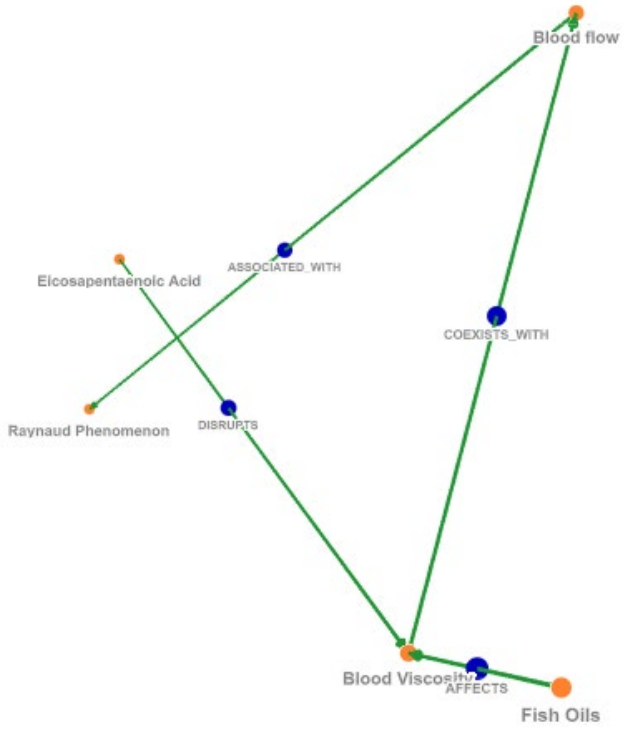


Figure A-17. Fish Oils – Raynaud Disease result 16 (with citations)

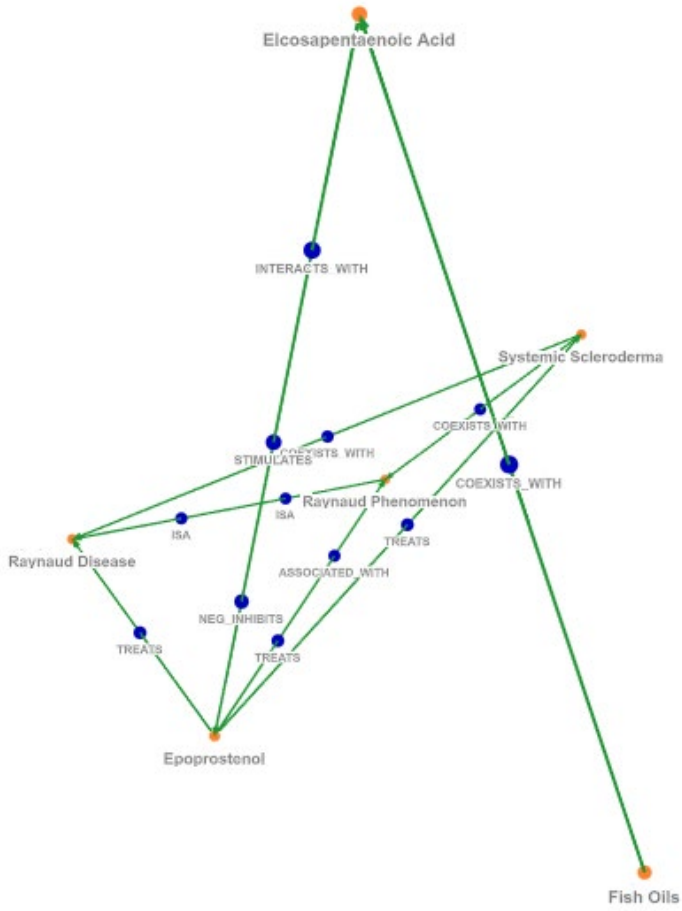


Figure A-18. Fish Oils – Raynaud Disease result 1 (no citations)

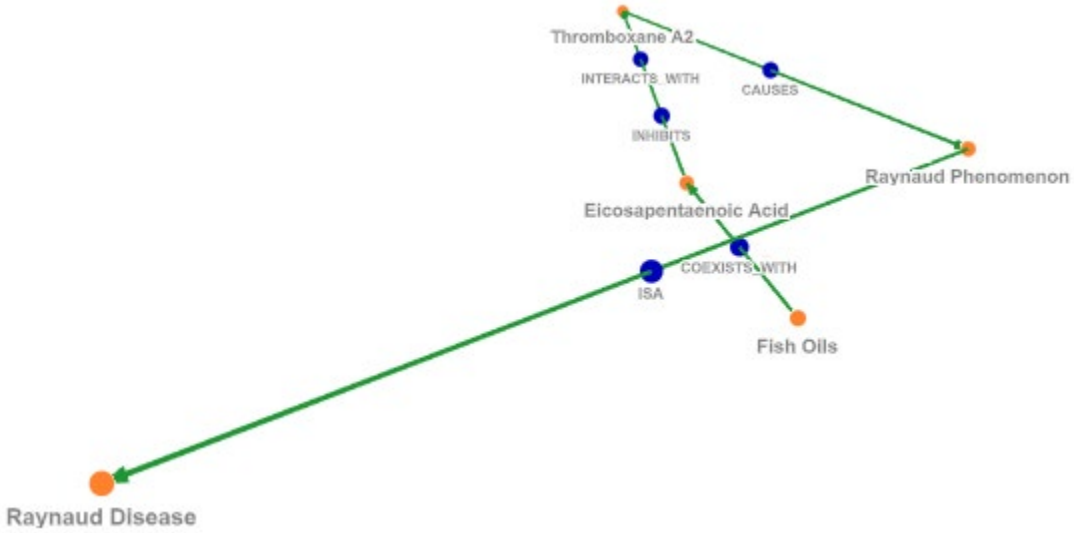


Figure A-19. Fish Oils – Raynaud Disease result 2 (no citations)

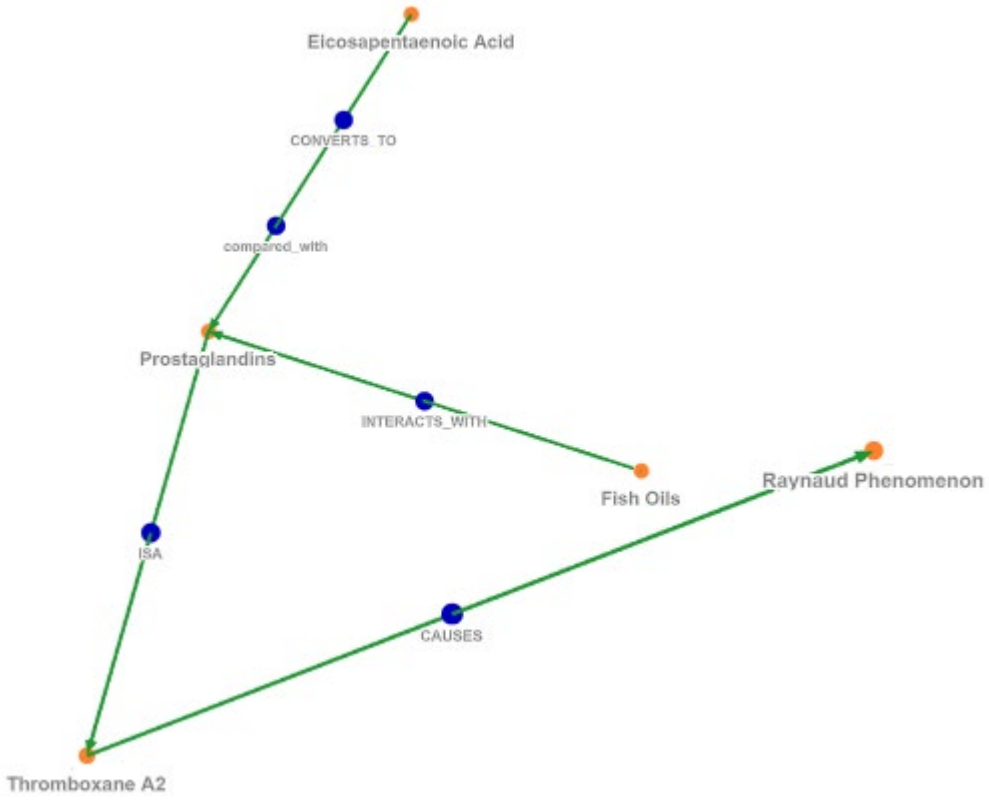


Figure A-20. Fish Oils – Raynaud Disease result 3 (no citations)

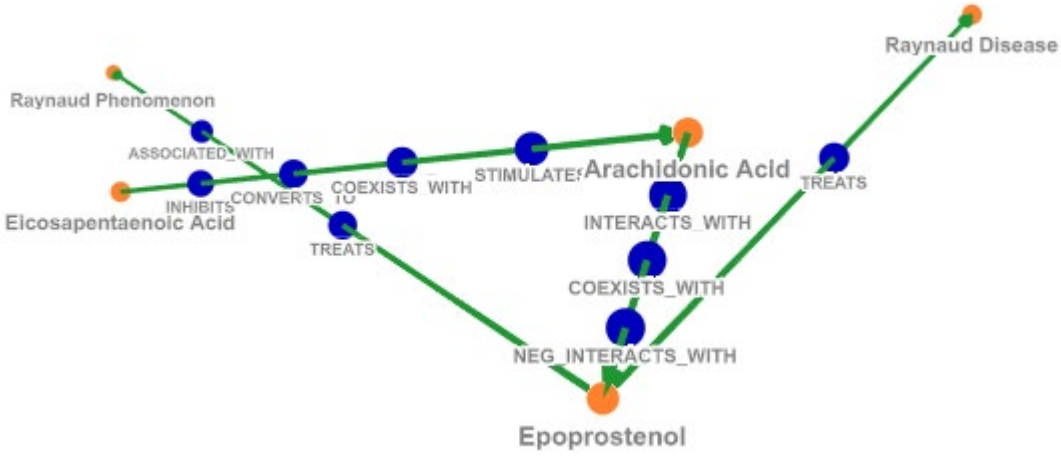


Figure A-21. Fish Oils – Raynaud Disease result 4 (no citations)

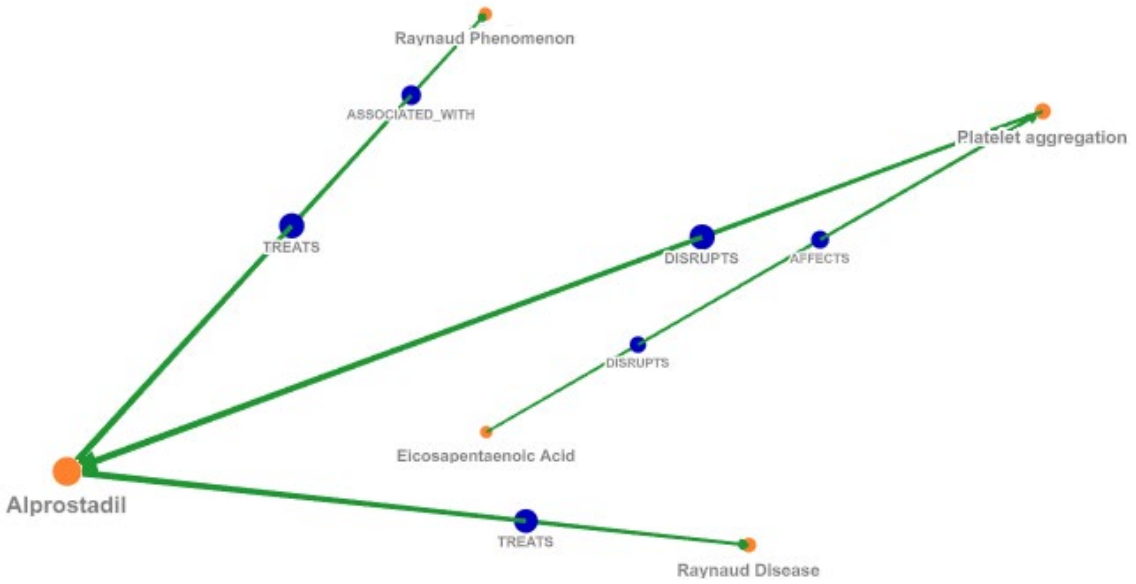


Figure A-22. Fish Oils – Raynaud Disease result 5 (no citations)

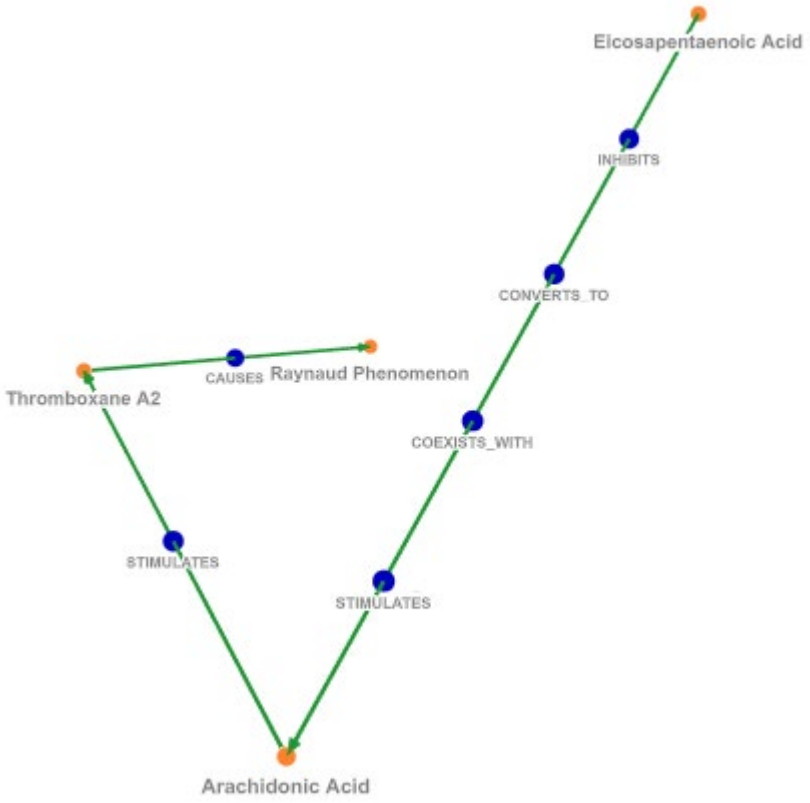


Figure A-23. Fish Oils – Raynaud Disease result 6 (no citations)

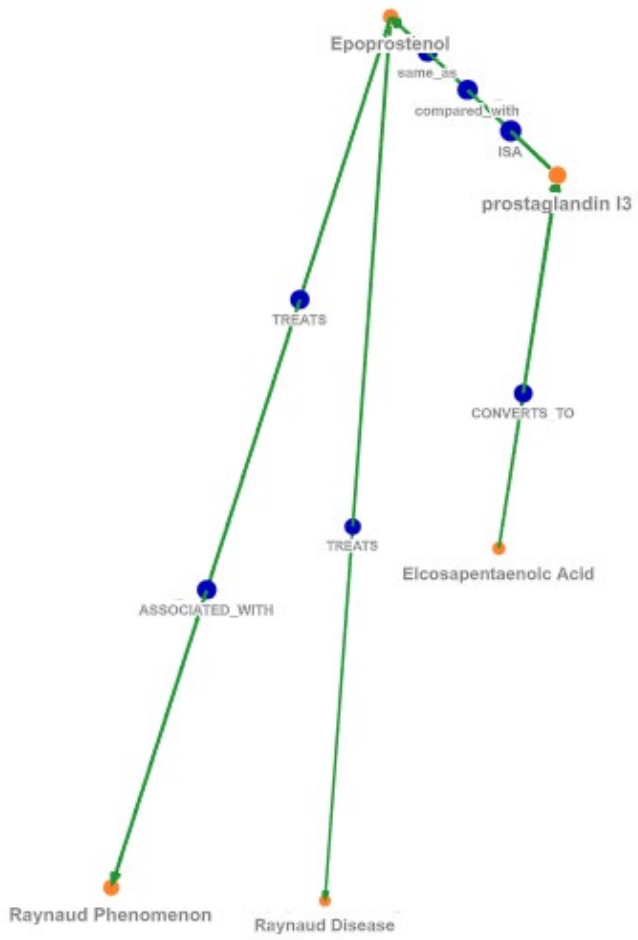


Figure A-24. Fish Oils – Raynaud Disease result 7 (no citations)

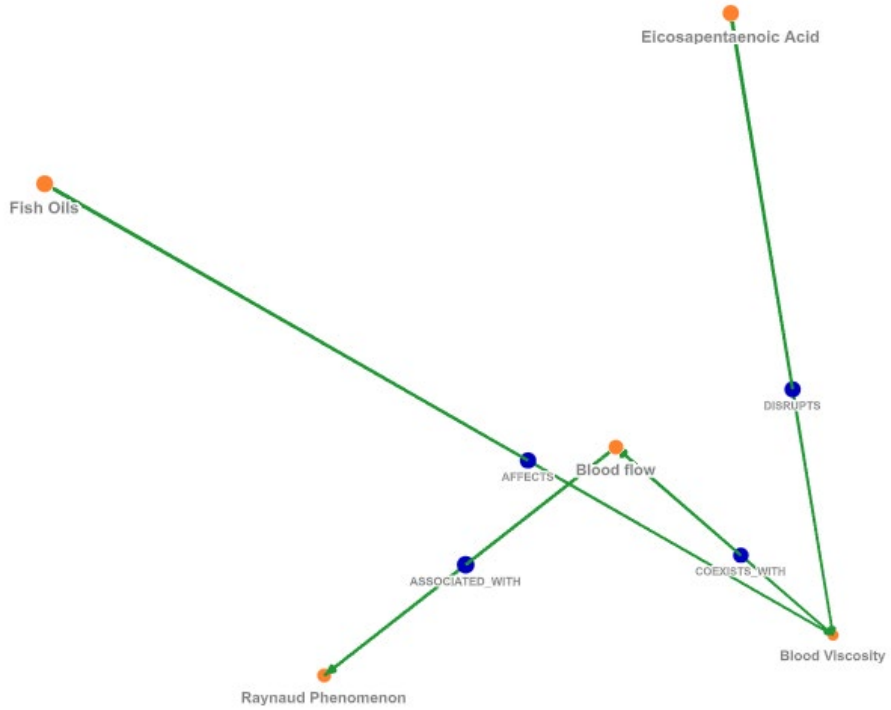


Figure A-25. Fish Oils – Raynaud Disease result 8 (no citations)

Table A-2. Fish Oils – Raynaud Disease generated predicates

[PMID114606]
Abstract "Both PGH2 and PGI2 could produce dose-dependent contraction or relaxation of isolated arteries."
[C0033567|Epoprostenol – CAUSES – C1140999|Contraction]
Abstract "As PGI2 is the most potent cerebral vasodilator drug tested, it may be of clinical use in the treatment of cerebral vasospasm."
[C0033567|Epoprostenol – ISA – C0042402|Vasodilator Agents] [C0033567|Epoprostenol – TREATS(SPEC) – C0265110|Cerebral Vasospasm] [C0042402|Vasodilator Agents – TREATS – C0265110|Cerebral Vasospasm]
[PMID190267]
Abstract "In contrast, when arachidonic acid is added directly to platelets, prior incubation with dBcAMP or PGE1 does not inhibit production of the prostaglandins or their metabolites."
[C0002335|Alprostadil – NEG_INHIBITS – C0033554|Prostaglandins] [C0003695|Arachidonic Acid – NEG_INHIBITS – C0033554|Prostaglandins]
Abstract "Furthermore, dBcAMP and PGE1 both inhibit platelet aggregation induced by either arachidonic acid or prostaglandin H2 without affecting the production of prostaglandin metabolites from these compounds."
[C0002335|Alprostadil – DISRUPTS – C0032176|Platelet aggregation] [C0003695|Arachidonic Acid – AUGMENTS – C0032176|Platelet aggregation] [C0072288|Prostaglandin H2 – AUGMENTS – C0032176|Platelet aggregation]
[PMID2408588]
Title "Inhibition of platelet aggregation by a new stable prostacyclin introduced in therapy of patients with progressive scleroderma."
[C0033567|Epoprostenol – DISRUPTS – C0032176|Platelet aggregation]

[PMID2412054]
Abstract “The effects of serotonin and its pharmacological antagonists on the physical flow properties of the blood have been studied far less than their effects on blood vessels, although they may be equally important.”
[C0036751|Serotonin – AFFECTS – C0005847|Blood Vessels]
Abstract “Indirect evidence suggests that in pathological circumstances serotonin may locally increase whole blood viscosity, particularly at low shear rates, decrease red cell deformability and increase the adhesiveness of white cells.”
[|C0036751|Serotonin – AUGMENTS – C0005848|Blood Viscosity]
Abstract “Specific serotonergic-antagonists, administered either orally or intravenously, normalize the increased whole blood viscosity and decreased blood filterability found in essential hypertension, following myocardial infarction and in severe leg ischaemia”
[C0948013|Increased blood viscosity – COEXISTS_WITH – C0085580|Essential Hypertension]
Abstract “8 Ketanserin given intravenously for seven days to patients with very severe leg ischaemia, significantly improves whole blood viscosity, increases red cell transit time and most dramatically decreases pore clogging.”
[C0022616|Ketanserin – AFFECTS – C0005848|Blood Viscosity]
[PMID364545]
Abstract “The activity of prostacyclin (PGI2), PGE1 or PGD2 as inhibitors of platelet aggregation in plasma from human, dog, rabbit, rat, sheep and horse was investigated.”
[C0002335|Alprostadil – DISRUPTS – C0032176|Platelet aggregation]
Abstract “5 Theophylline or dipyridamole potentiated the inhibition of human platelet aggregation by prostacyclin, PGE1 or PGD2.”

[C0002335|Alprostadil – DISRUPTS – C0032176|Platelet aggregation] [C0033567|Epoprostenol – DISRUPTS – C0032176|Platelet aggregation]
[PMID378308]
Abstract "...due to the formation of platelet aggregates it is concluded that PGI2 is a most potent anti-aggregatory prostaglandin…"
[C0033567|Epoprostenol – ISA – C0033554|Prostaglandins]
[PMID3883365]
Abstract "As prostacyclin (PGI2) is of benefit in the treatment of RS in SS, we have measured endogenous stable metabolites of PGI2 (PGI2m) in 42 patients with Raynaud's Phenomenon (RP) of varying aetiology (15 SS patients, 15 patients with Raynaud's Disease (RD) but no other symptoms, and 12 other RS patients with probable connective tissue disorder)."
[C0033567|Epoprostenol – TREATS - |C0034734|Raynaud Disease] [C0033567|Epoprostenol – TREATS(INFER) – C0034735|Raynaud Phenomenon] [C0034734|Raynaud Disease – COEXISTS_WITH – C0036421|Systemic Scleroderma]
[PMID3977414]
Abstract "An improved thermographic response to cold stress testing after treatment of Raynaud's syndrome with PGE1 has also been shown for the first time."
[C0002335|Alprostadil – TREATS – C0034734|Raynaud Disease]
[PMID531220]
Abstract – Taken from publisher's site. "In human platelet-rich plasma (PRP) eicosapentaenoic acid (EPA) inhibited platelet aggregation induced by a stable analogue of PGH2 (U46619), arachidonic acid, collagen or ADP."
[C0000545|Eicosapentaenoic Acid – DISRUPTS – C0032176|Platelet aggregation] [C0917834|U-44619 – ISA – C0072288|Prostaglandin H2]

Abstract – Taken from publisher's site. "EPA incubated with PRP did not induce the generation of a thromboxane (TXA)-like activity; indeed it prevented the formation of TXA2 induced by arachidonic acid or by collagen."
[C0000545|Eicosapentaenoic Acid – NEG_STIMULATES – C0040061|Thromboxanes] [C0003695|Arachidonic Acid – STIMULATES – C0040057|Thromboxane A2]
Abstract – Taken from publisher's site. "The anti-aggregatory activity of EPA was not influenced by inhibitors of cyclo-oxygenase and lipoxygenase."
[C0000545|Eicosapentaenoic Acid – INHIBITS – C0023837|Lipoxygenase] [C0000545|Eicosapentaenoic Acid – INHIBITS – C0033551|Prostaglandin-Endoperoxide Synthase]
[PMID6303363]
Abstract "Eicosapentaenoic acid (EPA), which is abundant in seafood, has been reported to be a potent antagonist of platelet aggregation and also to reduce the incidence of cardiovascular disorders."
[C0000545|Eicosapentaenoic Acid – DISRUPTS – C0032176|Platelet aggregation]
Abstract "We recently reported that EPA also reduces whole blood viscosity."
[C0000545|Eicosapentaenoic Acid – DISRUPTS – C0005848|Blood Viscosity]
[PMID6314583]
Abstract "EPA was not able to be converted to any prostaglandins (PGs) in murine and porcine smooth muscle cells."
[C0000545|Eicosapentaenoic Acid – CONVERTS_TO – C0033554|Prostaglandins]
Abstract "In rat, EPA was not only being converted to no PGI3, but also being a blocker to PGI2 synthesis in vascular cells."

[C0000545|Eicosapentaenoic Acid – CONVERTS_TO|C0072294|prostaglandin I3] [C0005773|Blood Cells – LOCATION_OF – C0033567|Epoprostenol] [C0072294|prostaglandin I3 – ISA – C0033567|Epoprostenol]
Abstract "Moreover, the rat EPA has much less activity in inhibiting thromboxane A2 (TXA2) synthesis in platelets."
[C0000545|Eicosapentaenoic Acid – INTERACTS_WITH – C0040057|Thromboxane A2]
Abstract "On the contrary, in human EPA was not only easily converted to PGI3 in vascular cells, but also blocking TXA2 synthesis in platelets."
[C0000545|Eicosapentaenoic Acid – CONVERTS_TO – C0072294|prostaglandin I3]
[PMID6320945]
Title "Beneficial effect of fish oil on blood viscosity in peripheral vascular disease."
[C0016157|Fish Oils – AFFECTS – C0005848|Blood Viscosity]
[PMID6339921]
Title "Controlled double-blind trial of nifedipine in the treatment of Raynaud's phenomenon"
[C0028066|Nifedipine – TREATS – C0034735|Raynaud Phenomenon]
Abstract – Taken from publisher's site. "The treatment of this condition is difficult.1 Nifedipine, a slow calcium-channel antagonist, causes vascular smooth-muscle relaxation and relief of arterial vasospasm."
[C0028066|Nifedipine – TREATS – C0085616|Vasospasm]
Abstract – Taken from publisher's site. "Nifedipine-induced vasodilation has been shown to result in a fall in peripheral vascular resistance and an increase in peripheral blood flow."

[C0042401|Vasodilation – COEXISTS_WITH – C1258192|Total Peripheral Resistance]
Abstract – Taken from publisher's site. "Nifedipine increases the skin blood flow by 50 to 150 per cent in normal persons and has also been shown in vitro to inhibit norepinephrine-induced vasospasm of rabbit and human peripheral arteries and veins."
[C0028066|Nifedipine – AUGMENTS – C0232338|Blood flow]
Abstract – Taken from publisher's site. "The calcium-channel blockers nifedipine and verapamil have …"
[C0028066|Nifedipine – ISA – C0006684|Calcium Channel Blockers]
[PMID6365102]
Title "Treatment of Raynaud's phenomenon with ketanserin, a selective antagonist of the serotonin2 (5-HT2) receptor."
[C0022616|Ketanserin – TREATS – C0034735|Raynaud Phenomenon]
Abstract "6 These findings lend support to the hypothesis that serotonin is an important element in the pathogenesis of systemic sclerosis."
[C0036751|Serotonin – CAUSES – C0036421|Systemic Scleroderma]
[PMID6432198]
Title "Treatment of Raynaud's phenomenon with ketanserin in patients with connective tissue disorders."
[C0022616|Ketanserin – TREATS - |C0034735|Raynaud Phenomenon]
Abstract "The results of this study suggest that orally administered ketanserin may be an effective and well tolerated treatment for Raynaud's phenomenon associated with connective tissue disorders, especially scleroderma."

[C0022616|Ketanserin – TREATS – C0034735|Raynaud Phenomenon]
[PMID6773615]
Abstract "These results support our suggestion that a prostaglandin, in particular PGI2, is required for hypercapnia to produce full cerebral vasodilatation."
[C0033554|Prostaglandins – COEXISTS_WITH – C0033567|Epoprostenol] [C0033567|Epoprostenol – ISA – C0033554|Prostaglandins]
[PMID6788326]
Title "Effect of prostaglandins I2 and E1 on red cell deformability in patients with Raynaud's phenomenon and systemic sclerosis."
[C0033567|Epoprostenol – TREATS(INFER) – C0034735|Raynaud Phenomenon] [C0033567|Epoprostenol – TREATS(INFER) – C0036421|Systemic Scleroderma]
[PMID6812750]
Abstract "This study suggests that ketanserin may be useful in the treatment of Raynaud's phenomenon"
[C0022616|Ketanserin – TREATS – C0034735|Raynaud Phenomenon]
[PMID6821892]
Abstract "Formation of prostaglandins derived from eicosapentaenoic acid and interference of eicosapentaenoic acid with formation and action of prostaglandins derived from arachidonic acid were evident in vitro."
[C0000545|Eicosapentaenoic Acid – CONVERTS_TO – C0033554|Prostaglandins] [C0003695|Arachidonic Acid – CONVERTS_TO – C0033554|Prostaglandins]
[PMID7003784]
Abstract "Prostaglandins E1 (PGE1) and I2 (prostacyclin, PGI2) are potent vasodilators and

inhibitors of platelet aggregation which have been reported to be of value in the treatment of peripheral ischemia.”
[C0002335|Alprostadil – DISRUPTS – C0032176|Platelet aggregation] [C0033567|Epoprostenol – ISA – C0042402|Vasodilator Agents]
[PMID7025341]
Title “Successful treatment of Raynaud’s Syndrome with prostacyclin.”
[C0033567|Epoprostenol – TREATS – C0034734|Raynaud Disease]
Abstract “We evaluated the effect of prostacyclin (PGI2) a potent antiplatelet and vasodilator agent in 5 female patients with Raynaud’s syndrome.”
[C0033567|Epoprostenol – TREATS(INFER) – C0034734|Raynaud Disease] [C0042402|Vasodilator Agents – TREATS(INFER) – C0034734|Raynaud Disease]
[PMID7037038]
Title “Treatment of Raynaud’s phenomenon by intravenous infusion of prostacyclin (PGI2)”
[C0033567|Epoprostenol – TREATS – C0034735|Raynaud Phenomenon]
[PMID7208950]
Abstract “The effect of altering the abundance of precursors and inhibitors of prostaglandin formation by dietary supplements of fish oil…”
[C0033554|Prostaglandins – INTERACTS_WITH – C0016157|Fish Oils]
[PMID7259326]
Abstract “PGE1 may therefore be suitable treatment for Raynaud’s phenomenon and the vascular insufficiency of systemic sclerosis and other connective tissue diseases.”
[C0002335|Alprostadil – TREATS – C0034735|Raynaud Phenomenon]

[PMID7427564]
Abstract “Prostaglandin E1, a vasodilator and potent inhibitor of platelet aggregation, was administered to 26 patients with severe vasospastic disease of the hands.”
[C0002335|Alprostadil – DISRUPTS – C0032176|Platelet aggregation] Abstract “Intensity of digital vasospasm induced by cold water challenge was not objectively affected by prostaglandin E1 despite an increased finger temperature after infusion.”
[C0002335|Alprostadil – AFFECTS – C0085616|Vasospasm]

Table A-3. Precision and recall at each result using citations for Fish Oil study

Result	New Intermediate Terms	Relevancy (Topic)	Precision	Recall
d0	Epoprostenol Alprostadil Vasodilator Agents Systemic Scleroderma	Platelet Aggregation Platelet Aggregation Vascular Reactivity Not Relevant	0.667	0.667
d1	Prostaglandins	Platelet Aggregation	0.667	0.667
d2	Thromboxane A2	Platelet Aggregation	0.667	0.667
d3	Platelet Aggregation	Platelet Aggregation	0.667	0.667
d4	N/A		0.667	0.667
d5	Aspirin	Platelet Aggregation	0.667	0.667
d6	N/A		0.667	0.667
d7	Arachidonic Acid	Not Relevant	0.5	0.667
d8	N/A		0.5	0.667
d9	Platelet Function	Platelet Aggregation	0.5	0.667
d10	Thromboxanes	Platelet Aggregation	0.5	0.667
d11	Myocardial Ischemia	Not Relevant	0.4	0.667
d12	Prostaglandin I3	Platelet Aggregation	0.4	0.667
d13	Drug Interactions	Not Relevant	0.333	0.667
d14	Blood Viscosity Ketanserin	Blood Viscosity Blood Viscosity	0.429	1.0
d15	Blood flow	Blood Viscosity	0.429	1.0

Table A-4. Precision and recall at each result not using citations for Fish Oil study

Result	New Intermediate Terms	Relevancy (Topic)	Precision	Recall
d0	Epoprostenol Systemic Scleroderma	Platelet Aggregation Not Relevant	0.5	0.333
d1	Thromboxane A2	Platelet Aggregation	0.5	0.333
d2	Prostaglandins	Platelet Aggregation	0.5	0.333
d3	Arachidonic Acid	Not Relevant	0.333	0.333
d4	Platelet Aggregation Alprostadil	Platelet Aggregation Platelet Aggregation	0.333	0.333
d5	N/A		0.333	0.333
d6	Prostaglandin I3	Platelet Aggregation	0.333	0.333
d7	Blood Viscosity Blood flow	Blood Viscosity Blood Viscosity	0.5	0.667

Appendix B

Magnesium – Migraine Disorders Study

Table B-1. Magnesium – Migraine Disorders Core Corpus Medline IDs

90971	3523056	3897128	6116859	6784664
96740	3523058	3898850	6141289	6826350
108361	3524092	3908832	6146019	6836300
330466	3530775	3918779	6150068	6861894
414910	3530778	3946808	6163401	6869560
418838	3531248	3962783	6170210	6870999
630136	3536167	4016924	6172961	6878586
739264	3537077	4016926	6206779	6881926
750727	3582557	4016930	6230778	7085796
1110377	3585352	4016941	6275679	7120030
1262918	3607427	4016945	6296376	7181451
2408700	3616619	4016946	6316376	7334315
2425960	3705931	4019939	6339166	7348028
2433912	3713254	4021225	6375330	7400981
2435656	3715946	4167805	6392763	7402146
2864494	3724238	4323509	6438124	7443079
2867984	3733214	4725298	6464059	10303918
3022872	3758107	4842482	6481797	13295838
3107334	3758108	4978139	6503357	13597818
3160475	3763042	4990911	6525007	13645235
3300911	3771208	5514170	6609739	14212747
3308769	3797213	5573439	6620432	14338128
3458981	3801897	5703635	6643036	14399531
3513507	3827683	5763574	6680619	
3514549	3827684	5956342	6689894	
3515057	3880896	6115200	6708731	
3521194	3881117	6116625	6751554	

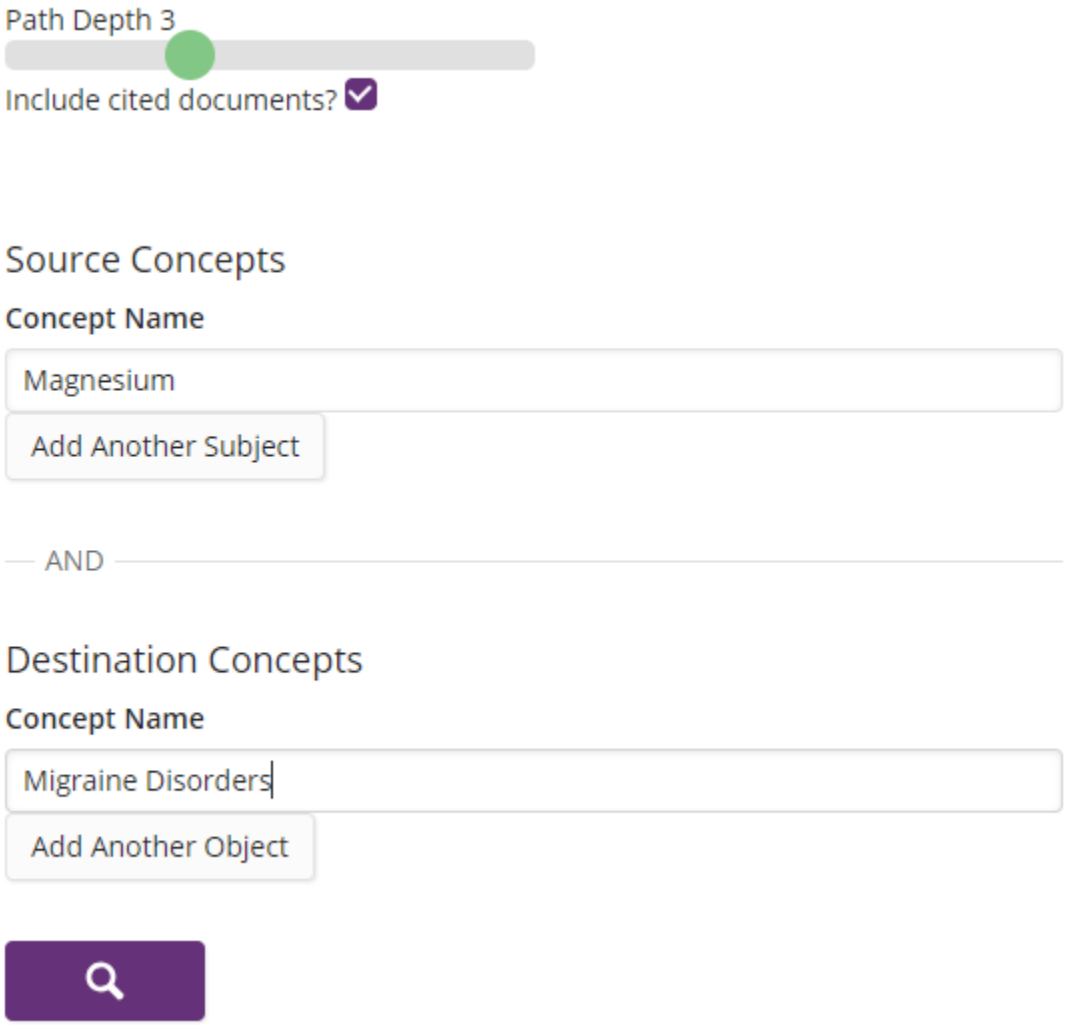


Figure B-1. Magnesium – Migraine Disorders search parameters

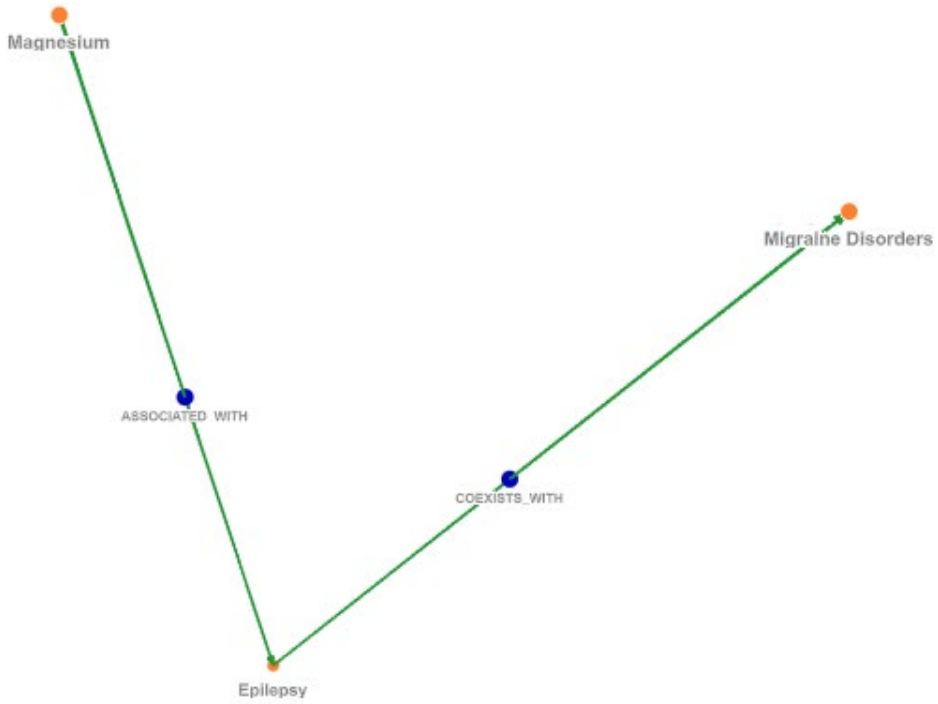


Figure B-2. Magnesium – Migraine Disorders result 1 (with citations)

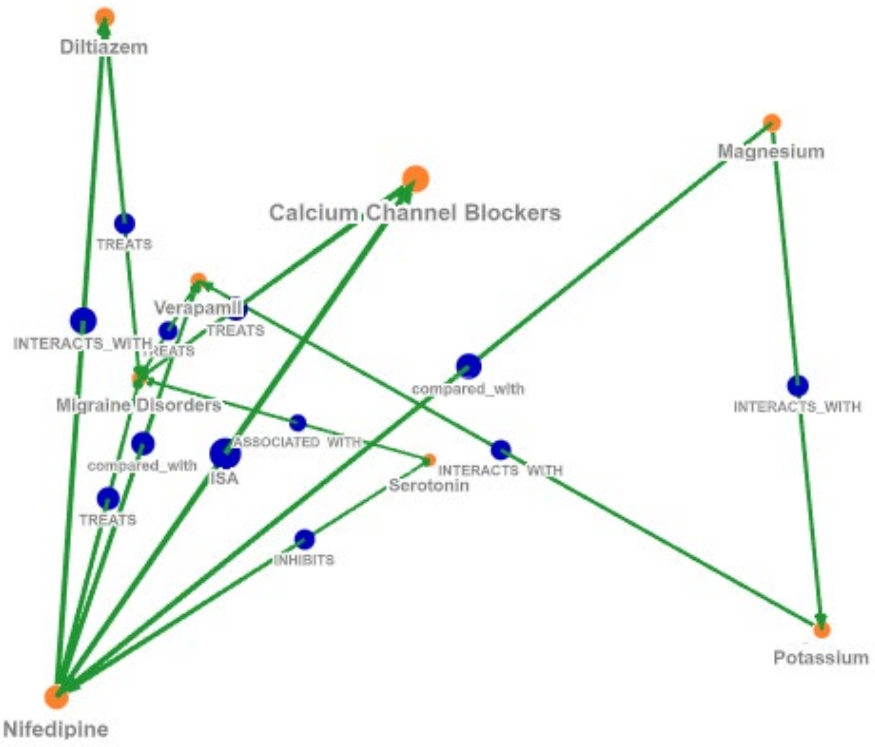


Figure B-3. Magnesium – Migraine Disorders result 2 (with citations)

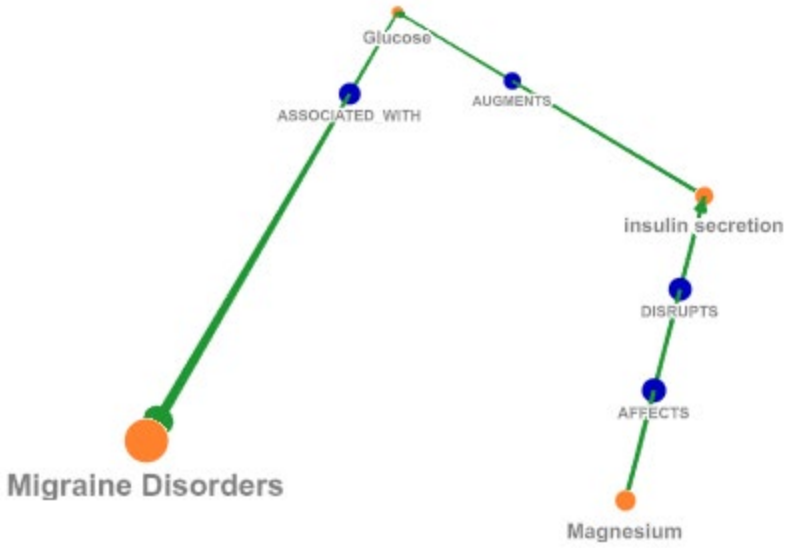


Figure B-4. Magnesium – Migraine Disorders result 3 (with citations)

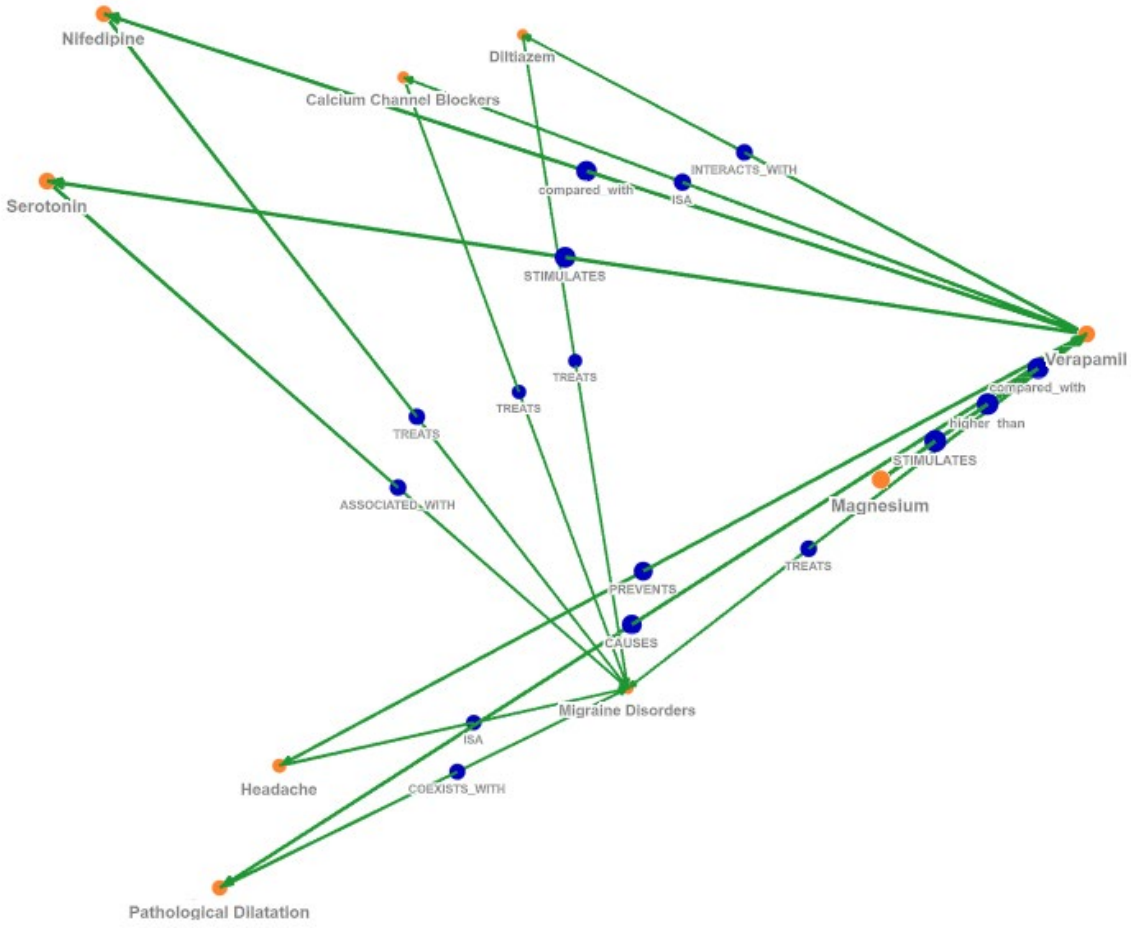


Figure B-5. Magnesium – Migraine Disorders result 4 (with citations)

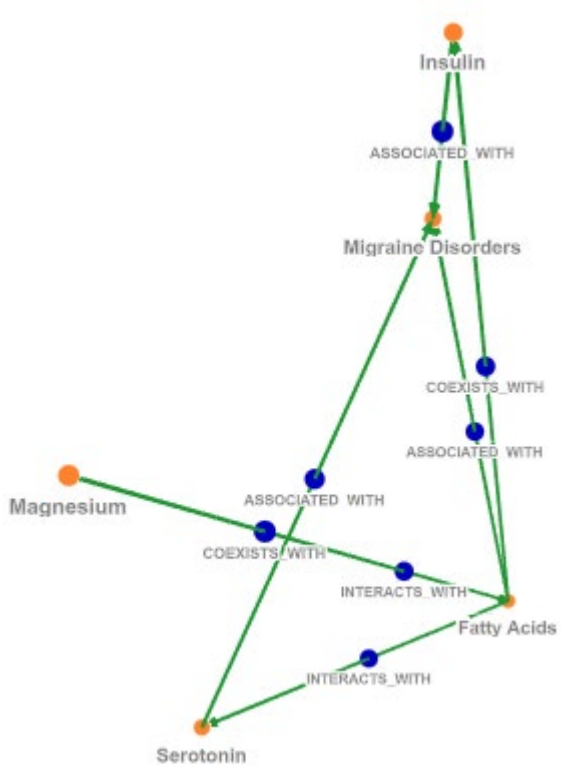


Figure B-6. Magnesium – Migraine Disorders result 5 (with citations)

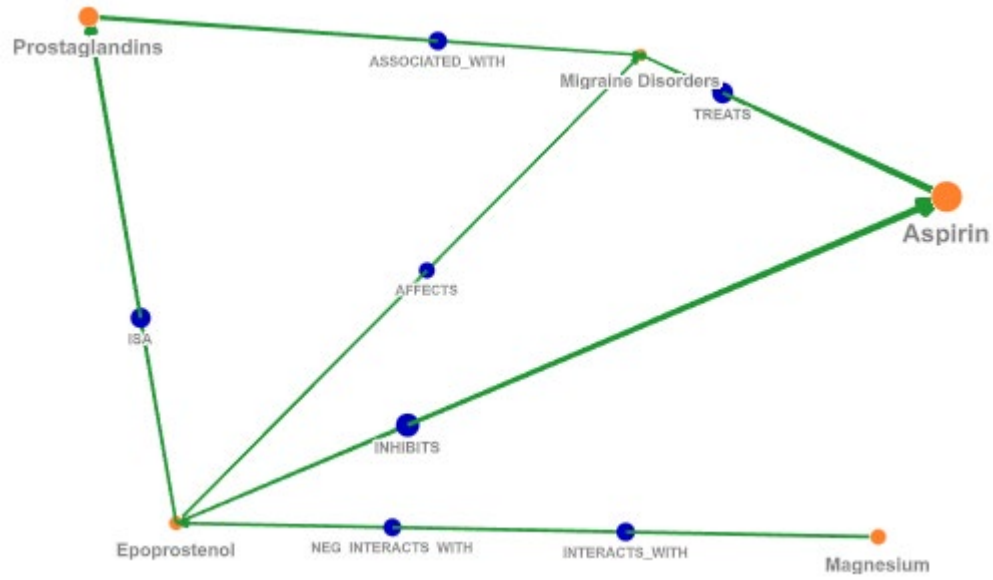


Figure B-7. Magnesium – Migraine Disorders result 6 (with citations)

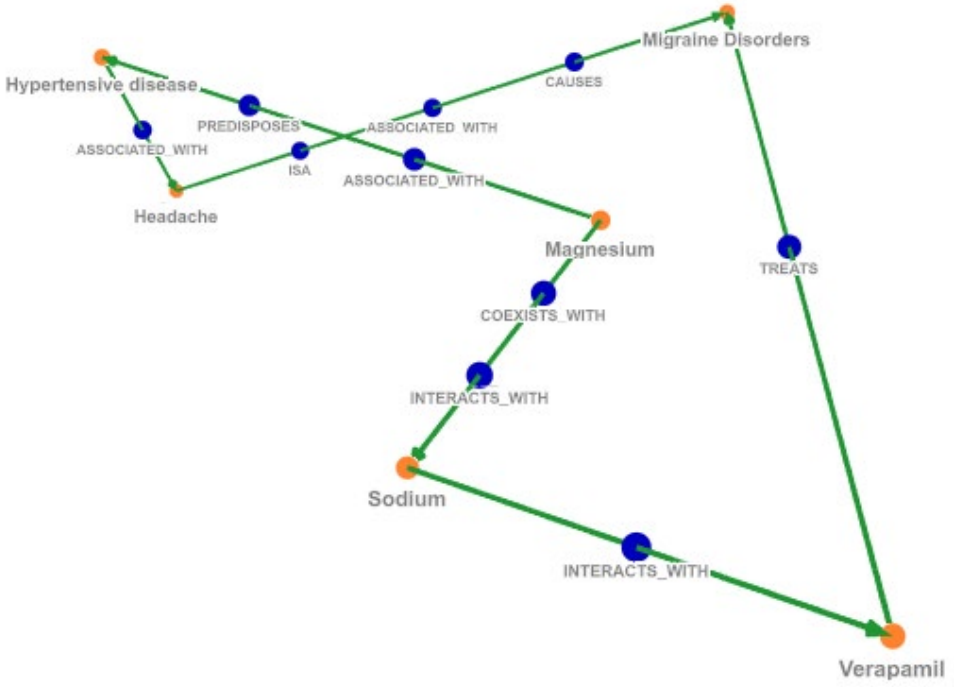


Figure B-8. Magnesium – Migraine Disorders result 7 (with citations)

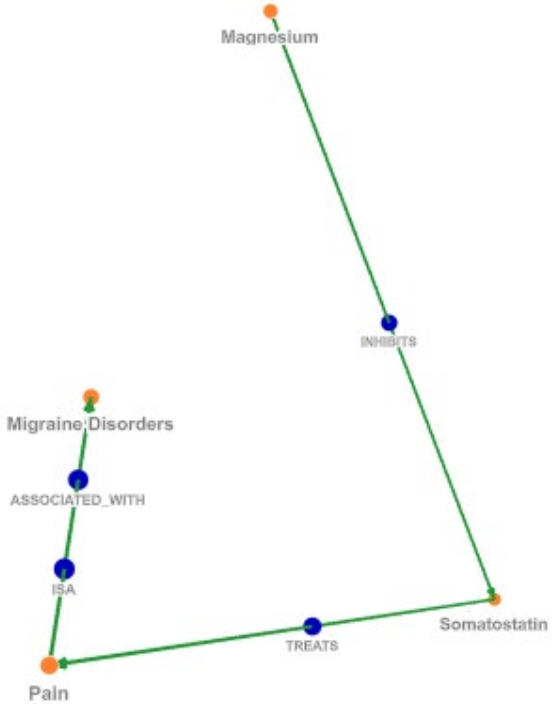


Figure B-9. Magnesium – Migraine Disorders result 8 (with citations)

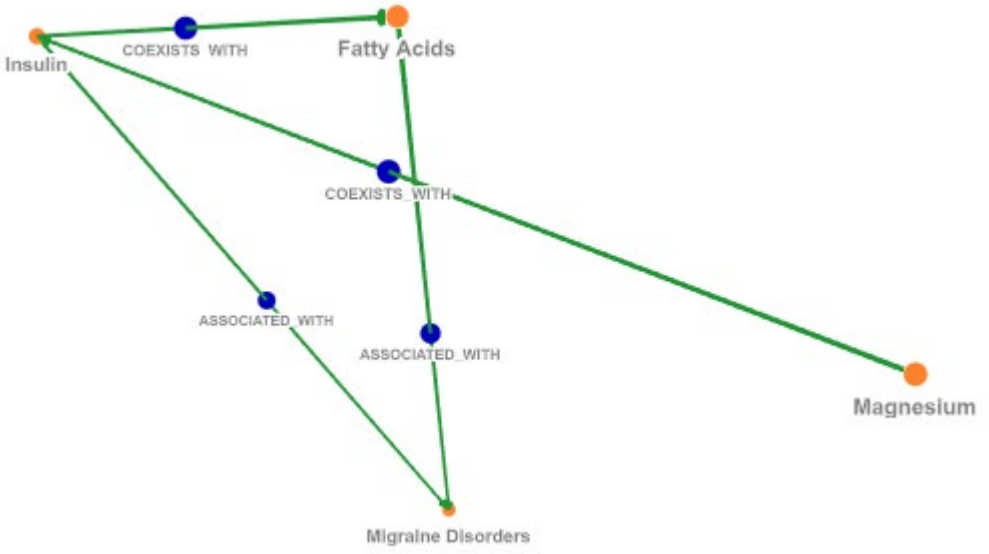


Figure B-10. Magnesium – Migraine Disorders result 9 (with citations)

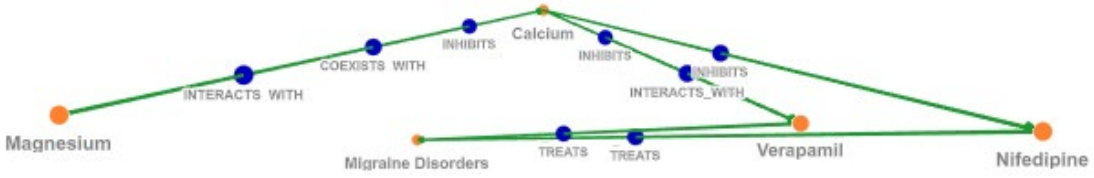


Figure B-11. Magnesium – Migraine Disorders result 10 (with citations)

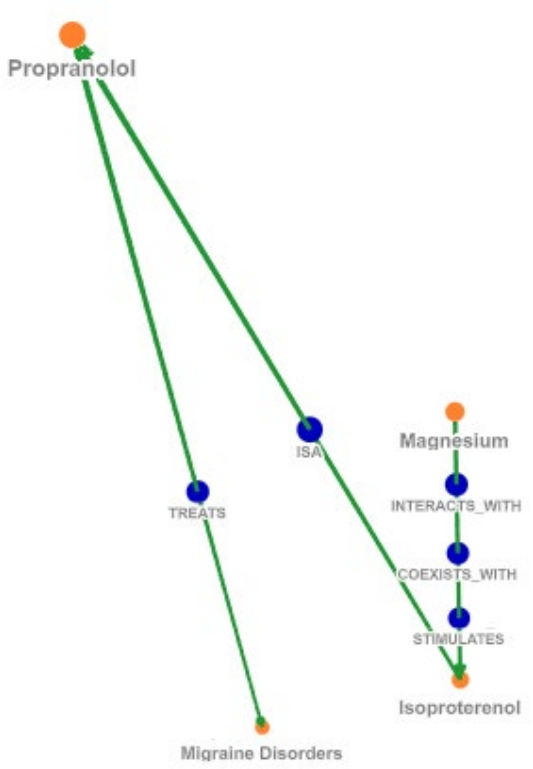


Figure B-12. Magnesium – Migraine Disorders result 11 (with citations)

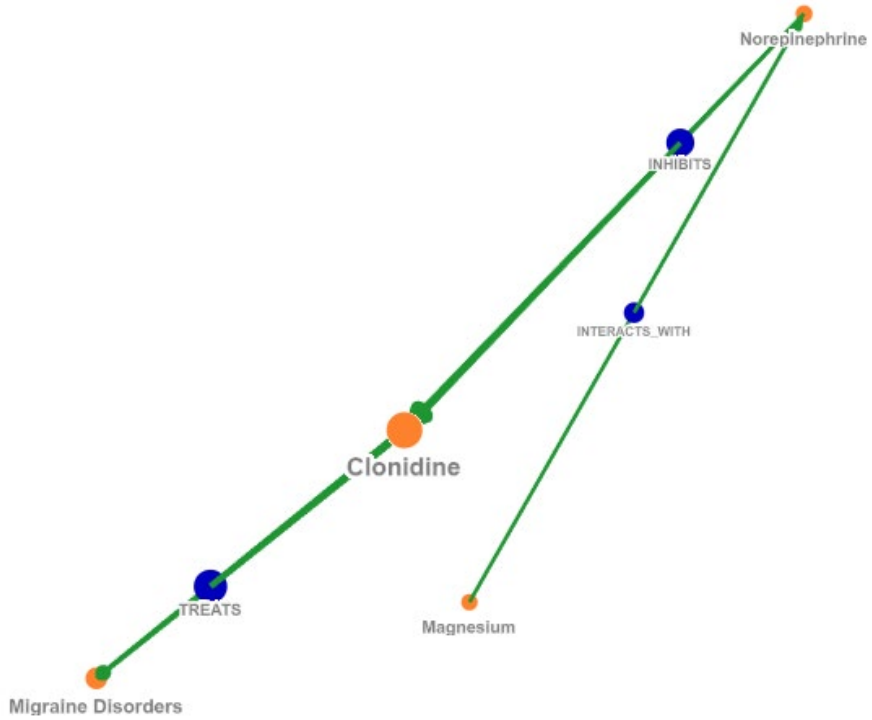


Figure B-13. Magnesium – Migraine Disorders result 12 (with citations)

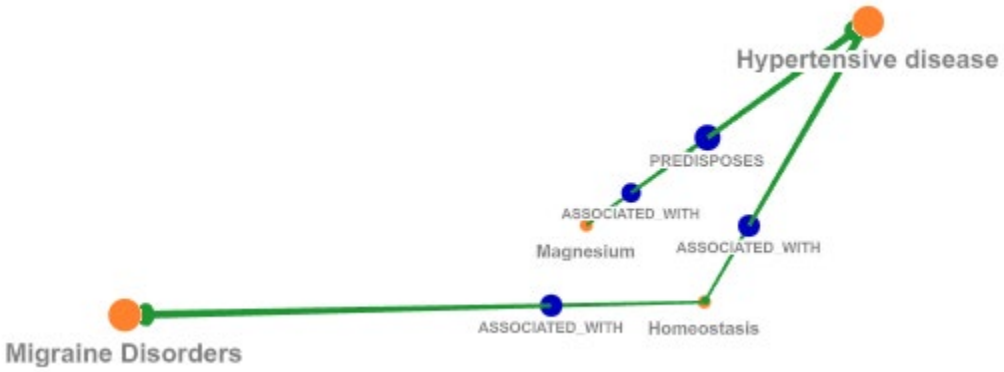


Figure B-14. Magnesium – Migraine Disorders result 13 (with citations)

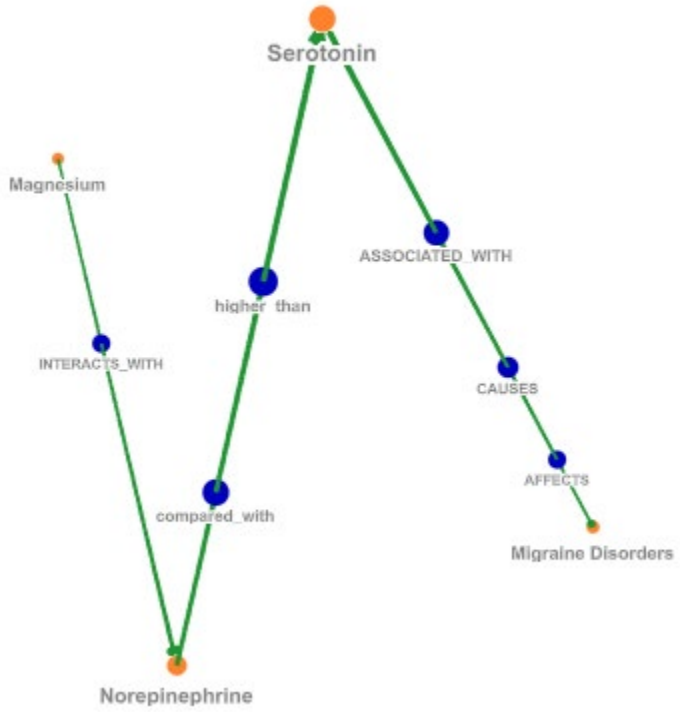


Figure B-15. Magnesium – Migraine Disorders result 14 (with citations)

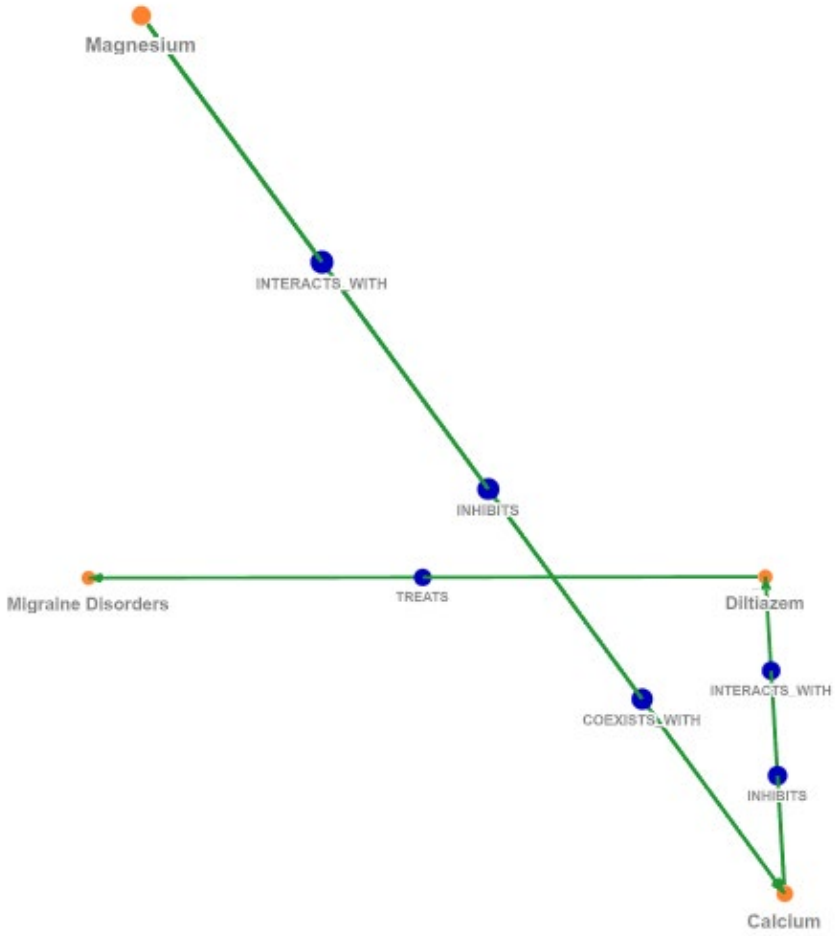


Figure B-16. Magnesium – Migraine Disorders result 15 (with citations)

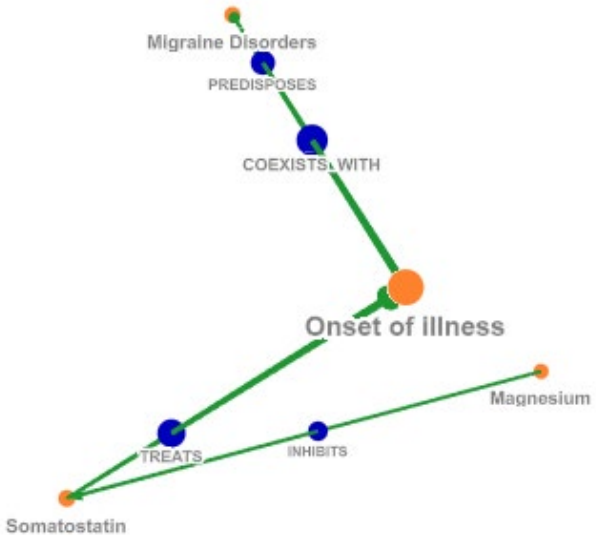


Figure B-17. Magnesium – Migraine Disorders result 16 (with citations)

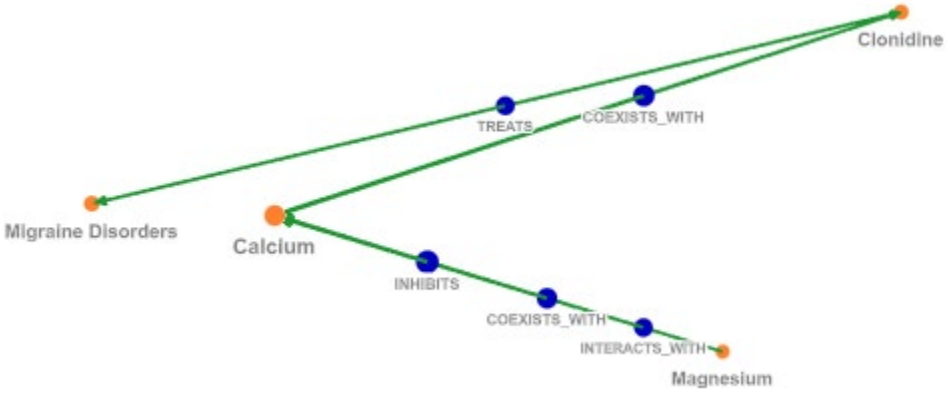


Figure B-18. Magnesium – Migraine Disorders result 17 (with citations)

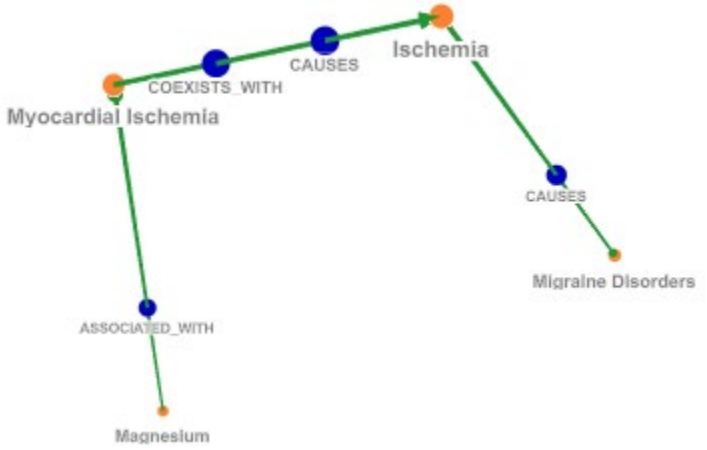


Figure B-19. Magnesium – Migraine Disorders result 18 (with citations)

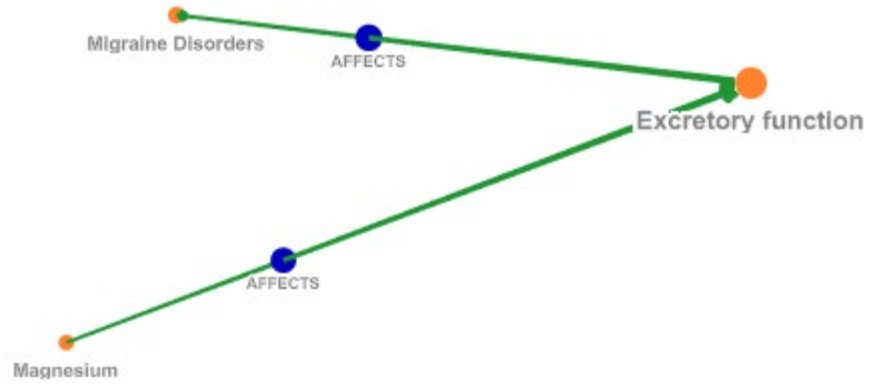


Figure B-20. Magnesium – Migraine Disorders result 1 (no citations)

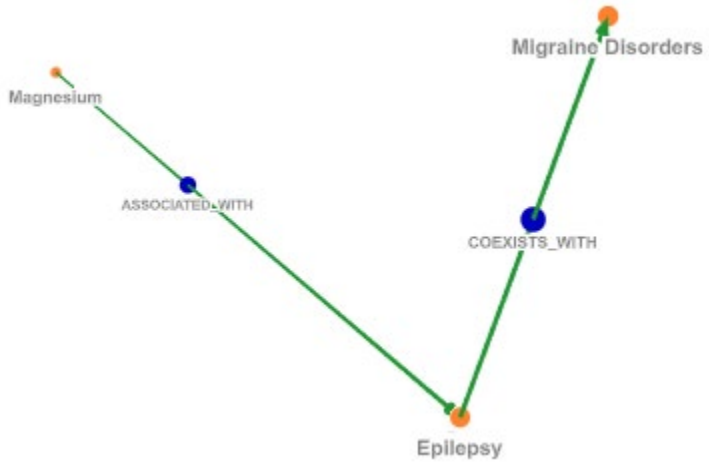


Figure B-21. Magnesium – Migraine Disorders result 2 (no citations)

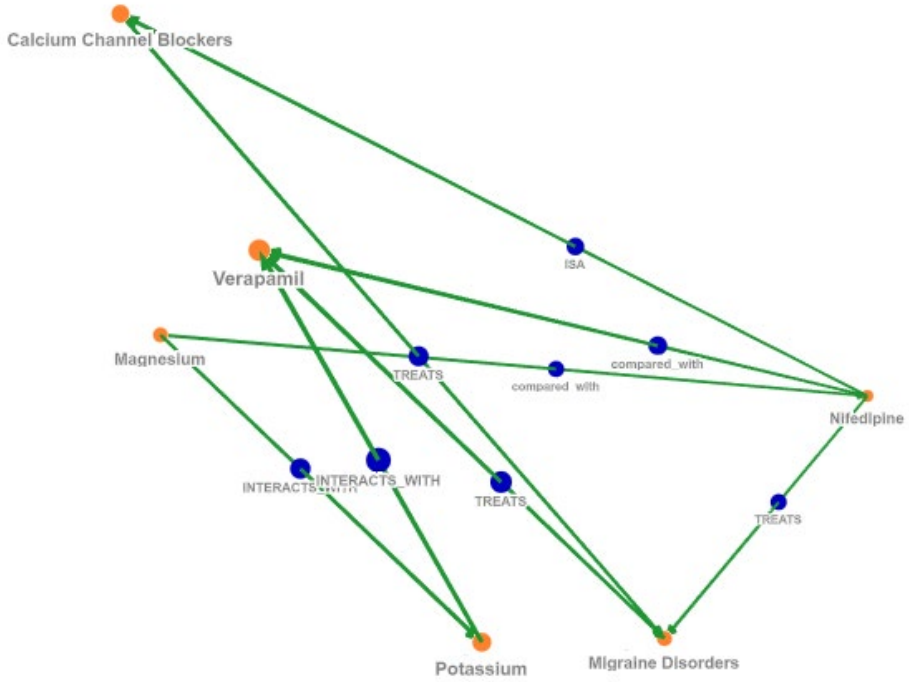


Figure B-22. Magnesium – Migraine Disorders result 3 (no citations)

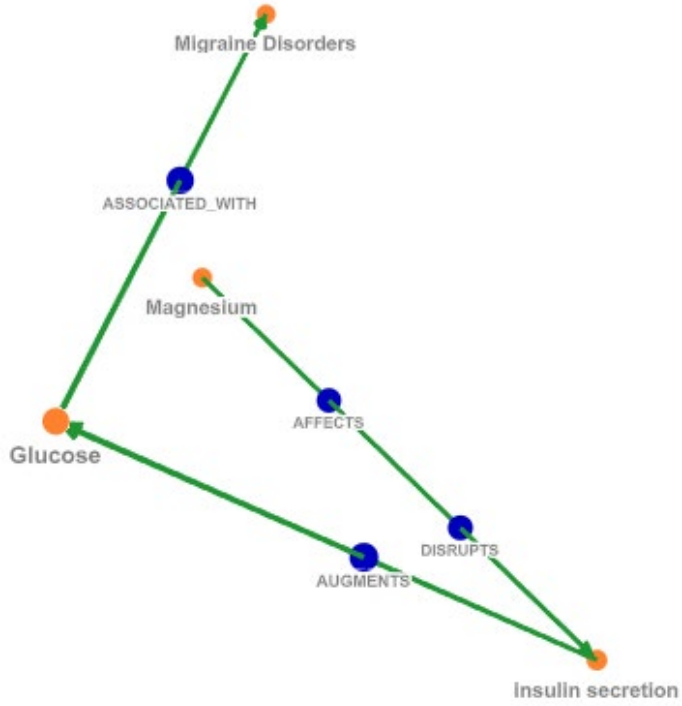


Figure B-23. Magnesium – Migraine Disorders result 4 (no citations)

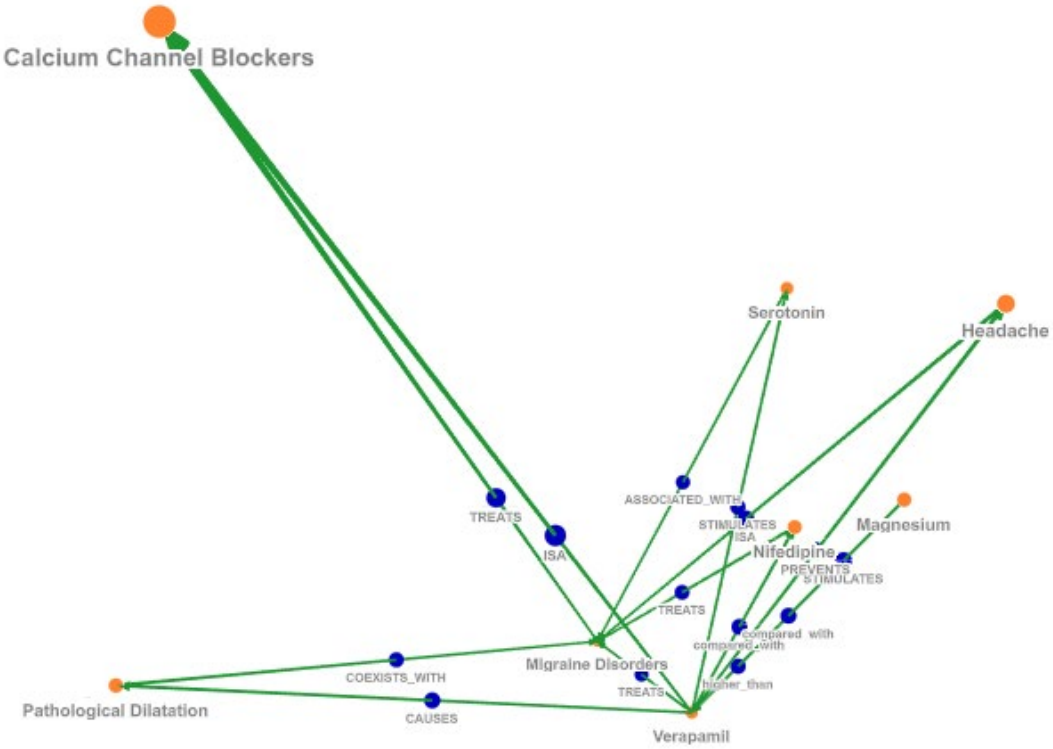


Figure B-24. Magnesium – Migraine Disorders result 5 (no citations)

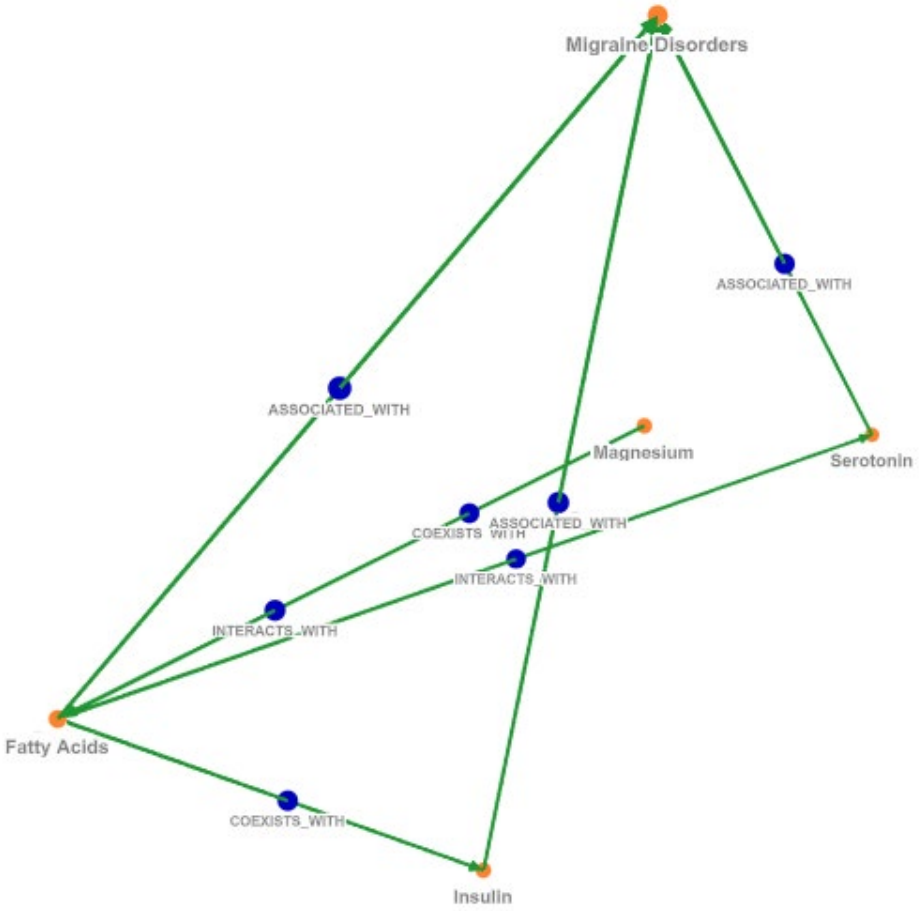


Figure B-25. Magnesium – Migraine Disorders result 6 (no citations)

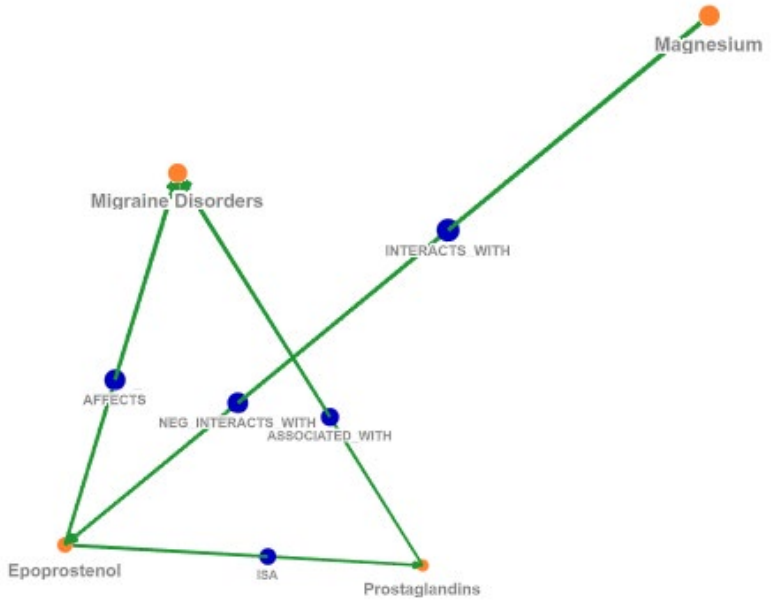


Figure B-26. Magnesium – Migraine Disorders result 7 (no citations)

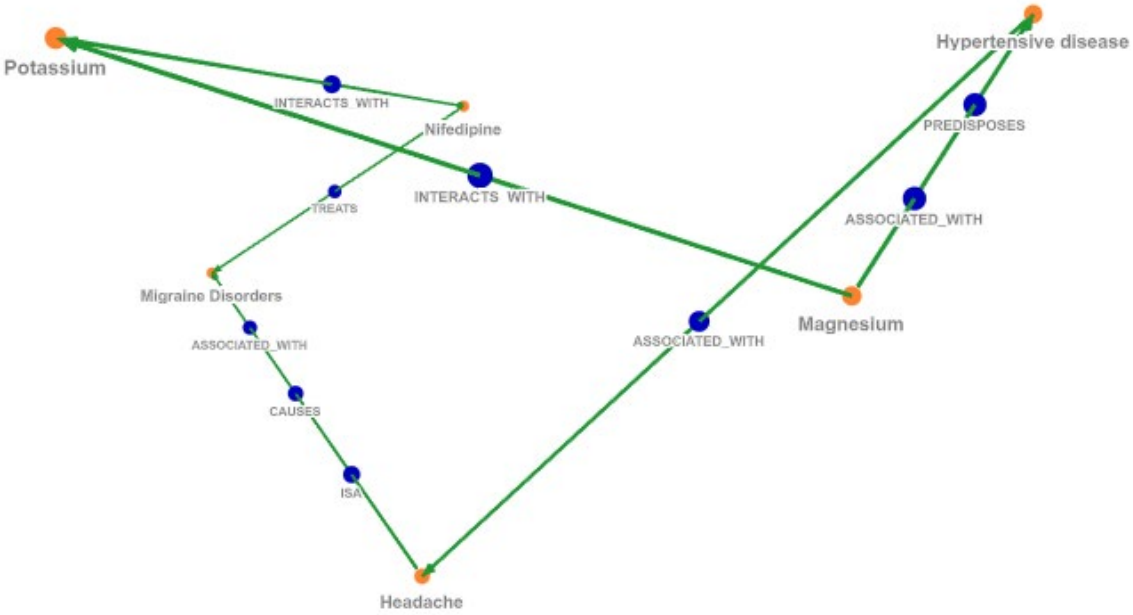


Figure B-27. Magnesium – Migraine Disorders result 8 (no citations)

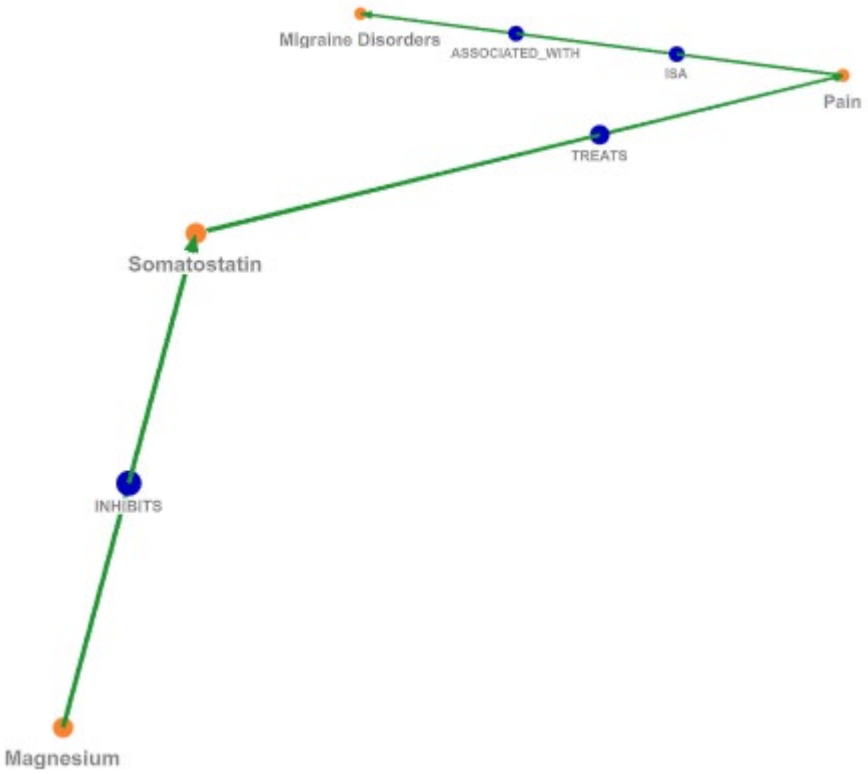


Figure B-28. Magnesium – Migraine Disorders result 9 (no citations)

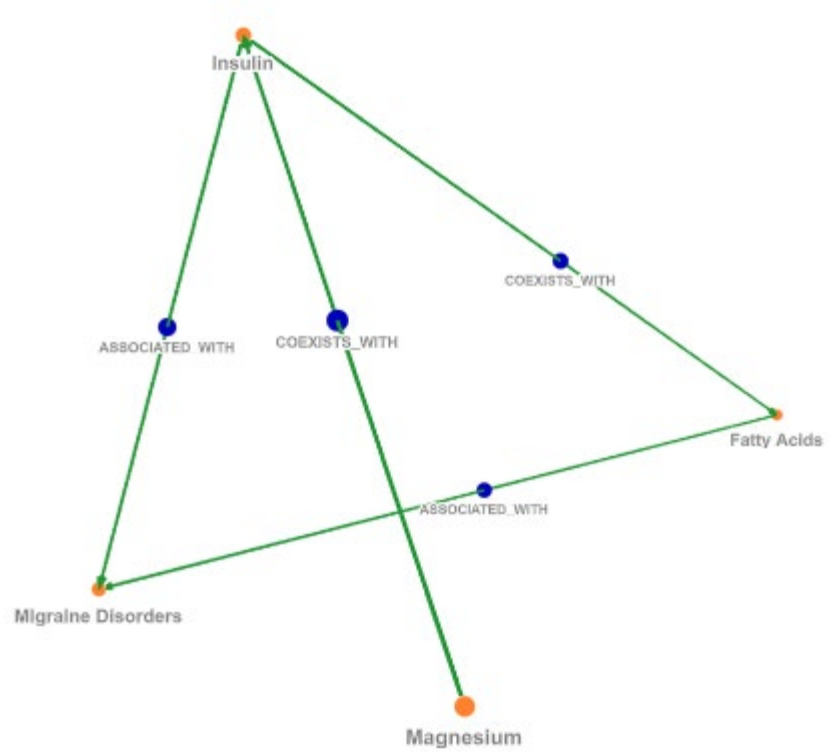


Figure B-29. Magnesium – Migraine Disorders result 10 (no citations)

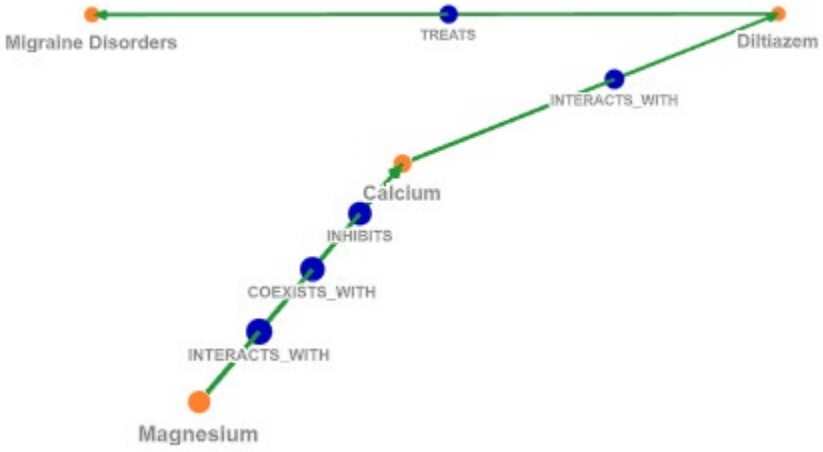


Figure B-30. Magnesium – Migraine Disorders result 11 (no citations)

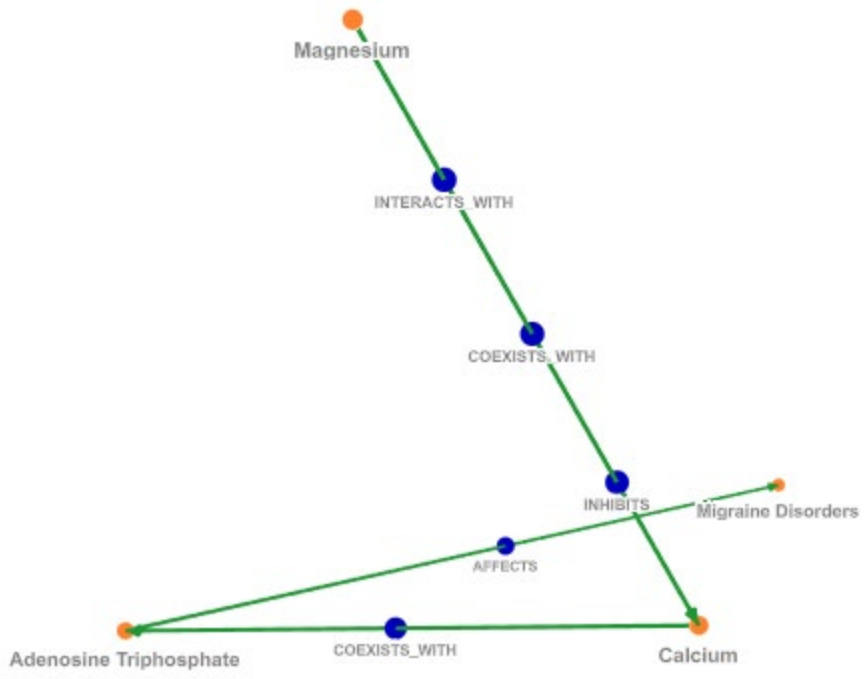


Figure B-31. Magnesium – Migraine Disorders result 12 (no citations)

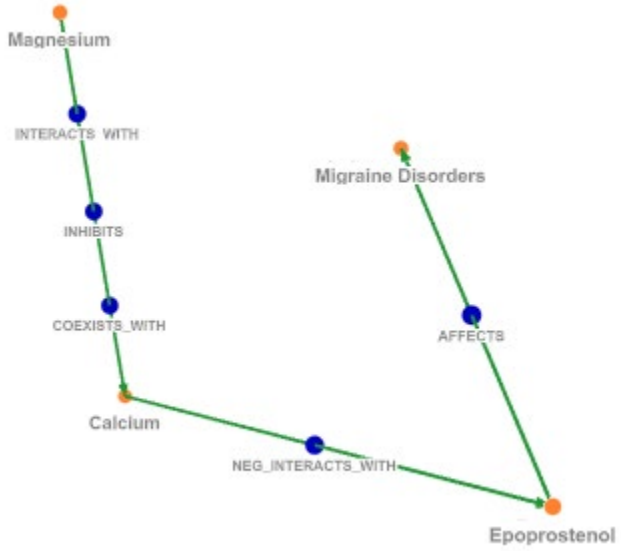


Figure B-32. Magnesium – Migraine Disorders result 13 (no citations)

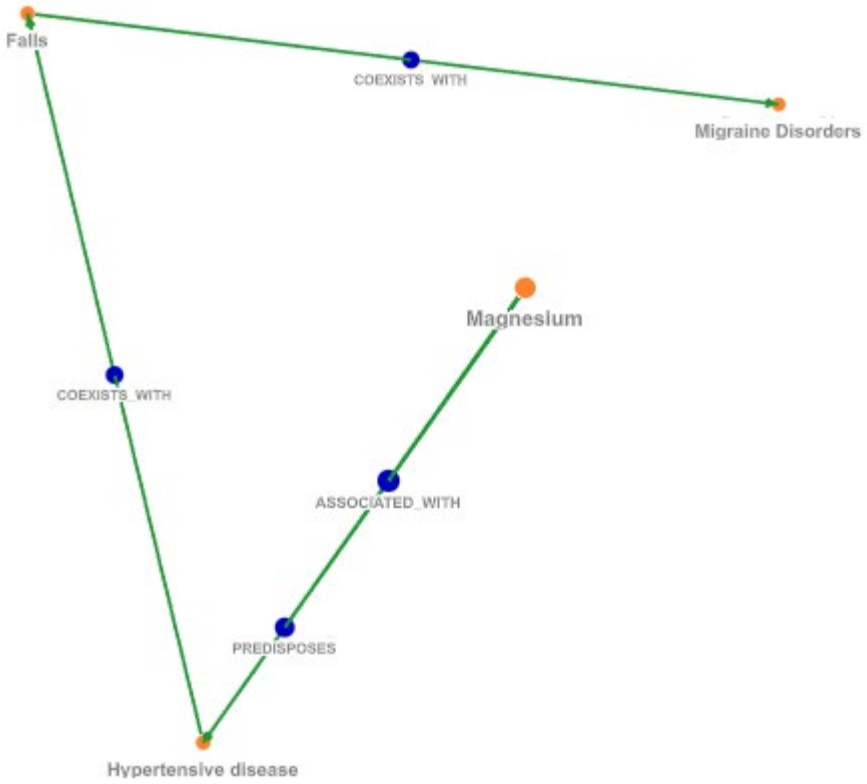


Figure B-33. Magnesium – Migraine Disorders result 14 (no citations)

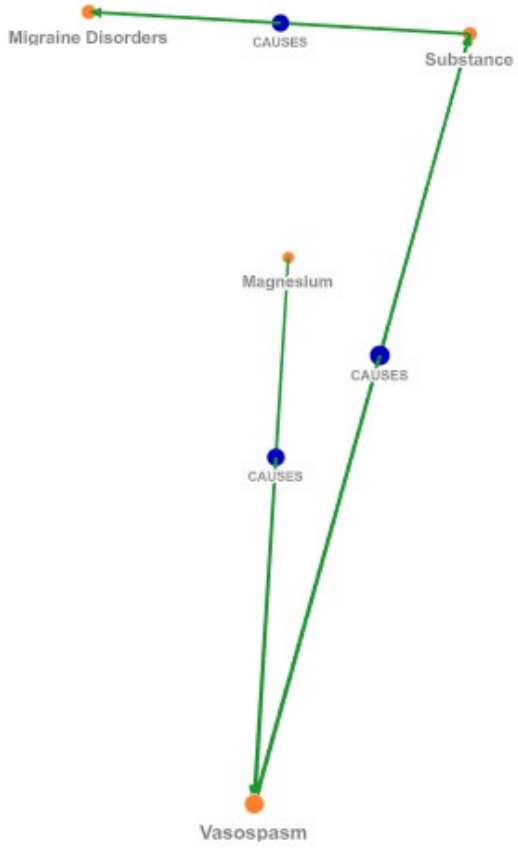


Figure B-34. Magnesium – Migraine Disorders result 15 (no citations)

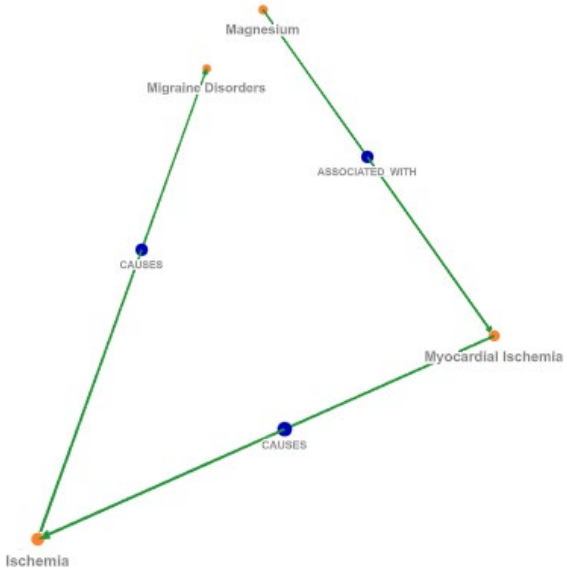


Figure B-35. Magnesium – Migraine Disorders result 16 (no citations)

Table B-2. Magnesium – Migraine Disorders generated predicates

[PMID108361]
Title "Cerebrospinal fluid concentrations of magnesium and inorganic phophate in epilepsy."
[C0024467|Magnesium – ASSOCIATED_WITH – C0014544|Epilepsy]
[PMID1138583]
Title "Increased platelet aggregation in early diabetus mellitus"
SemRep did not derive a predicate from this statement, but it is worth noting.
Abstract "In view of the tendency toward vascular disease in diabetes mellitus, we studied platelet aggregation in 15 normal, 7 prediabetic, 12 latent, and 20 frankly diabetic subjects."
[C0042373|Vascular Diseases – COEXISTS_WITH – C0011849|Diabetes Mellitus] [C0042373|Vascular Diseases – COEXISTS_WITH – C0241863|Diabetic]
Abstract "We conclude that platelet aggregation may be increased early in diabetes mellitus and may be involved in the genesis of diabetic microangiopathy."
[C0011849|Diabetes Mellitus – AFFECTS – C0025945|Microangiopathy, Diabetic]
[PMID1146499]
Title "The significance of blood serotonin levels in migraine."
[C1445983|Blood serotonin measurement – DIAGNOSES – C0149931|Migraine Disorders]
Abstract "If serotonin is effective in relieving migraine pain, this is probably due to extracellular serotonin acting on the cardiovascular system."

[C0036751|Serotonin – ASSOCIATED_WITH – C0030193|Pain] [C0036751|Serotonin – ASSOCIATED_WITH(SPEC) – C0149931|Migraine Disorders]
[PMID1153064]
Abstract "EEG changes suggest–according to the author–that migraine is a primary cerebral and only secondarily a vascular disorder."
[C0149931|Migraine Disorders – ISA – C0042373|Vascular Diseases]
[PMID1254466]
Title "Serotonin (5HT) in migraine: levels in whole blood in and between attacks."
[C0036751|Serotonin – ASSOCIATED_WITH – C0149931|Migraine Disorders]
[PMID1262918]
Title "Cerebral arterial spasm. Part 4: in vitro effects of temperature, serotonin analogues, large nonphysiological concentrations of serotonin, and extracellular calcium and magnesium on serotonin-induced contractions of the canine basilar artery."
[C0006675|Calcium – AFFECTS – C1140999|Contraction] [C0024467|Magnesium – AFFECTS – C1140999|Contraction] [C0036751|Serotonin – AFFECTS – C1140999|Contraction]
Abstract "Extracellular magnesium, in contrast, was shown to inhibit serotonin-induced contractions."
[C0024467|Magnesium – DISRUPTS – C1140999|Contraction]
[PMID13214278]
Introductory Paragraph "Among this group are headaches associated with hypertension, arteriosclerosis, infection, brain tumor, hematoma, allergy, and cranial trauma."

[C0018681|Headache – ASSOCIATED_WITH – C0020538|Hypertensive disease]
[PMID14338128]
Title "DISORDERS OF MAGNESIUM METABOLISM IN EPILEPSY."
[C0024467|Magnesium – ASSOCIATED_WITH – C0014544|Epilepsy]
Introductory Paragraph "The possibility that convulsions may occur in Mg deficiency led some investigators to study the metabolism of this metal in the epilepsies, and a trend to low blood concentrations was usually found."
[C0009951|Convulsions – OCCURS_IN – C0024473|Magnesium Deficiency]
[PMID184066]
Abstract – Taken from publisher's site. "Since biochemical abnormalities reported herein were common to occlusive CVD and migraine headache, it seems probable that they are due to ischemia associated with both conditions and possibly related to the resultant disorder of cerebral energy metabolism."
[C0022116|Ischemia – CAUSES – C0149931|Migraine Disorders]
[PMID20784580]
Title "The Pain Pathways in Migraine"
[C0030193|Pain – ASSOCIATED_WITH – C0149931|Migraine Disorders]
[PMID2424267]
Abstract "The cardiovascular effects of different calcium channel blockers (CCB), exemplified by nifedipine, verapamil and diltiazem, are not identical."
[C0028066|Nifedipine – INTERACTS_WITH – C0006684|Calcium Channel Blockers]

[C0042523|Verapamil – INTERACTS_WITH – C0006684|Calcium Channel Blockers]
[PMID2425960]
Abstract "Three such drugs, nifedipine, verapamil, and diltiazem, are currently available in the United States, although none are specifically approved for use in migraine."
[C0028066|Nifedipine – TREATS – C0149931|Migraine Disorders] [C0042523|Verapamil – TREATS – C0149931|Migraine Disorders]
Abstract "Nifedipine and nimodipine also appear to be valuable for the treatment of cluster headache."
[C0028066|Nifedipine – TREATS – C0009088|Cluster Headache] [C0028094|Nimodipine – TREATS – C0009088|Cluster Headache]
Abstract "Two case reports describing favorable responses to flunarizine in childhood hemiplegic migraine are the only available data concerning the utility of these drugs in "complicated" migraine syndromes."
[C0016295|Flunarizine – TREATS – C0270862|Hemiplegic migraine]
[PMID2433912]
Abstract "Opiates and somatostatin inhibit the release of substance P from primary sensory neurones and relieve both pain and autonomic symptoms of cluster headache attack."
[C0037659|Somatostatin – TREATS – C0030193|Pain]
Abstract "Plasma substance P-like immunoreactivity was decreased during spontaneous attack of cluster headache and migraine and during histamine precipitated attack of cluster headache. Taken together these data suggest that substance P and endogenous opioids could be implicated in the pathophysiology of cluster headache and migraine."
[PMID2452018]
Title "Substance P in the human iris: possible involvement in echothiophate-induced miosis

in cluster headache.”
[C0038585|Substance P – CAUSES – C0009088|Cluster Headache]
[PMID2981405]
Abstract – Taken from publisher’s site. “Verapamil is a calcium-channel blocking agent that has been used extensively in the treatment of coronary vasospasm and supraventricular arrhythmias.”
[C0042523|Verapamil – ISA – C0006684|Calcium Channel Blockers]
[PMID3458981]
Title “Vasomotor effects of magnesium: a comparison with nifedipine and verapamil of in vitro reactivity in feline cerebral and peripheral arteries.”
[C0024467|Magnesium – compared_with – C0028066|Nifedipine]
Abstract “The order of potency for eliciting relaxation was: nifedipine greater than verapamil greater than magnesium.”
[C0042523|Verapamil – compared_with – C0024467|Magnesium]
[PMID3521194]
Title “Migraine treatment with calcium channel blockers.”
[C0006684|Calcium Channel Blockers – TREATS – C0149931|Migraine Disorders]
[PMID3585352]
Title Laterality of pain in migraine distinguished by interictal rates of habituation of electrodermal responses to visual and auditory stimuli.

[C0030193|Pain – ASSOCIATED_WITH – C0149931|Migraine Disorders]
[PMID361756]
Abstract Prostacyclin (PGI(2)) is an unstable prostaglandin which inhibits platelet aggregation and serotonin release and causes vasodilation.
[C0033554|Prostaglandins – DISRUPTS – C0032176|Platelet aggregation]
[PMID3629724]
Abstract "A stimulating effect of serotonin, guanine nucleotides and sodium fluoride is found as well as a dependence of the catalytic activity of adenylate cyclase on magnesium and manganese ions."
[C0001492|Adenylate Cyclase – INTERACTS_WITH – C0024467|Magnesium]
[PMID372242]
Abstract Endothelial cells synthesize prostacyclin (PGI(2)), an unstable prostaglandin that inhibits platelet aggregation and serotonin release.
[C0033554|Prostaglandins – DISRUPTS – C0032176|Platelet aggregation]
[PMID379918]
Abstract "However, in 1 X 10(-5)M concentrations PGI2 contracted the arterial muscle and did not antagonize contractions induced by serotonin or PGF2 alpha."
[C0012471|Dinoprost – CAUSES – C1140999|Contraction] [C0036751|Serotonin – CAUSES – C1140999|Contraction]
[PMID386807]
Abstract – Taken from publisher's site. "Substantial species variability has been described for the effects of extracellular

magnesium upon tension development, which arises out of striking calcium-magnesium antagonism at an extracellular site in the rat and guinea pig.”
[C0024467|Magnesium – AFFECTS – C0233494|Tension]
[PMID3898850]
Abstract – Introductory paragraph obtained from publisher’s site and used. “Prostacyclin is the most potent naturally occurring vasodilator and inhibitor of platelet aggregation.”
[C0033567|Epoprostenol – DISRUPTS – C0032176|Platelet aggregation]
Abstract – Introductory paragraph obtained from publisher’s site and used. “Magnesium is known to exert an inhibitory effect on coagulation and platelet function, but, so far, little is known about its influence on prostacyclin metabolism. Therefore the effect of magnesium sulfate on vascular prostacyclin synthesis and platelet prostacyclin interaction was investigated.”
[C0024467|Magnesium – INTERACTS_WITH – C0033567|Epoprostenol]
[PMID390330]
Abstract “Other direct studies, from our laboratory, indicate that [Mg2+]o regulates calcium exchange and content of vascular smooth muscle.”
SemRep did not derive a predicate from this statement, but it is worth noting.
Abstract “In summary, the concept to be presented suggests that a deficiency in dietary Mg2+ is a key factor in the high incidence of mortality noted in SDIHD in nations of the Western world.”
SemRep did not derive a predicate from this statement, but it is worth noting.
Abstract ”The hypomagnesemia produces progressive vasoconstriction, vasospasm and ischemia, which, given time, will lead to SDIHD.”
[C0022116|Ischemia – CAUSES – C0151744|Myocardial Ischemia] [C0085616|Vasospasm – CAUSES – C0151744|Myocardial Ischemia]

[PMID4016946]
Title “Relevance of prostaglandins in migraine.”
[C0033554|Prostaglandins – ASSOCIATED_WITH – C0149931|Migraine Disorders]
Abstract “Prostaglandins (PG), particularly PGE, may be linked to the pathophysiology of migraine in several important ways. PGE1 may “simulate” a migraine attack in healthy volunteers. In animal experiments and in human infusions, PGEs cause vasodilation and hyperalgesia, both typical reactions of inflammation.”
[C0033559|Prostaglandins E – ISA – C0033554|Prostaglandins]
Abstract “The view that vascular headache is an “inflammatory reaction” allows the best concept concerning the local role of PGs and the effectiveness of PG-inhibitors in the treatment of migraine.
[C0033529|Prostaglandin Antagonists – TREATS – C0149931|Migraine Disorders]
[PMID4260015]
Title “Magnesium metabolism from the viewpoint of cardioangiology. II. Findings in magnesium metabolism in vascular diseases”
[C0024467|Magnesium – ASSOCIATED_WITH – C0042373|Vascular Diseases]
[PMID450213]
Abstract “We performed in vitro experiments with a small volume chamber to determine the inhibitory effect of nifedipine on serotonin-, phenylephrine-, and potassium-induced contractions of canine basilar and femoral arteries.”
[C0028066|Nifedipine – INHIBITS – C0036751|Serotonin]
Abstract “In contrast, nifedipine did not significantly inhibit the serotonin- and phenylephrine-

induced contractions of the femoral artery but did inhibit potassium-induced contractions of the femoral artery. Calcium-induced contractions of the basilar artery were also inhibited by nifedipine.”
SemRep did not derive a predicate from this statement, but it is worth noting.
[PMID4760464]
Abstract – Taken from publisher’s site. “Reserpine pretreatment did not alter the effect of magnesium on contractility to those agents both in polarizing and calcium-free depolarizing solutions.”
[C0024467|Magnesium – NEG_AFFECTS – C1140999|Contraction]
Abstract – Taken from publisher’s site. “However, such pretreatment modified the effect of magnesium on sensitivity of aortas to acetylcholine and norepinephrine.”
[C0024467|Magnesium – INTERACTS_WITH – C0001041|Acetylcholine] [C0024467|Magnesium – INTERACTS_WITH – C0028351|Norepinephrine]
Abstract – Taken from publisher’s site. “These results suggest that magnesium and reserpine act to alter calcium binding, and its availability to the contractile apparatus, by different mechanisms.”
[C0024467|Magnesium – AFFECTS – C1516144|Calcium Binding]
[PMID4814375]
Abstract “Methylsergide reversibly blocked the artery’s response to serotonin and caused a contraction of the basilar artery.”
[C0025842|Methysergide – AFFECTS – C0871261|response] [C0036751|Serotonin – CAUSES – C1140999|Contraction]
[PMID4978139]
Introduction – Taken from paper itself. “It is the purpose of this paper to examine the question of the relation of migraine and epilepsy by an analysis of 1,800 cases of migraine seen at the Northcott Neurological Centre since 1951 with particular reference to those cases in which migraine and epilepsy were manifested in the same individual.”

[C0014544|Epilepsy – COEXISTS_WITH – C0149931|Migraine Disorders]
[PMID527222]
Title "Hypomagnesaemia in diabetes."
[C0151723|Hypomagnesemia – COEXISTS_WITH – C0011847|Diabetes]
Abstract "Although its significance is unclear, hypomagnesaemia could conceivably predispose to ischaemic heart disease in diabetes."
[C0151723|Hypomagnesemia – PREDISPOSES – C0151744|Myocardial Ischemia]
[PMID5297855]
Title "Plasma serotonin in migraine and stress."
[C0036751|Serotonin – ASSOCIATED_WITH – C0149931|Migraine Disorders]
Abstract – Taken from publisher's site/paper "Interest in the possible relationship of serotonin to the migraine syndrome had quickened following Sicuteri's preliminary trial of methysergide, a serotonin "antagonist," in the treatment of migraine."
[C0025842|Methysergide – TREATS(SPEC) – C0149931|Migraine Disorders] [C0036753|Serotonin Antagonists – TREATS – C0149931|Migraine Disorders]
Abstract – Taken from publisher's site/paper "Finally, total plasma serotonin was investigated after the intramuscular injection of reserpine 2.5 mg, which has been reported to induce migraine headache in susceptible patients."
[C0036751|Serotonin – CAUSES – C0149931|Migraine Disorders]
Abstract – Taken from publisher's site/paper "The effect of injected serotonin on spontaneous and reserpine-induced headaches was noted and correlated with serotonin levels."

[C0036751|Serotonin – ASSOCIATED_WITH – C0018681|Headache]
[PMID5483751]
Title “Influence of magnesium and cysteine on vasopressin-induced contractions in various canine blood vessels.”
[C0024467|Magnesium – AUGMENTS – C0005847|Blood Vessels]
[PMID5658323]
Title “Plasma serotonin levels in migraine.”
[C0036751|Serotonin – ASSOCIATED_WITH – C0149931|Migraine Disorders]
[PMID5674982]
Title “Magnesium-calcium antagonism in the contraction of arterioles”
[C0006684|Calcium Channel Blockers – TREATS – C1140999|Contraction]
Abstract – Taken from publisher’s site. “Increasing the level of Mg in the blood inhibits the tone and contractility of the resistance vessels in the muscle.”
SemRep did not derive a predicate from this statement, but it is worth noting.
[PMID5682248]
Abstract – Taken from publisher’s site “Diabetes mellitus was the most common condition associated with hypomagnesemia.”
[C0151723|Hypomagnesemia – COEXISTS_WITH – C0011849|Diabetes Mellitus]
[PMID572225]

Abstract “FFA fell by a maximum of 65% or 0.44 meq/liter and Mg fell by a maximum of 0.31 meq/liter during the glucose-insulin infusions.”
[C0015688|Fatty Acids, Nonesterified – COEXISTS_WITH – C0021641|Insulin] [C0024467|Magnesium – COEXISTS_WITH – C0021641|Insulin]
Abstract “The sharp divergent changes in FFA and Mg after cessation of nicotinic acid infusion support the prime role of FFA-effecting movements of Mg and the thesis that FFA bind Mg.”
[C0015688|Fatty Acids, Nonesterified – INTERACTS_WITH – C0024467|Magnesium]
[PMID5803690]
Introduction – Taken from paper itself. “According to the classical theory the first (prodromal) phase in migraine is caused by ischemia within the internal carotid system and the second phase (headache) by a vasodilation, especially within the external carotid system.”
[C0022116|Ischemia – CAUSES – C0149931|Migraine Disorders]
[PMID597797]
Abstract “These findings do not exclude the possibility of distension of the larger intracranial arteries during migraine headache, but the angiographic evidence, however limited, does not support this speculation.”
[C0012359|Pathological Dilatation – COEXISTS_WITH – C0149931|Migraine Disorders]
[PMID6022537]
Title “Serotonin, the carotid body, and cranial vessels in migraine.”
[C0036751|Serotonin – ASSOCIATED_WITH – C0149931|Migraine Disorders]
[PMID6116625]

Abstract – Taken from article on www.researchgate.net "Lowering (or removal) of magnesium ions enhances reactivity of a number of arterial, arteriolar, and venous vessels to several neurohumoral agents (e.g., catecholamines, angiotensin, acetylcholine, serotonin) and K+; relaxations induced by prostaglandins and beta-adrenergic stimulation are often attenuated in the absence of magnesium ions."
[C0033554|Prostaglandins – STIMULATES – C0596876|Magnesium Ions]
[PMID6117893]
Abstract "Prostacyclin is a potent vasodilator and the most potent inhibitor of platelet aggregation so far described."
[C0033567|Epoprostenol – DISRUPTS – C0032176|Platelet aggregation]
Abstract "Additionally, since prostacyclin powerfully inhibits platelet aggregation and promotes their disaggregation, this agent could have an important use in the therapy of conditions in which increased platelet aggregation takes place and in which, perhaps, a prostacyclin deficiency exists."
[C0033567|Epoprostenol – DISRUPTS – C0032176|Platelet aggregation] [C0033567|Epoprostenol – AUGMENTS – C0032176|Platelet aggregation]
[PMID6141690]
Title "Magnesium requirement for somatostatin inhibition of insulin secretion."
[C0024467|Magnesium – INHIBITS – C0037659|Somatostatin]
Abstract "We observed that in the absence of magnesium, somatostatin concentrations of 0.5 and 2.0 ng/ml were without effect on first phase insulin secretion. However, these same somatostatin levels produced 50% or more inhibition of insulin secretion in the presence of magnesium at 0.3 or 0.6 mEq/l."
[C0024467|Magnesium – DISRUPTS – C1256369|insulin secretion]
Abstract "Therefore, magnesium ion is necessary for the full inhibitory effect of somatostatin to

occur.”
[C0596876|Magnesium Ions – INHIBITS – C0037659|Somatostatin]
[PMID6149922]
Abstract “A number of CHD and myocardial ischaemic syndromes such as unstable angina pectoris, sudden death ischaemic heart disease, acute myocardial infarction and ventricular arrhythmias have been associated with losses of myocardial magnesium and potassium.”
[C0024467|Magnesium – ASSOCIATED_WITH – C0151744|Myocardial Ischemia]
Abstract “Mg++ ions are essential for regulation of Na+ and K+ transport across cell membranes, including those found in cardiac and vascular smooth muscle cells. Loss of cellular Mg++ results in loss of critically important phosphagens: MgATP and creatine phosphate. Thus, under conditions where cellular Mg++ is depleted (e.g. hypoxia, ischaemia, anoxia), the Na+-K+ pump and phosphagen stores will be compromised, leading to alterations in resting membrane potentials.”
N/A
Abstract “Considerable evidence has accumulated to indicate that the extracellular concentration of Mg++ is important in control of arterial tone and blood pressure via pressure via regulation of vascular membrane Mg++-Ca++ exchange sites. A reduction in the extracellular Mg++ concentration can produce hypertension, coronary vasospasm and potentiation of vasoconstrictor agents by allowing excess entry of Ca++; concomitantly, the potency of vasodilator agents is reduced.”
N/A
[PMID6171999]
Abstract “Indirect evidence has abundantly been presented to support the view that substance P (SP) is involved in the vasodilatation following activation of fine calibre pain fibres (Lembeck & Holzer 1979).”

[C0038585|Substance P – ASSOCIATED_WITH – C0595862|Vasodilation disorder]
Abstract “We have also determined whether (D-Pro2, D-Phe7, D-Trp9)-SP specifically blocks the vascular effects of SP.”
[C0038585|Substance P – AFFECTS – C0005847|Blood Vessels]
[PMID6184468]
Title “The effects of substance P on histamine and 5-hydroxytryptamine release in the rat.”
[C0038585|Substance P – INTERACTS_WITH – C0019588|Histamine]
Abstract “Substance P (SP) induces histamine release from isolated rat peritoneal mast cells at concentrations of 0.1-10 muM.
[C0038585|Substance P – CAUSES – C0019595|Histamine Release]
Abstract “Extracellular calcium (0.1-1 mM), magnesium (1-10 mM) and cobalt (0.01-0.1 mM) all inhibit SP-induced histamine release when added before the peptide.”
[C0024467|Magnesium – DISRUPTS – C0019595|Histamine Release]
[PMID6193886]
Title “Substance P and endogenous opioids: how and where they could play a role in cluster headache.”
[C0038585|Substance P – AFFECTS – C0009088|Cluster Headache]
Abstract “Substance P appears to be involved in the transmission of pain signals from the periphery to the spinal cord and brain stem. Substance P containing neurons are responsible for the neurogenic vasodilation identical to that obtained by substance P release evoked by antidromic stimulation of these fibres.”

N/A
[PMID6202853]
Abstract “In NCB -20 cells, this voltage-sensitive 45Ca2+ uptake was inhibited selectively by organic calcium antagonists such as nitrendipine, cinnarizine, verapamil, and diltiazem (IC50 values = 6.4, 750, 1800, and 4500 nM, respectively).”
[C0008803|Cinnarizine – ISA – C0006684|Calcium Channel Blockers] [C0012373|Diltiazem – ISA – C0006684|Calcium Channel Blockers] [C0028127|Nitrendipine – ISA – C0006684|Calcium Channel Blockers] [C0042523|Verapamil – ISA – C0006684|Calcium Channel Blockers]
[PMID6275679]
Title “Serotonin and cyclic nucleotides in migraine.”
[C0036751|Serotonin – ASSOCIATED_WITH – C0149931|Migraine Disorders]
[PMID6279807]
Abstract “The relation of circulating insulin and glucagon concentrations to effects of magnesium deficiency was explored in experiment 3. The results of these experiments suggest that magnesium deficiency alters PEPCK activity by affecting secretion of pancreatic hormones.”
[C0021641|Insulin – AFFECTS – C0024473|Magnesium Deficiency]
[PMID6339937]
Title “Nifedipine in the treatment of migraine in patients with Raynaud’s phenomenon.”
[C0028066|Nifedipine – TREATS – C0149931|Migraine Disorders]
[PMID6358126]

Abstract – Taken from publisher's site. "Since Wolff's original proposal regarding the vascular etiology of cyclic head pain, evidence has accumulated that the prodromes of migraine are due to cerebral vasoconstriction and headaches of both cluster and migraine are due to painful dilatation."
[C0018681|Headache – CAUSES – C0149931|Migraine Disorders] [C0149931|Migraine Disorders – ISA – C0018681|Headache] [C1167661|Cerebral vasoconstriction – CAUSES – C0149931|Migraine Disorders]
Abstract – Taken from publisher's site. "Nifedipine and verapamil provided equivalent relief for cluster but produced more side effects, and were less effective, than nimodipine in control of migraine."
[C0028066|Nifedipine – TREATS – C0149931|Migraine Disorders] [C0042523|Verapamil – TREATS – C0149931|Migraine Disorders]
Abstract – Taken from publisher's site. "Theories regarding their pathogenesis include cyclic release of vasoactive substances from platelets and/or other sources (such as serotonin, catecholamines, histamine, acetyl choline, prostaglandins, substance P, endogenous opiates). Drugs which modify receptors (such as methysergide, alpha and beta blockers, antihistaminics, anticholinergics, steroids and non-steriodal anti-inflammatory agents) have had some therapeutic success in migraine but provide little benefit for cluster patients. Ca2+ entry blockers (including nimodipine, nifedipine, verapamil) theoretically should diminish cephalic vasoconstriction and –dilatation no matter what their cause."
SemRep did not derive a predicate from this statement, but it is worth noting.
Abstract – Taken from publisher's site. "Within 10 days migraine prodromes became infrequent and after 2-4 weeks headache frequency was significantly reduced for migraine and within 4-6 weeks for cluster."
[C0149931|Migraine Disorders – ISA – C0018681|Headache]
[PMID6358127]
Abstract – Taken from publisher's site. "The calcium channel antagonists are a recently developed class of vasodilators that prevent the influx of calcium into vascular smooth muscle. The unique pharmacologic effects of these agents provide a theoretical basis for their use in the treatment of migraine."
[C0450442|Agent – TREATS – C0149931|Migraine Disorders]

Abstract – Taken from publisher's site. "Clinically, two specific calcium antagonists (flunarizine, cinnarizine) and two non-specific antagonists (cyproheptadine, amitriptyline) are effective in the treatment of migraine."
[C0006684|Calcium Channel Blockers – TREATS – C0149931|Migraine Disorders]
[PMID6375330]
Introductory Paragraph – Taken from article. "In a sense, magnesium may be considered nature's physiologic calcium blocker."
[C0024467|Magnesium – INHIBITS – C0006675|Calcium]
[PMID6392763]
Title "Reversible retinal vasospasm in magnesium-treated hypertension despite no significant change in blood pressure."
[C0085616|Vasospasm – COEXISTS_WITH – C0020538|Hypertensive disease]
[PMID6399344]
Abstract "Experimental magnesium deficiency induces arterial damage, a loss of magnesium and potassium and an increase in the calcium and sodium content of the cell."
[C0024467|Magnesium – COEXISTS_WITH – C0006675|Calcium]
Abstract "Magnesium and potassium depletion have also been reported in diabetes and the vascular implications of this should be considered."
[C0024467|Magnesium – ASSOCIATED_WITH – C0011847|Diabetes]
[PMID6407598]
Abstract "The effect of magnesium on blood pressure may be direct or through influences on the

internal balance of potassium, sodium, and calcium.”
[C0006675|Calcium – INTERACTS_WITH|C0024467|Magnesium|bacs,elii|bacs||
[PMID6503357]
Title “Inhibition of calcium by magnesium in the contraction of rat aortic smooth muscle.”
[C0024467|Magnesium – ASSOCIATED_WITH – C1140999|Contraction] [C0024467|Magnesium – INHIBITS – C0006675|Calcium]
[PMID6538269]
Abstract “Cyproheptadine is equipotent (IC50 = 41 to 45 nM) in blocking contractions of canine basilar artery segments induced by serotonin, norepinephrine, potassium, or calcium.”
[C0006675|Calcium – CAUSES – C1140999|Contraction] [C0028351|Norepinephrine – CAUSES – C1140999|Contraction] [C0036751|Serotonin – CAUSES – C1140999|Contraction]
Abstract “Propranolol, at concentrations to 10 micromolar, had minimal effect on vessel contractions.”
[C0033497|Propranolol – AFFECTS – C1140999|Contraction]
Abstract “We conclude that the primary action of cyproheptadine in preventing induced contractions of the canine basilar artery is antagonism of calcium channels.This action is unique among drugs used for migraine prophylaxis and may have important implications for the treatment of headache and other neurologic disorders.”
SemRep did not derive a predicate from this statement, but it is worth noting.
[PMID6539877]
Title “Verapamil in prophylactic therapy of migraine.”
[C0042523|Verapamil – TREATS – C0149931|Migraine Disorders]

Abstract “Verapamil significantly reduced both headache frequency and duration with few side effects.”
[C0042523|Verapamil – PREVENTS – C0018681|Headache]
[PMID65562]
Abstract “No differences were found in plasma tryptophan, glucose, insulin, and free fatty acid levels in the migraine/no-migraine categories.”
[C0017725|Glucose – ASSOCIATED_WITH – C0149931|Migraine Disorders] [C0021641|Insulin – ASSOCIATED_WITH – C0149931|Migraine Disorders]
[PMID6680619]
Title “Pharmacologic inhibition of cerebral vasospasm in ischemia, hallucinogen ingestion, and hypomagnesemia: barbiturates, calcium antagonists, and magnesium.”
[C0006684|Calcium Channel Blockers – DISRUPTS – C0265110|Cerebral Vasospasm] [C0024467|Magnesium – DISRUPTS – C0265110|Cerebral Vasospasm] [C0265110|Cerebral Vasospasm – COEXISTS_WITH – C0022116|Ischemia] [C0265110|Cerebral Vasospasm – COEXISTS_WITH – C0151723|Hypomagnesemia]
Abstract “Experiments indicate that several different calcium antagonists have vasodilatory properties which may be expressed selectively on different organ vascular beds. Magnesium is a competitive calcium antagonist, and alterations in the extracellular content of this ion have profound effects on cerebral vascular resistance.”
[C0006684|Calcium Channel Blockers – AFFECTS – C0042380|Vascular resistance]
[PMID6708731]
Abstract “Flunarizine is a calcium entry blocking drug possessing antihypoxic activity in animal models of cerebral and peripheral ischemia-anoxia and has clinical usefulness in circulatory disorders of both central and peripheral origin.”
[C0016295|Flunarizine – ISA – C0006684|Calcium Channel Blockers]
Abstract “This report compares the activity of flunarizine and verapamil, another calcium entry blocking drug, on the central nervous system (CNS) and peripheral consequences of cytotoxic hypoxia induced by high and low doses of KCN.”

[C0042523|Verapamil – ISA – C0006684|Calcium Channel Blockers]
Abstract "Since low doses of the cyanide ion render respiration quicker and deeper by an action on chemoreceptive cells in peripheral arteries, the effect of verapamil against the hypoxic effect of KCN is mediated by an action in the periphery."
[C0042523|Verapamil – AFFECTS – C0242184|Hypoxia]
Abstract "In summary, we have shown that the physiological consequences of cytotoxic hypoxia can be affected by calcium entry blocking drugs having site-specific activities."
[C0006684|Calcium Channel Blockers – AFFECTS – C0242184|Hypoxia]
[PMID6715160]
Title "Relative potency and selectivity of calcium antagonists used in the treatment of migraine."
[C0006684|Calcium Channel Blockers – TREATS – C0149931|Migraine Disorders]
Abstract – Taken from publisher's site. "Calcium channel antagonists selectively block intracerebral vasoconstriction and appear clinically effective in the treatment of migraine."
[C0243076|antagonists – TREATS – C0149931|Migraine Disorders]
[PMID6802656]
Title "Renal hypomagnesaemia in human diabetes mellitus: its relation to glucose homeostasis."
[C0151723|Hypomagnesemia – COEXISTS_WITH – C0011849|Diabetes Mellitus]
Abstract "These data indicate that the net tubular reabsorption of magnesium is decreased in diabetic patients in presence of hyperglycaemia, leading to hypermagnesiuria and hypomagnesaemia."
SemRep did not derive a predicate from this statement, but it is worth noting.
[PMID6830162]
Title "Extracranial vascular changes and the source of pain in migraine headache."

[C0030193|Pain – ASSOCIATED_WITH – C0149931|Migraine Disorders]
[PMID6847018]
Title "Calcium and magnesium nutrition in human hypertension."
[C0006675|Calcium – ASSOCIATED_WITH – C0020538|Hypertensive disease]
Abstract "Many studies suggest that reduced consumption of calcium or magnesium is associated with an increased risk of developing hypertension and cardiovascular disease."
[C0024467|Magnesium – PREDISPOSES – C0020538|Hypertensive disease] [C0006675|Calcium – PREDISPOSES – C0020538|Hypertensive disease]
[PMID6864566]
Abstract "These results show that monoamines can modulate long-term changes in synaptic function in the dentate gyrus, and suggest that 5-HT is more potent in this respect than NA."
[C0036751|Serotonin – compared_with – C0028351|Norepinephrine] [C0036751|Serotonin – higher_than – C0028351|Norepinephrine]
[PMID6869560]
Abstract "The vasodilator effect of Mg2+ on pial arterioles was enhanced in the presence of the calcium antagonist verapamil (0.5 micrograms/ml), despite the fact that verapamil by itself caused a 12-13% arteriolar dilation".
[C0042523|Verapamil – CAUSES – C0012359|Pathological Dilatation] [C0042523|Verapamil – ISA – C0006684|Calcium Channel Blockers]
[PMID6948814]
Abstract "In isolated human pial arteries (diameter 0.4-0.5 mm), contractions were produced by potassium, noradrenaline, serotonin, and prostaglandin F2 alpha."
[C0028351|Norepinephrine – CAUSES – C1140999|Contraction] [C0036751|Serotonin – CAUSES – C1140999|Contraction]
Abstract "Both nifedipine and nimodipine effectively inhibited contraction elicited by noradrenaline and serotonin in pial arteries."

[C0028066|Nifedipine – DISRUPTS – C1140999|Contraction] [C0028094|Nimodipine – DISRUPTS – C1140999|Contraction] [C0028351|Norepinephrine – AUGMENTS – C1140999|Contraction] [C0036751|Serotonin – AUGMENTS – C1140999|Contraction]
Abstract "Both nifedipine and nimodipine effectively inhibited contractions induced by calcium in pial arteries pretreated in a calcium-free medium and depolarised by potassium."
[C0006675|Calcium – CAUSES – C1140999|Contraction] [C0028066|Nifedipine – DISRUPTS – C1140999|Contraction]
[PMID6992233]
Abstract "Graft failure, as determined by reduction and ultimate cessation of renal blood flow and urine production, can be abrogated in the short term by prostacyclin which is the most potent inhibitor of platelet aggregation yet discovered."
[C0033567|Epoprostenol – DISRUPTS – C0032176|Platelet aggregation]
[PMID7014260]
Abstract "Further, authentic 6-keto-PGE1, like PGI2, escapes pulmonary degradation and is a potent inhibitor of platelet aggregation."
[C0033567|Epoprostenol – DISRUPTS – C0032176|Platelet aggregation]
[PMID7026501]
Title "The effect of infused prostacyclin in migraine and cluster headache."
[C0033567|Epoprostenol – AFFECTS – C0009088|Cluster Headache] [C0033567|Epoprostenol – AFFECTS – C0149931|Migraine Disorders]
Abstract – Taken from publisher's site. Prostacyclin appears to be the most active of the vasodilating prostaglandins with inflammatory and hyperalgesic properties.
[C0033567|Epoprostenol – ISA – C0033554|Prostaglandins]
Abstract – Taken from publisher's site. Its actions are very similar to those of prostaglandin E1, which has long been known to cause "vascular-type" headaches.
[C0002335|Alprostadil – CAUSES – C0042376|Vascular Headaches]

[PMID7048862]
Abstract “Prostacyclin (PGI2), a strong vasodilator of cerebral vessels and potent inhibitor of platelet aggregation, was infused intravenously into seven cats after induction of prolonged vasospasm by hourly application of oxyhaemoglobin solution into the subarachnoid space round the basilar artery.”
[C0033567|Epoprostenol – DISRUPTS – C0032176|Platelet aggregation]
[PMID7060640]
Abstract “5-Hydroxytryptamine amplified the contractions evoked by threshold concentrations of histamine, angiotensin II and prostaglandin F2 alpha; in all three cases, the amplification was antagonized by comparable concentrations of ketanserin.”
[C0036751|Serotonin – AUGMENTS – C1140999|Contraction]
Abstract “The inhibition by ketanserin of the amplifying effect of 5-hydroxytryptamine on vascular responses may help explain the antihypertensive properties of the compound.”
[C0022616|Ketanserin – INHIBITS – C0036751|Serotonin]
[PMID7068108]
Abstract – taken from paper itself “Hypomagnesemia has been reported to occur in diabetes mellitus in the course of recovery from ketoacidoses, as well as during insulin maintenance therapy.”
[C0151723|Hypomagnesemia – COEXISTS_WITH – C0011849|Diabetes Mellitus]
[PMID7150872]
Title “Withdrawal of magnesium enhances coronary arterial spasms produced by vasoactive agents.”
[C0024467|Magnesium – AUGMENTS – C0010073|Coronary Artery Vasospasm]
Abstract “The calcium channel blocking agent, verapamil (10(-6)M), inhibited completely contractile responses to KCl; contractile responses elicited by angiotensin II and 5-HT were attenuated by verapamil”
[C0036751|Serotonin – STIMULATES – C0042523|Verapamil]
[PMID7174304]

Title “Long-term study of propranolol in the treatment of migraine.”
[C0033497|Propranolol – TREATS – C0149931|Migraine Disorders]
[PMID7195070]
Abstract “A specific calcium antagonist, verapamil, readily prevented (and reversed) PCP-induced vasospasm.”
[C0042523|Verapamil – ISA – C0006684|Calcium Channel Blockers] [C0042523|Verapamil – PREVENTS – C0085616|Vasospasm]
[PMID7297597]
Abstract “Previous studies on isolated blood vessels indicate that (1) withdrawal of magnesium ([Mg2+]0) induces calcium-dependent contractile responses, and (2) Mg2+]0 and verapamil (VE) inhibit calcium influx across the cell membrane.”
[C0024467|Magnesium – STIMULATES – C0042523|Verapamil]
Abstract “The present study, using isolated rat aortic strips and portal veins, was designed to assess the interactions of [Mg2+]0 and VE on increases in active tension and contractility induced by withdrawal of [Mg2+]0.[Mg2+]0 was found to: (1) enhance VE-induced inhibition of portal vein amplitude, and (2) antagonize VE-induced enhanced frequency of spontaneous phasic contractions in portal veins.”
[C0042523|Verapamil – AUGMENTS – C1140999|Contraction] [C0042523|Verapamil – CAUSES – C1140999|Contraction]
Abstract “Both [Mg2+]0 and VE could prevent and reverse the increases in active tension developed in aortic smooth muscle upon removal of [Mg2+]0.[Mg2+]0 markedly potentiated the inhibitory effect of verapamil on calcium-induced contractions of K+-depolarized aorta but not in K+-depolarized portal vein.”
[C0042523|Verapamil – DISRUPTS – C1140999|Contraction] [C0042523|Verapamil – TREATS – C0233494|Tension]
[PMID7305671]
Title ”Hypomagnesemia and vasoconstriction: possible relationship to etiology of sudden death ischemic heart disease and hypertensive vascular diseases.”

[C0085616|Vasospasm – CAUSES – C0020538|Hypertensive disease] [C0085616|Vasospasm – CAUSES – C0151744|Myocardial Ischemia]
Abstract ”Hypomagnesemia and vasoconstriction: possible relationship to etiology of sudden death ischemic heart disease and hypertensive vascular diseases.”
[C0085616|Vasospasm – CAUSES – C0020538|Hypertensive disease]
Abstract “In-vitro experiments are presented which indicate that the concentration of extracellular magnesium ions ([Mg2+]o) can exert profound influences on the contractility and reactivity of arteries, arterioles and veins from a number of regional vasculatures in several mammalian species, including man.”
SemRep did not derive a predicate from this statement, but it is worth noting.
Abstract “Hypomagnesemia can potentiate the contractile activity of a variety of neurohumoral substances and induce vasospasm. Hypermagnesemia can do the reverse, i.e., induce hyporeactivity, relaxation and vasodilatation.”
SemRep did not derive a predicate from this statement, but it is worth noting.
Abstract “Data are reviewed which suggest that certain vascular diseases (e.g., sudden-death ischemic heart disease, hypertension, eclampsia, diabetes mellitus) are associated with a Mg-deficiency.”
SemRep did not derive a predicate from this statement, but it is worth noting.
[PMID7310427]
Title “Asymmetry of the aura and pain in migraine”
[C0030193|Pain – ASSOCIATED_WITH – C0149931|Migraine Disorders]
[PMID7437647]
Title “Differential effects of the calcium antagonist, verapamil, on lumen sizes of terminal arterioles and muscular venules in the rat mesenteric, pial and skeletal muscle microvasculatures.”
[C0042523|Verapamil – INTERACTS_WITH – C0006684|Calcium Channel Blockers] [C0042523|Verapamil – ISA – C0006684|Calcium Channel Blockers]
Abstract

“Local application of verapamil (1 to 100 micrograms) to cremaster muscle arterioles and venules of the rat induces dose-dependent vasodilatation and increased perfusion of capillaries.”
[C0042523|Verapamil – CAUSES – C0042401|Vasodilation]
Abstract “Although these direct in situ microvascular findings do indicate that verapamil can induced dilatation of microscopic resistance and capacitance vessels in skeletal muscle, our data do not support the concept that verapamil induces non-specific peripheral vasodilatation.”
[C0042523|Verapamil – CAUSES – C1328540|Peripheral vasodilatation]
[PMID7443079]
Abstract – Taken from publisher’s site. “Withdrawal of magnesium causes vasospasm while elevated magnesium produces relaxation of tone in cerebral arteries.”
[C0024467|Magnesium – CAUSES – C0085616|Vasospasm]
[PMID773081]
Title “Short-term clinical trial of phopranolol in racemic form (Inderal), D-propranolol and placebo in migraine.”
[C0033497|Propranolol – TREATS – C0149931|Migraine Disorders]
Abstract “The results indicate that beta-receptor blocking properties, but possibly also properties other than the beta-blocking ones, may be of importance for the anti-migraine effect of propranolol.”
SemRep did not derive a predicate from this statement, but it is worth noting.
[PMID89390]
Abstract “The effectiveness of tolfenamic acid in acute migraine attacks accords with the postulated role of prostaglandins in migraine.”
[C0033554|Prostaglandins – ASSOCIATED_WITH – C0149931|Migraine Disorders]
[PMID939020]
Abstract “The aim of the studies was to develop a solution which, if infused into the coronary

vessels just prior to the onset of ischemia, would rapidly induce arrest and would also counteract several of the deleterious cellular changes known to occur during myocardial ischemia.”
[C0022116|Ischemia – COEXISTS_WITH – C0151744|Myocardial Ischemia]
[PMID953864]
Abstract “When taken in conjunction with other evidence these results suggest that the diuretics exert their vascular effects by inhibiting prostaglandin synthesis whereas prolactin acts by stimulating such synthesis.”
[C0033554|Prostaglandins – AFFECTS – C0005847|Blood Vessels]

Table B-3. Precision and recall using citations for Magnesium study

Result	New Intermediate Terms	Relevancy (Topic)	Precision	Recall
d0	Epilepsy	Epilepsy	1.0	0.083
d1	Potassium Verapamil Dilitazam Nifedipine Calcium Channel Blockers Serotonin	Calcium Channel Blockers Calcium Channel Blockers Calcium Channel Blockers Calcium Channel Blockers Calcium Channel Blockers Serotonin	1.0	0.25
d2	Glucose Insulin Secretion	Diabetes Diabetes	1	0.333
d3	Pathological Dilatation	Vascular Mechanisms	1	0.417
d4	Insulin Fatty Acids	Diabetes Not Relevant	0.833	0.417
d5	Epoprostenol Prostaglandins Aspirin	Platelet Aggregation Prostaglandins Platelet Aggregation	0.875	0.583
d6	Sodium Hypertensive Disease	Calcium Channel Blockers Stress / Type A	0.889	0.667
d7	Pain Somatostatin	Inflammation Inflammation	0.9	0.75
d8	N/A		0.9	0.75
d9	Calcium	Calcium Channel Blockers	0.9	0.75
d10	Isoproterenol Propranolol	Vascular Mechanisms Vascular Mechanisms	0.9	0.75
d11	Norepinephrine Clonidine	Serotonin Vascular Mechanisms	0.9	0.75
d12	Homeostasis	Not Relevant	0.818	0.75
d13	N/A		0.818	0.75
d14	N/A		0.818	0.75
d15	Onset of illness	Not Relevant	0.75	0.75
d16	N/A		0.75	0.75
d17	Myocardial Ischemia Ischemia	Hypoxia Hypoxia	0.769	0.833

Table B-4. Precision and recall not using citations for Magnesium study

Result	New Intermediate Terms	Relevancy (Topic)	Precision	Recall
d0	Excretory Function	Not Relevant	0.0	0.0
d1	Epilepsy	Epilepsy	0.5	0.083
d2	Potassium Verapamil Nifedipine Calcium Channel Blockers	Calcium Channel Blockers Calcium Channel Blockers Calcium Channel Blockers Calcium Channel Blockers	0.667	0.167
d3	Glucose Insulin Secretion	Diabetes Diabetes	0.75	0.25
d4	Serotonin Pathological Dilatation	Serotonin Vascular Mechanisms	0.833	0.417
d5	Insulin Fatty Acids	Diabetes Not Relevant	0.714	0.417
d6	Epoprostenol Prostaglandins	Platelet Aggregation Prostaglandins	0.778	0.583
d7	Hypertensive Disease	Stress / Type A	0.8	0.667
d8	Pain Somatostatin	Inflammation Inflammation	0.818	0.75
d9	N/A		0.818	0.75
d10	Dilitazam Calcium	Calcium Channel Blockers Calcium Channel Blockers	0.818	0.75
d11	ATP	Not Relevant	0.75	0.75
d12	N/A		0.75	0.75
d13	Falls	Not Relevant	0.692	0.75
d14	Vasospasm	Vascular Mechanisms	0.692	0.75
d15	Myocardial Ischemia Ischemia	Hypoxia Hypoxia	0.714	0.833

Appendix C

Somatomedins – Arginine Study

Table C-1. Somatomedins – Arginine Core Corpus Medline IDs

412936	3075738	3690631	5459997	6754563
425782	3080468	3700933	5786514	6790230
430225	3080558	3731567	5841101	6846217
448444	3096900	3758932	5859253	6890417
492275	3099250	3782436	5877589	6989150
515169	3109352	3803997	5917038	6998997
606683	3127426	3833530	5929465	7007997
617386	3140744	3865530	5944094	7020137
632300	3259505	3884753	6176592	7038514
684614	3262126	3902942	6190641	7042736
686863	3288047	3918429	6202444	7045779
722344	3310600	3918442	6243390	7058674
744062	3315866	3923266	6262917	7079756
794768	3322823	4038409	6299718	7189195
839329	3380847	4056323	6324479	7193684
870020	3434631	4058978	6343062	7194884
893668	3499612	4158217	6363657	7194952
908965	3514981	4164428	6370513	7195462
939198	3525601	4205831	6373069	7197311
1110767	3533330	4430939	6414318	7197688
1120749	3546937	4538721	6416888	7198145
1254283	3558728	4550398	6427525	7198682
2536517	3558729	4568402	6428892	7213491
2648317	3569122	4624624	6540317	7264776
2840748	3652525	4631334	6583688	7327721
2897974	3653346	4696902	6609081	7350218
2899993	3654912	4919910	6639868	7391725
2938688	3665113	5059775	6681614	13429201
2959373	3668688	5165405	6684729	13491695
2972888	3683183	5347431	6749885	14328387

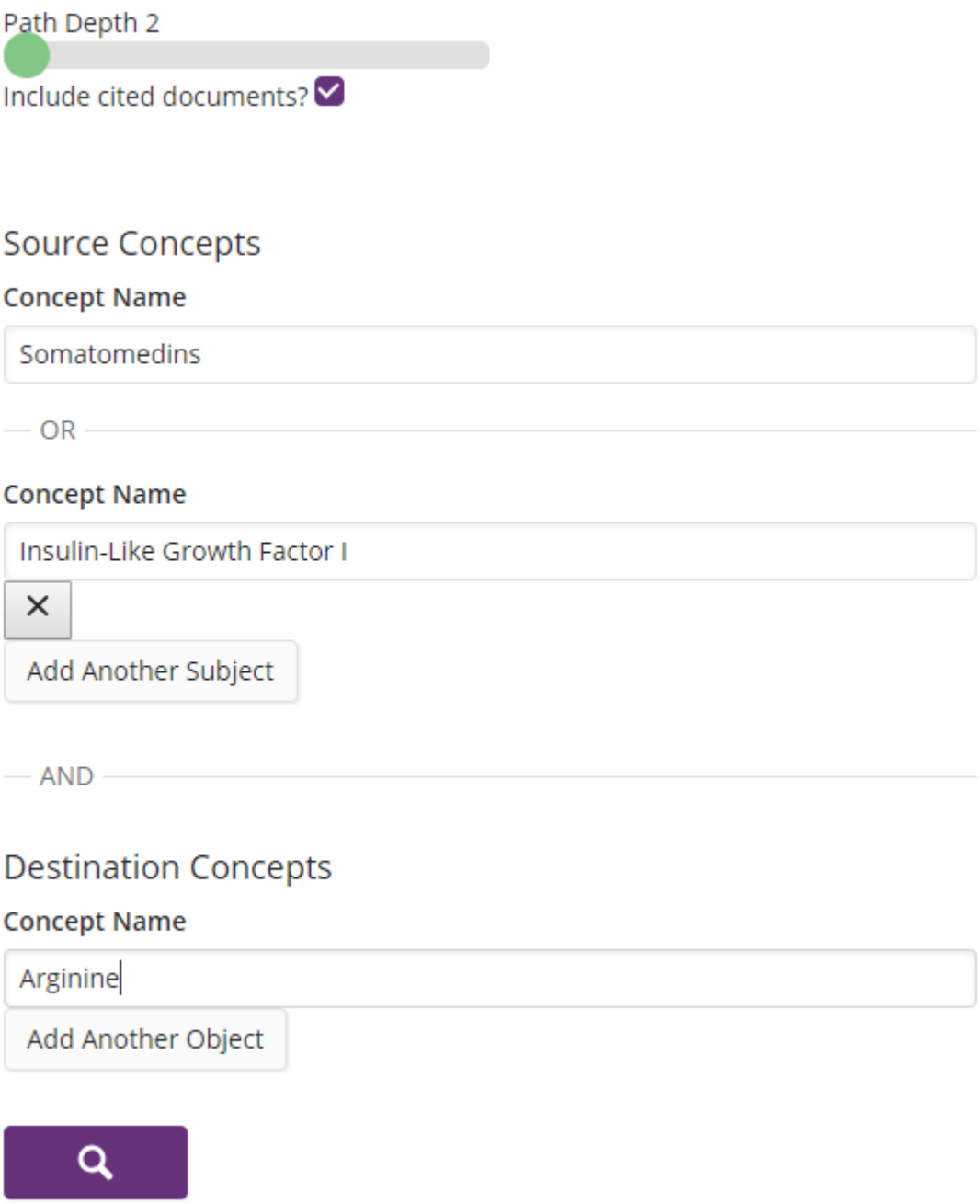


Figure C-1. Somatomedins – Arginine search parameters

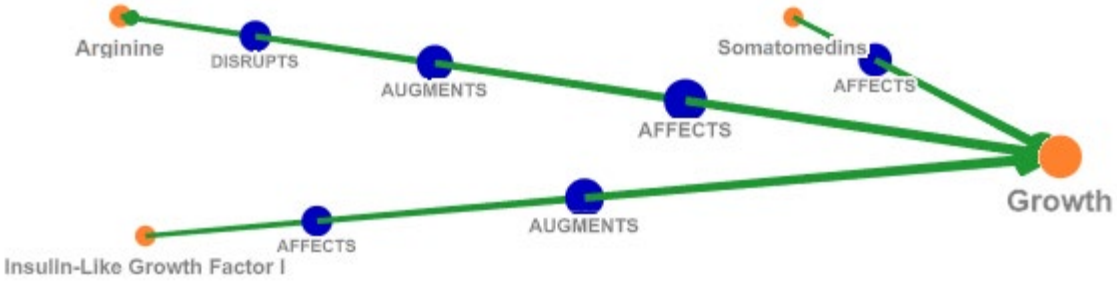


Figure C-2. Somatomedins – Arginine result 1 (with citations)

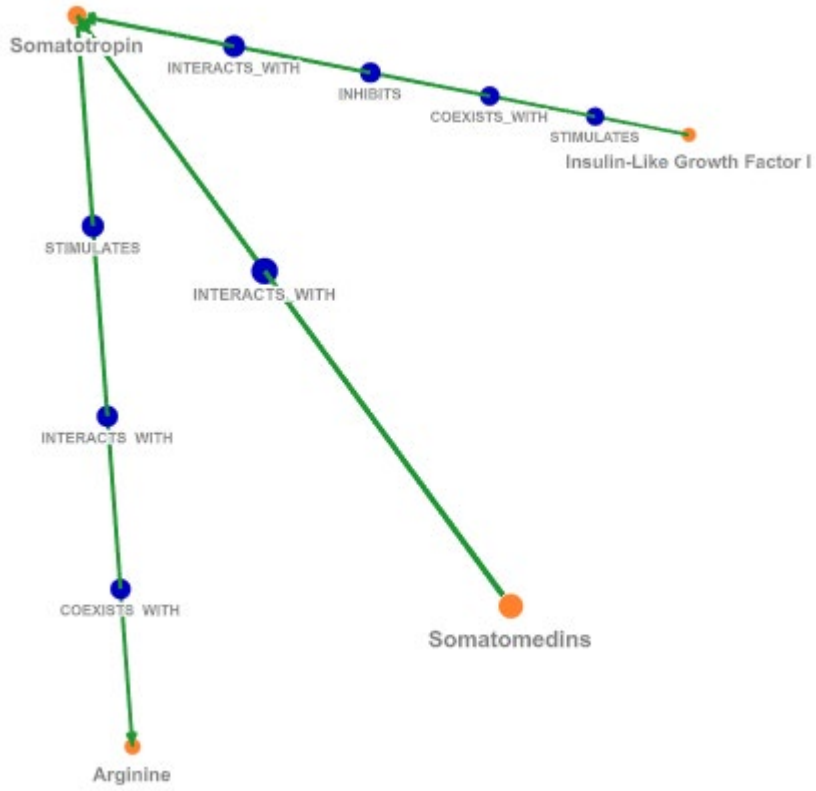


Figure C-3. Somatomedins – Arginine result 2 (with citations)

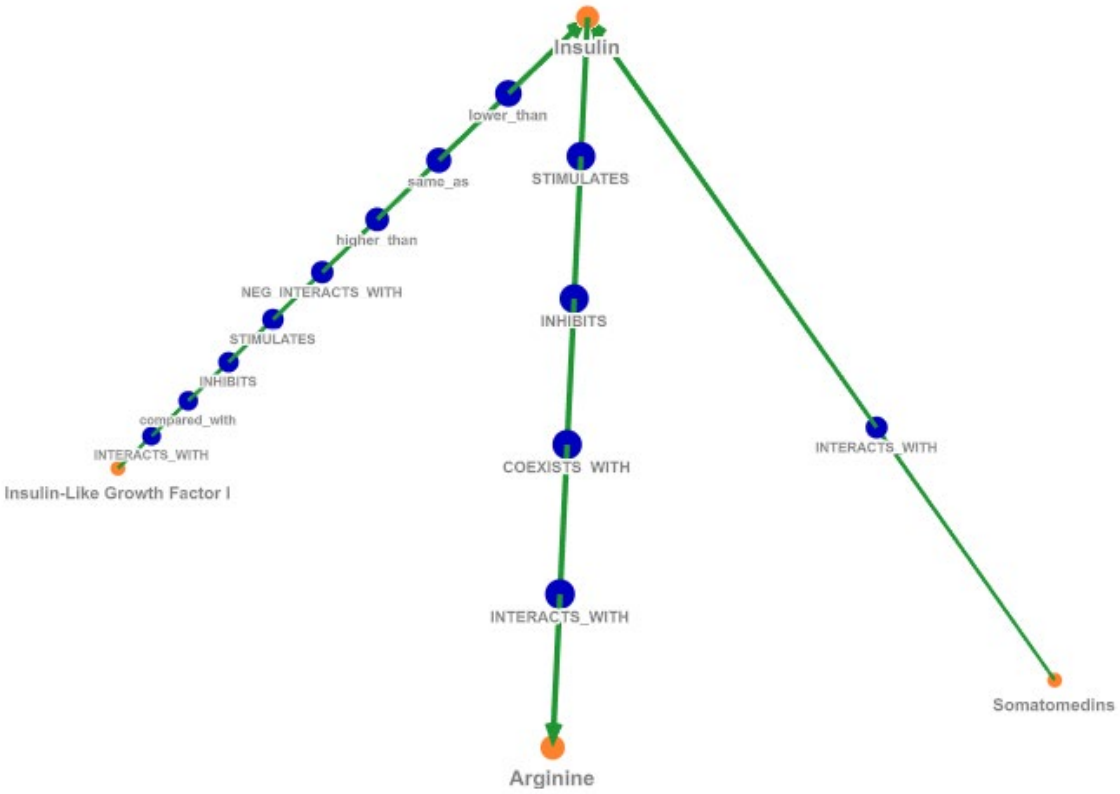


Figure C-4. Somatomedins – Arginine result 3 (with citations)

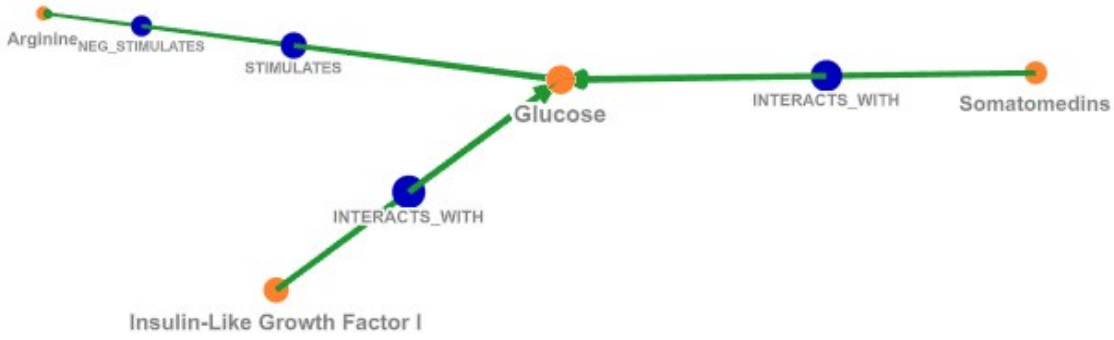


Figure C-5. Somatomedins – Arginine result 4 (with citations)

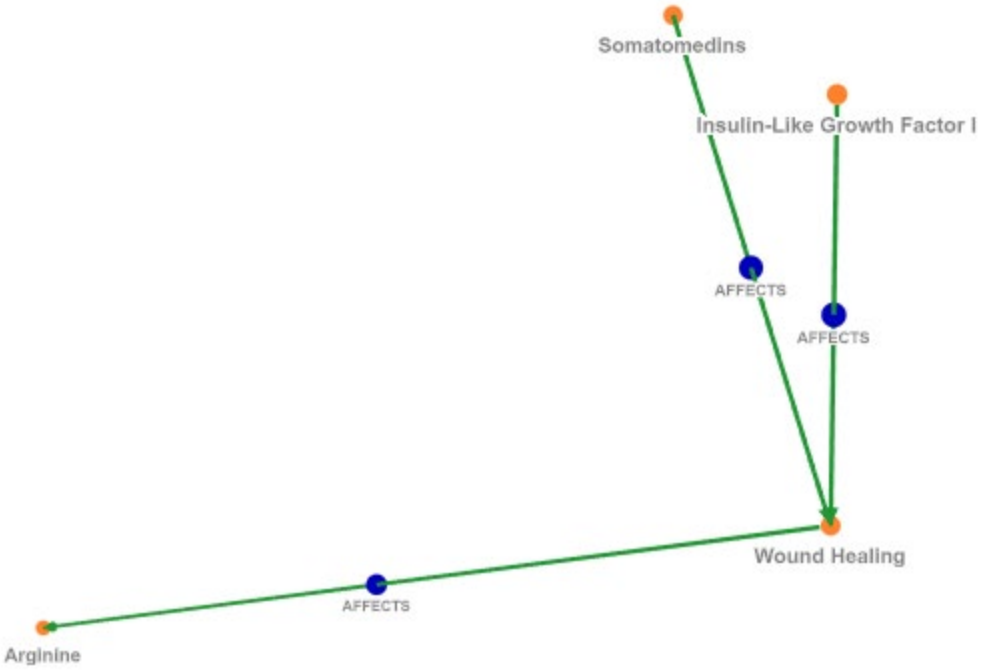


Figure C-6. Somatomedins – Arginine result 5 (with citations)

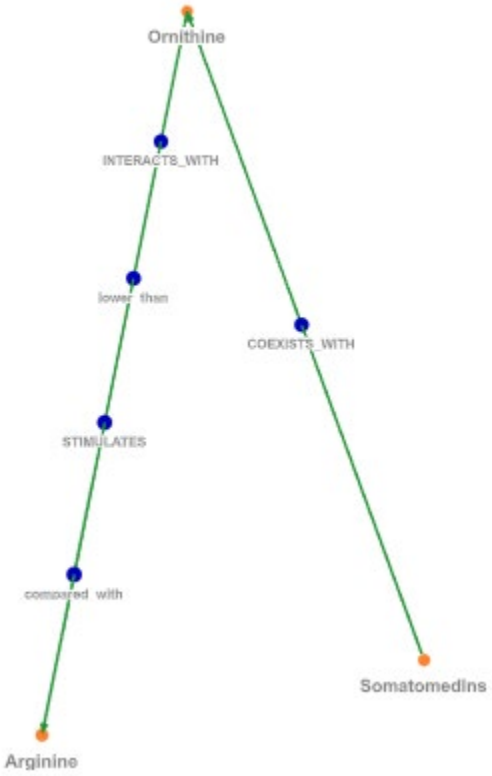


Figure C-7. Somatomedins – Arginine result 6 (with citations)

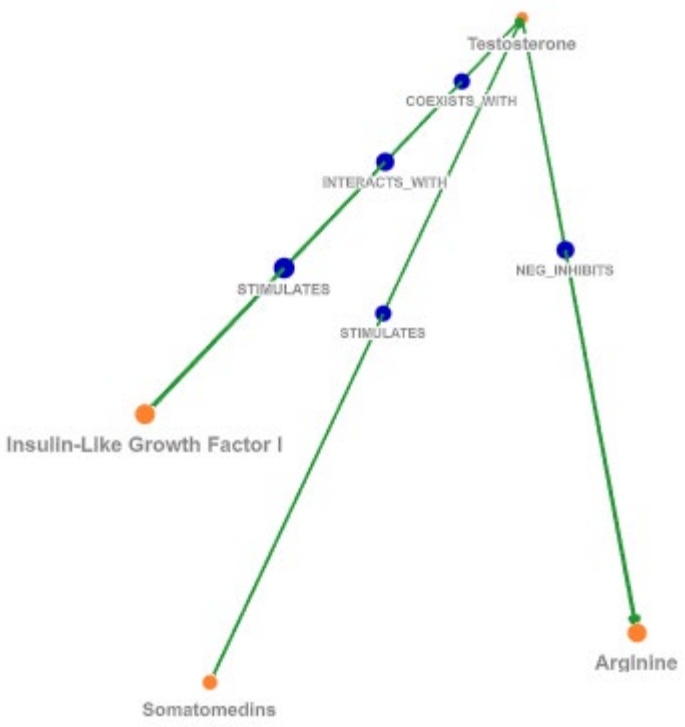


Figure C-8. Somatomedins – Arginine result 7 (with citations)

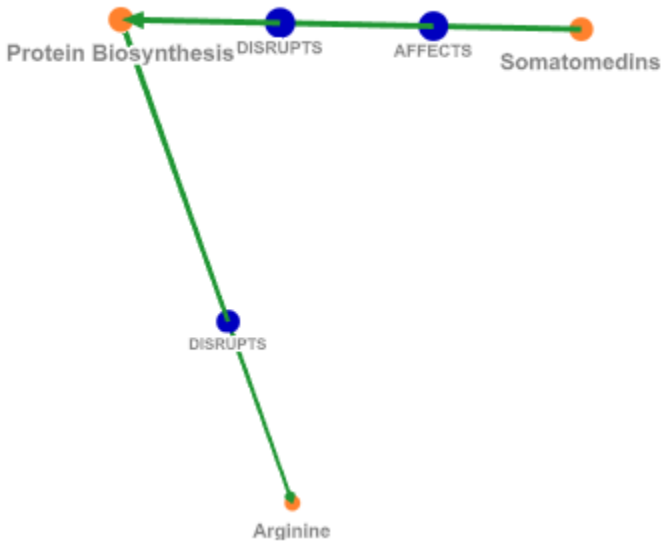


Figure C-9. Somatomedins – Arginine result 8 (with citations)

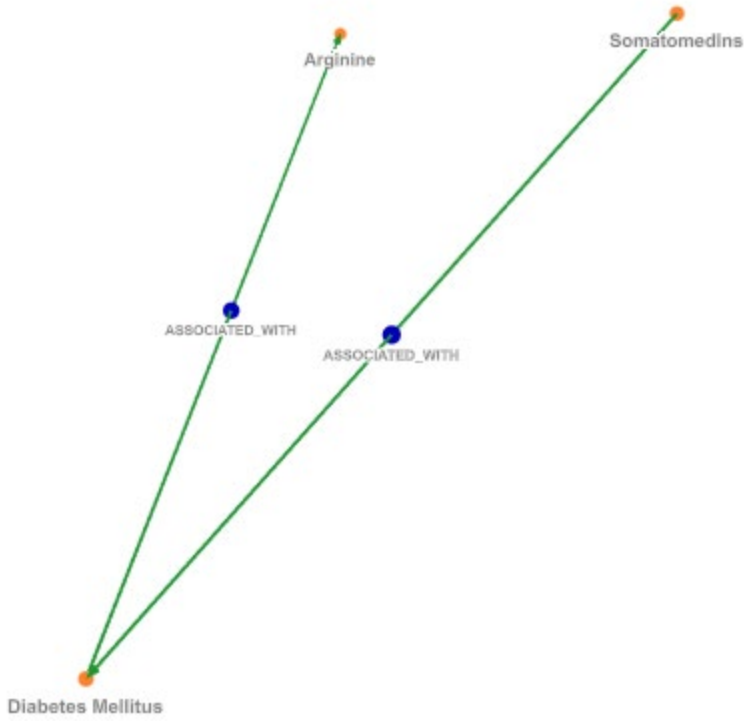


Figure C-10. Somatomedins – Arginine result 9 (with citations)

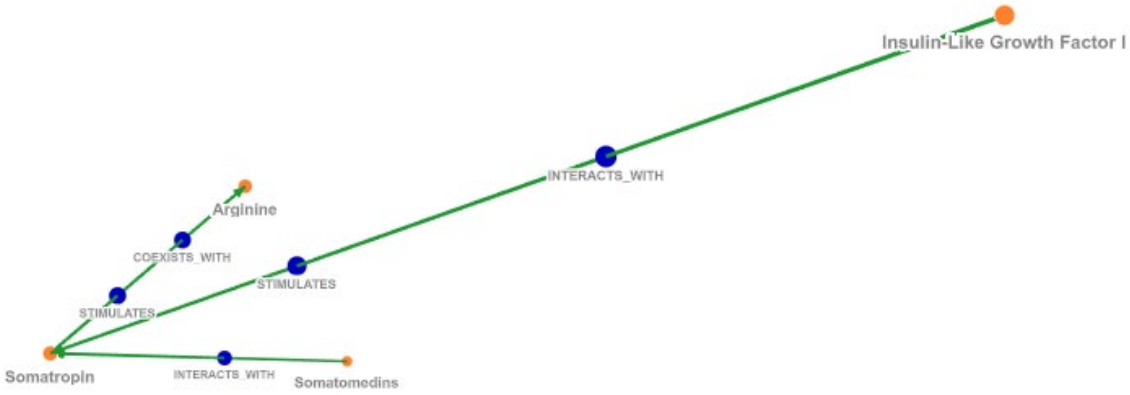


Figure C-11. Somatomedins – Arginine result 10 (with citations)

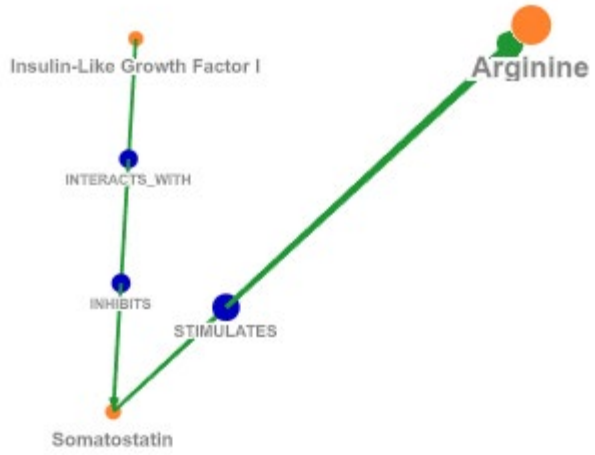


Figure C-12. Somatomedins – Arginine result 11 (with citations)

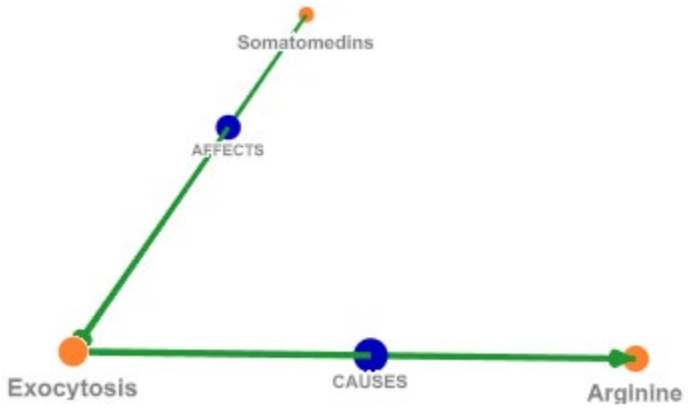


Figure C-13. Somatomedins – Arginine result 12 (with citations)

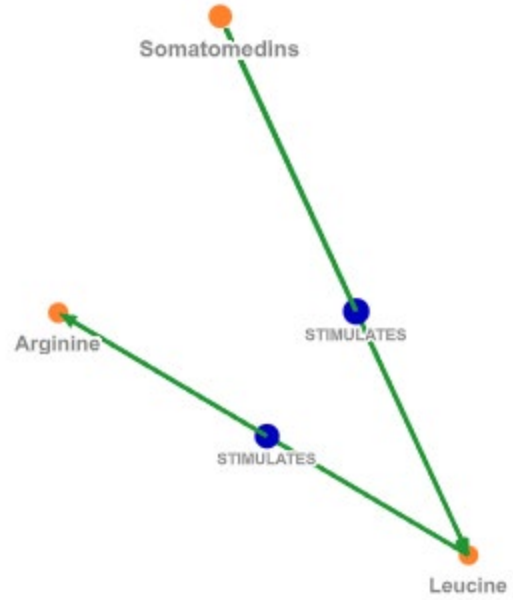


Figure C-14. Somatomedins – Arginine result 13 (with citations)

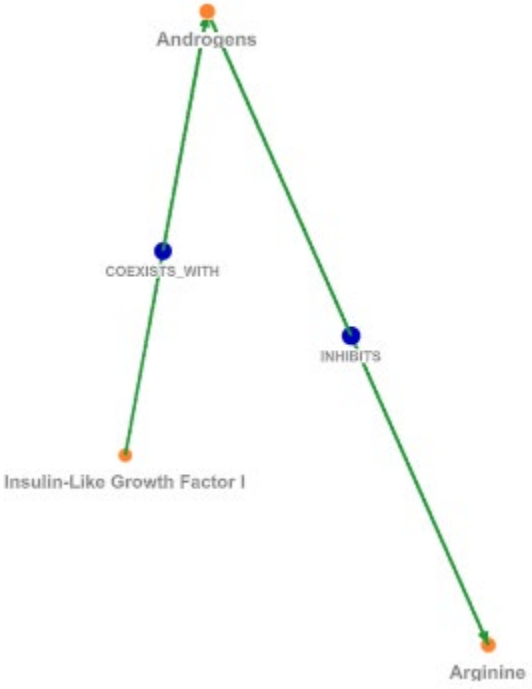


Figure C-15. Somatomedins – Arginine result 14 (with citations)

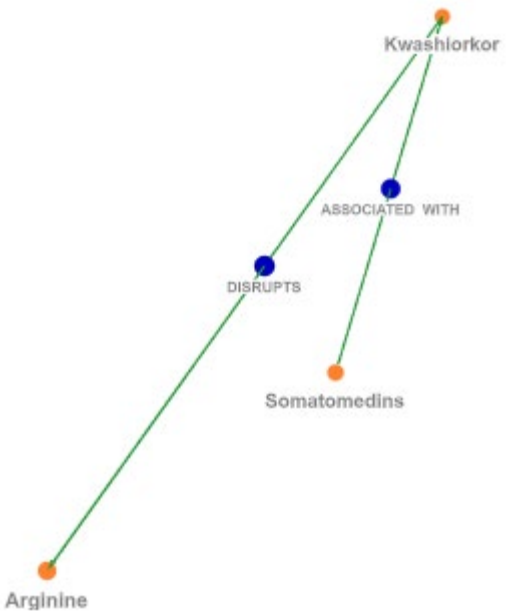


Figure C-16. Somatomedins – Arginine result 15 (with citations)

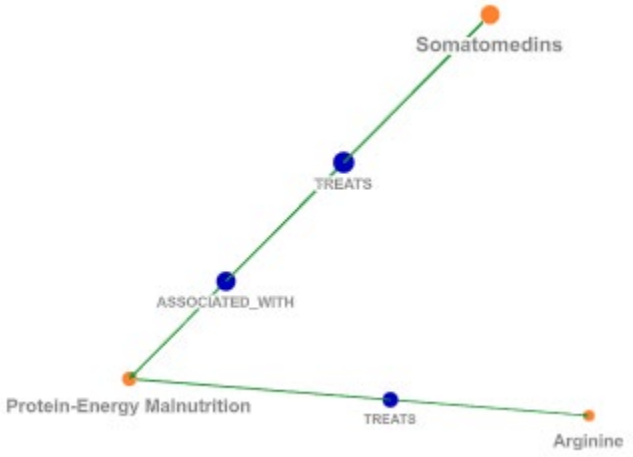


Figure C-17. Somatomedins – Arginine result 16 (with citations)

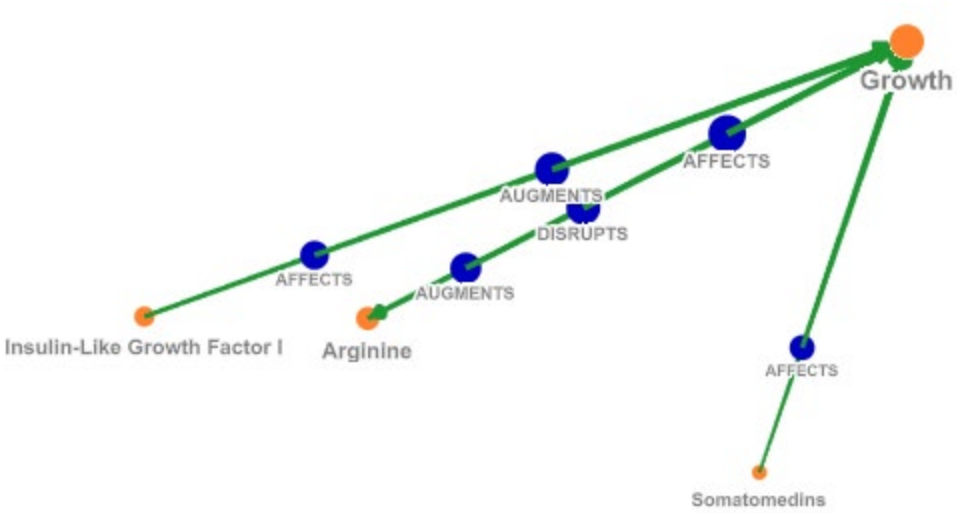


Figure C-18. Somatomedins – Arginine result 1 (no citations)

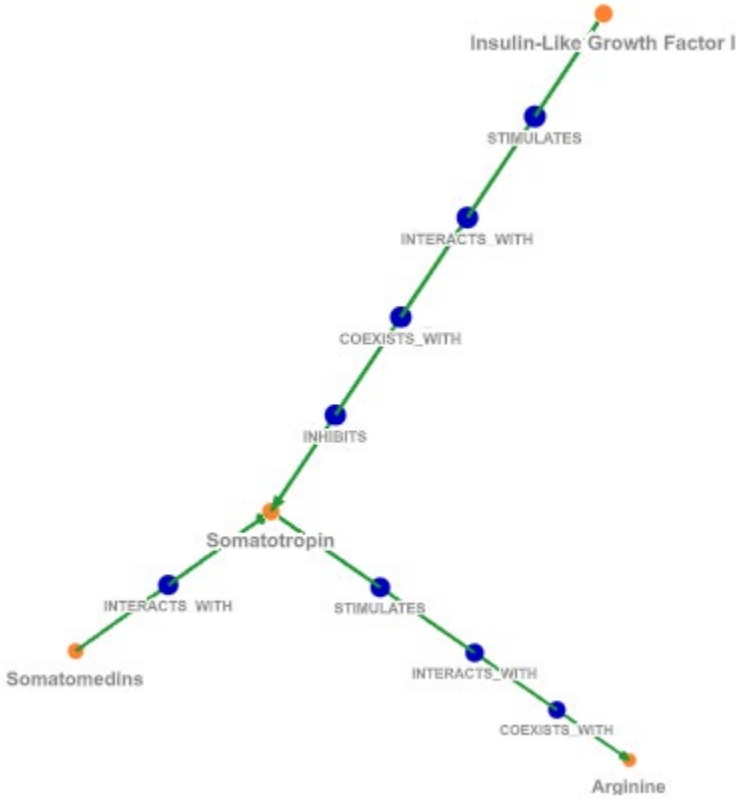


Figure C-19. Somatomedins – Arginine result 2 (no citations)

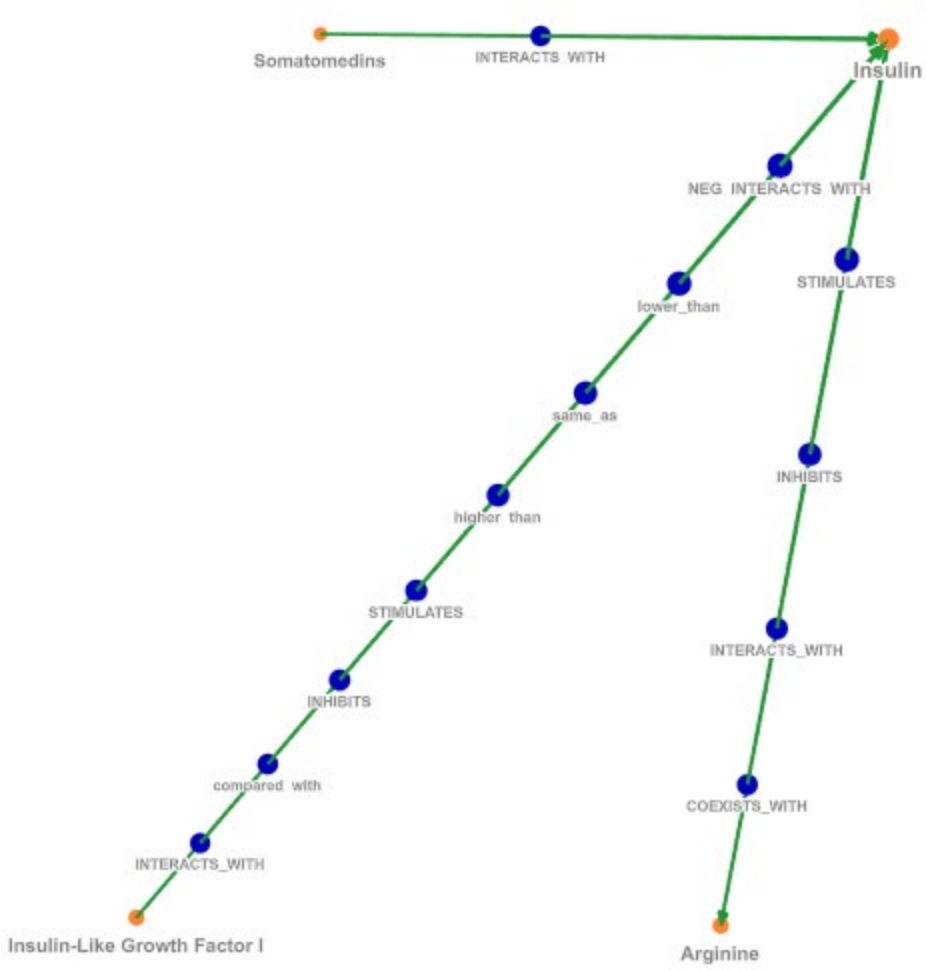


Figure C-20. Somatomedins – Arginine result 3 (no citations)

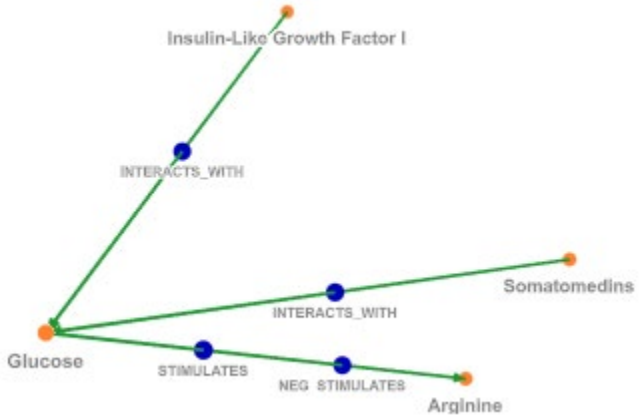


Figure C-21. Somatomedins – Arginine result 4 (no citations)

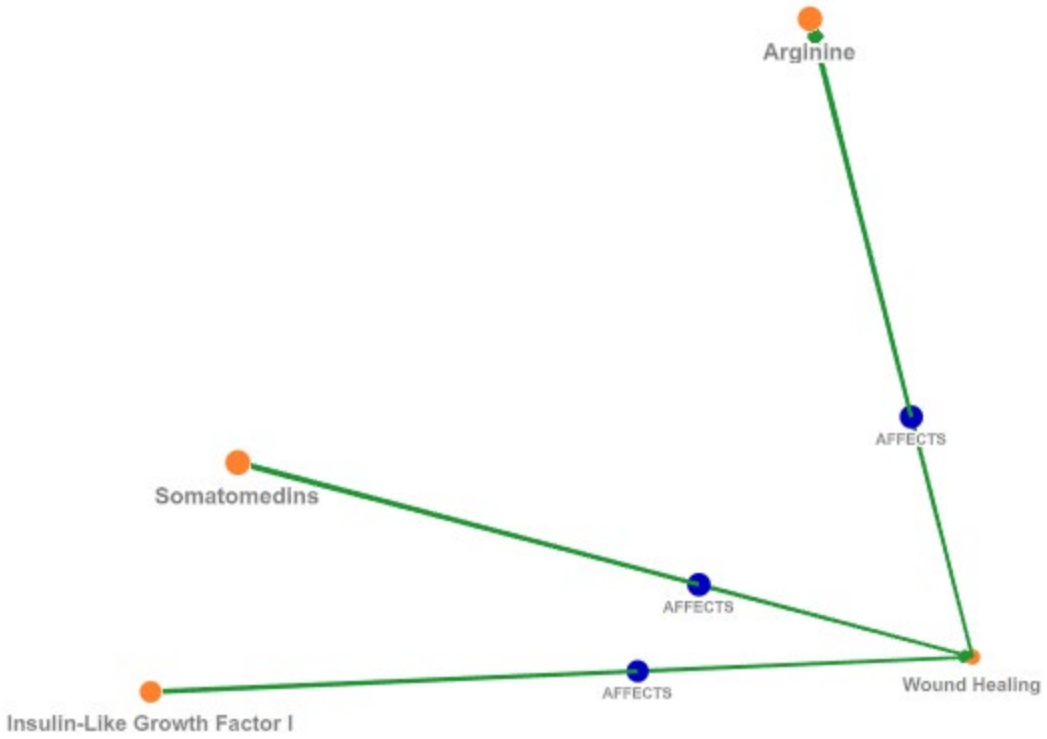


Figure C-22. Somatomedins – Arginine result 5 (no citations)

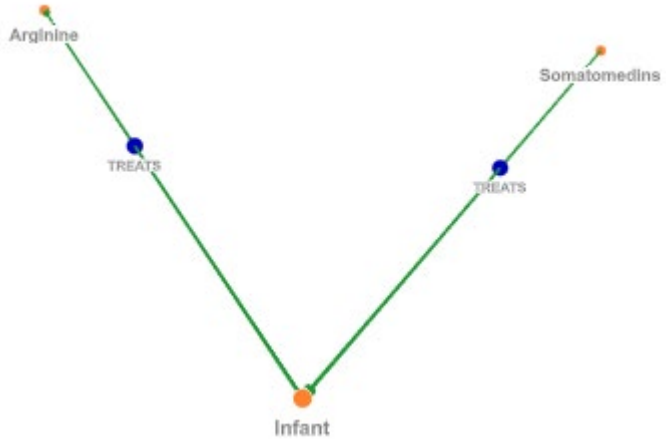


Figure C-23. Somatomedins – Arginine result 6 (no citations)

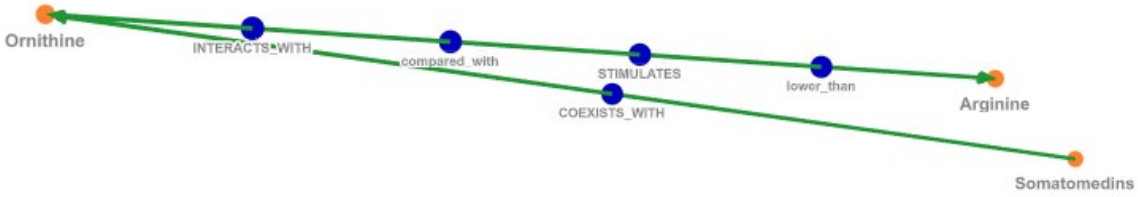


Figure C-24. Somatomedins – Arginine result 7 (no citations)

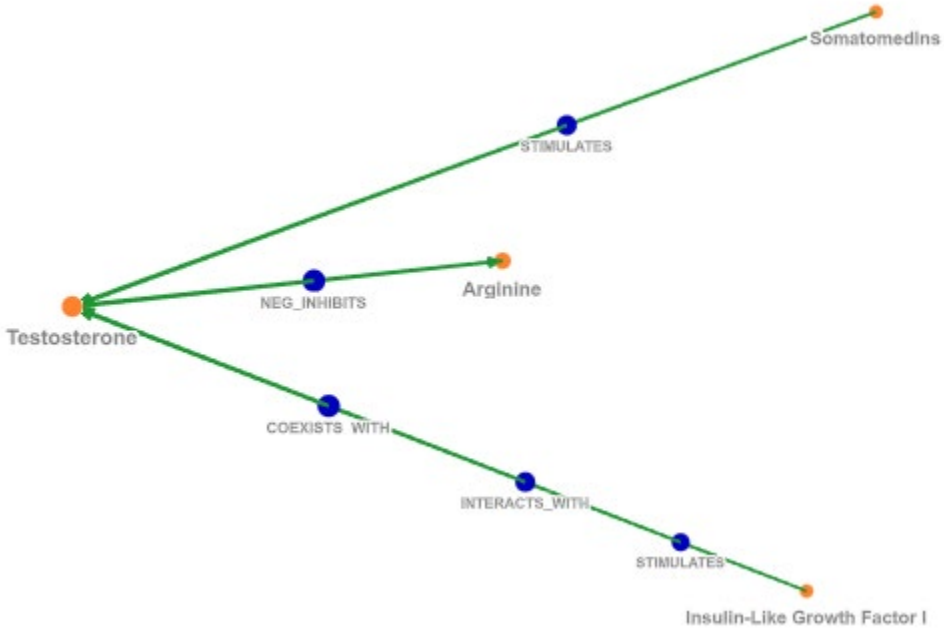


Figure C-25. Somatomedins – Arginine result 8 (no citations)

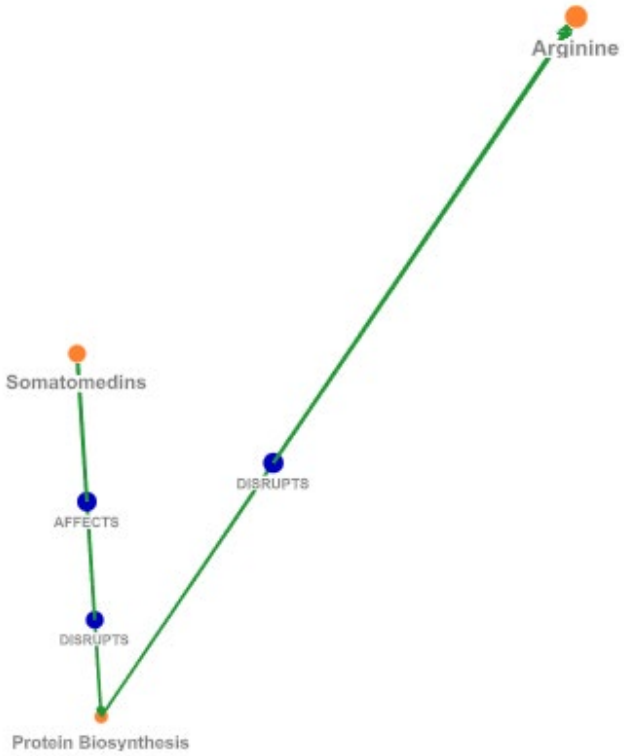


Figure C-26. Somatomedins – Arginine result 9 (no citations)

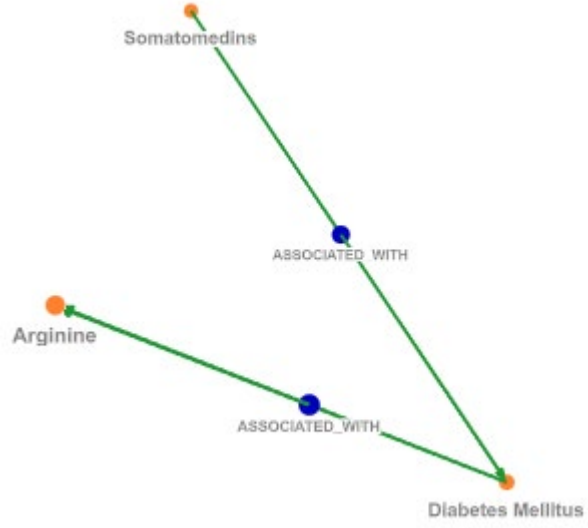


Figure C-27. Somatomedins – Arginine result 10 (no citations)

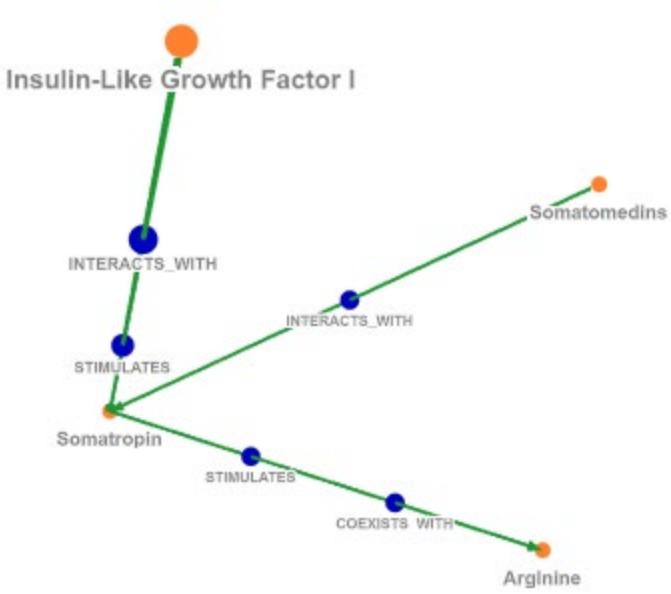


Figure C-28. Somatomedins – Arginine result 11 (no citations)

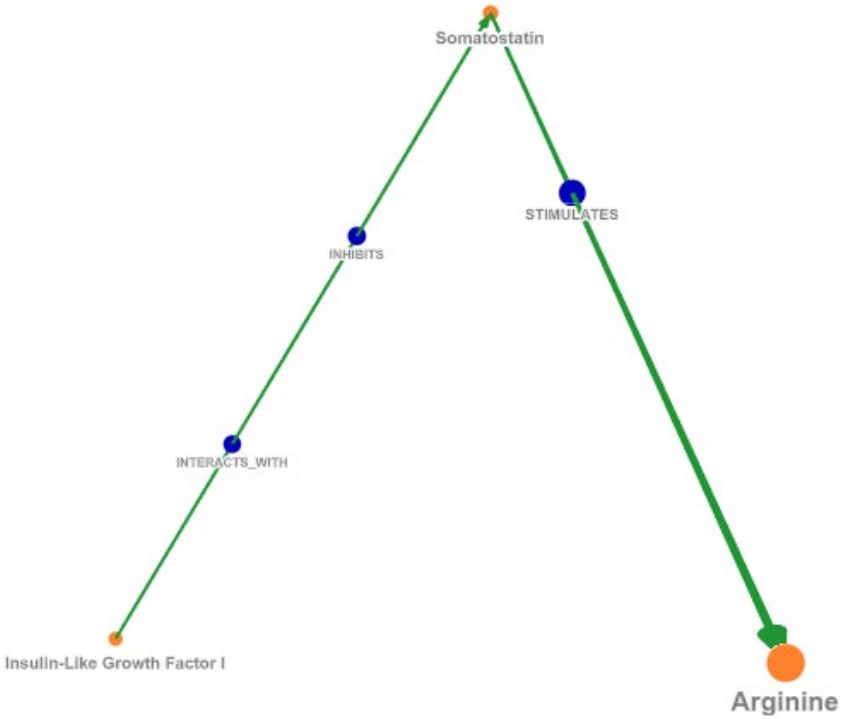


Figure C-29. Somatomedins – Arginine result 12 (no citations)

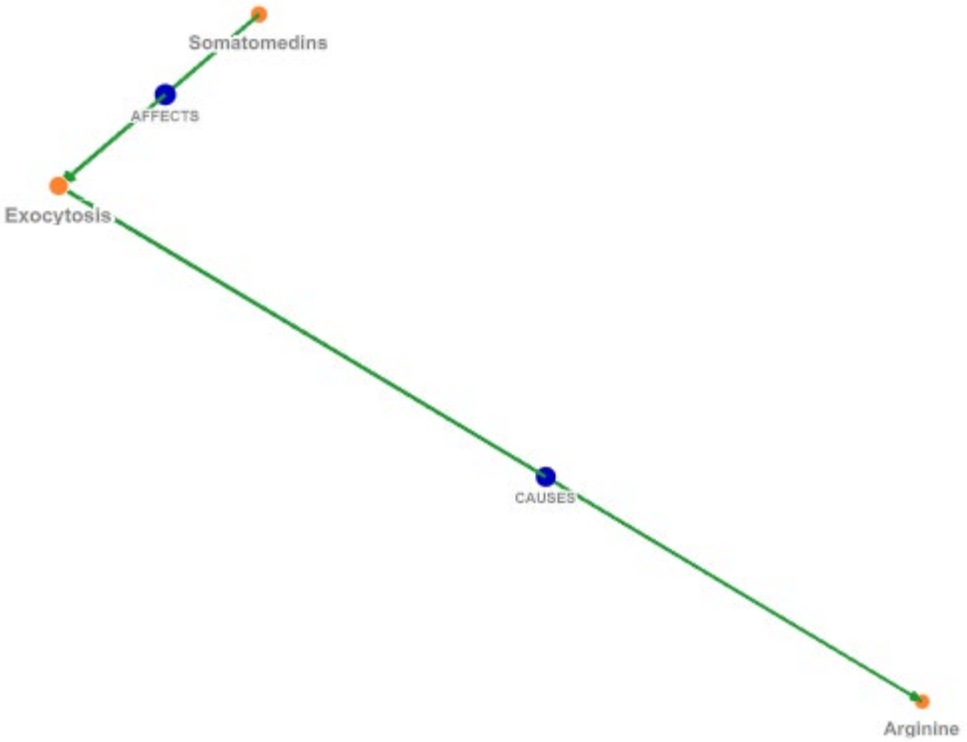


Figure C-30. Somatomedins – Arginine result 13 (no citations)

Figure C-31. Somatomedins – Arginine result 14 (no citations)

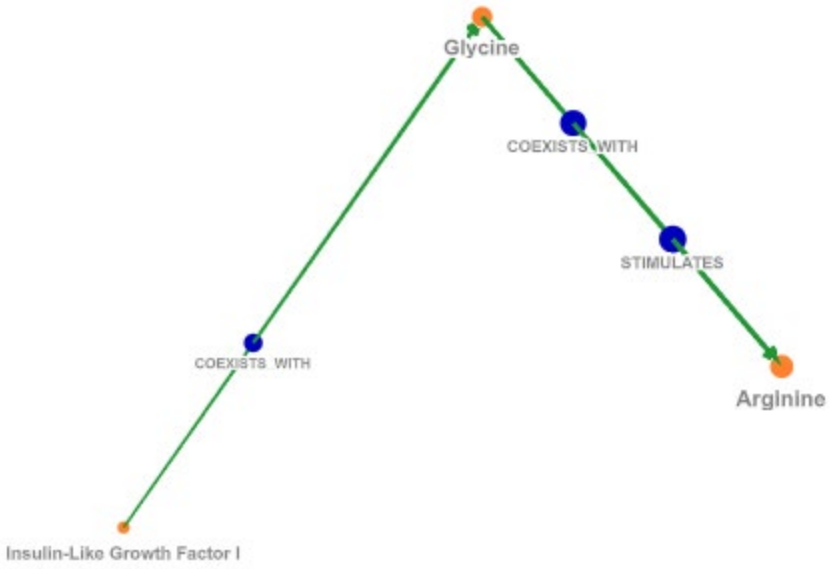


Figure C-32. Somatomedins – Arginine result 15 (no citations)

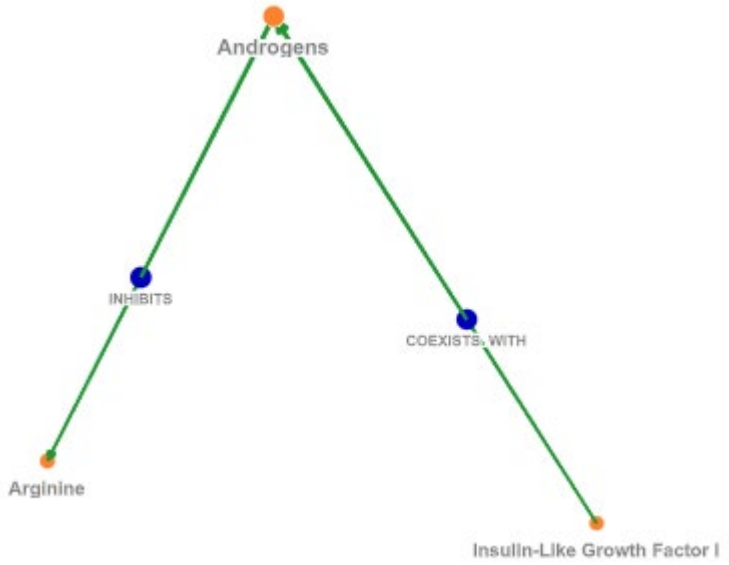


Figure C-33. Somatomedins – Arginine result 16 (no citations)

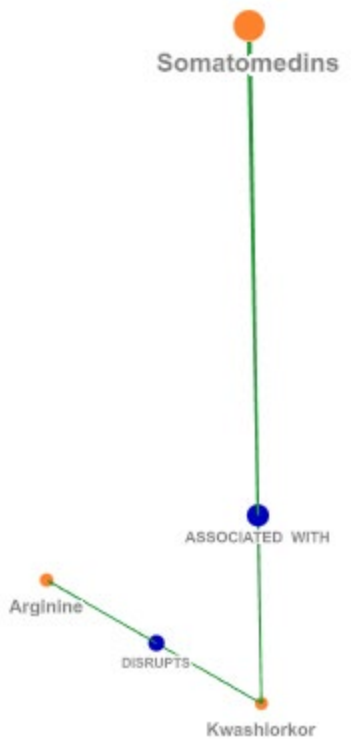


Figure C-34. Somatomedins – Arginine result 17 (no citations)

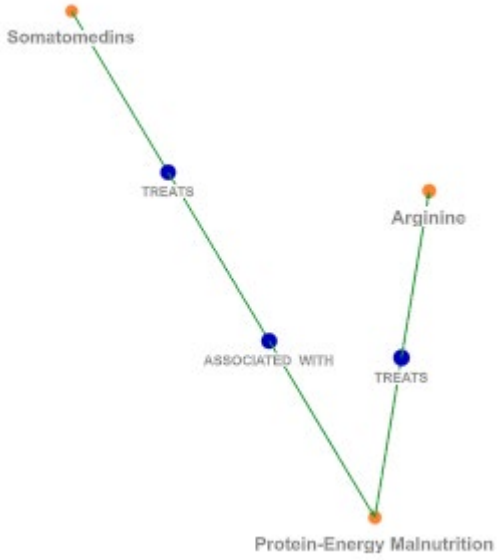


Figure C-35. Somatomedins – Arginine result 18 (no citations)

Table C-2. Somatomedins – Arginine generated predicates

[PMID105670]
Title “Plasma somatomedin activity in protein calorie malnutrition.”
[C0037657|Somatomedins – ASSOCIATED_WITH – C0033677|Protein-Energy Malnutrition]
[PMID107298]
Abstract – Taken from publisher’s site. “These data suggest that dietary composition has a direct effect on plasma SM activity and that the severe growth retardation associated with protein malnutrition may be related to its additional effect on serum SM.”
[C0037657|Somatomedins – ASSOCIATED_WITH – C0022806|Kwashiorkor]
[PMID1128255]
Abstract – Taken from publisher’s site. “Somatomedin (SM) is a growth hormone (GH)-dependent peptide, circulating in normal serum, which directly causes skeletal growth.”
[C0037657|Somatomedins – AUGMENTS – C0018270|Growth]
Abstract – Taken from publisher’s site. “It is hypothesized that GH produces skeletal growth indirectly by causing the generation of a “sulfation factor,” now called “SM,” which acts directly to cause proliferation of cartilage.”
[C0037663|Somatotropin – CAUSES – C0018270|Growth]
[PMID1128683]
Abstract – Taken from publisher’s site. “In the whole animal a major function of GH is to stimulate the longitudinal growth of the skeleton, and GH acts by controlling the production of a second series of hormones, the somatomedins.”

[C0037663|Somatotropin – AUGMENTS – C0018270|Growth]
[PMID1261514]
Abstract "When compared with the results without hormone, the addition of 1000 muU/ml of insulin per ml of medium during the 2nd to 6th period led to a significant increase in perfusate somatomedin activity at all periods."
[C0021641|Insulin – STIMULATES – C0037657|Somatomedins]
Abstract "Differences in insulin concentration may explain some clinical situations in which somatomedin concentrations cannot be correlated with GH levels."
[C0021641|Insulin – COEXISTS_WITH – C0037657|Somatomedins]
[PMID132381]
Abstract "Since diabetes mellitus is a condition in which poor growth occurs despite elevation of plasma GH, we have attempted to determine if poor growth in diabetes, as in malnutrition, could be associated with a decrease in somatomedin activity."
[C0018270|Growth – ASSOCIATED_WITH – C0011847|Diabetes] [C0018270|Growth – ASSOCIATED_WITH – C0162429|Malnutrition]
Abstract "Administration of insulin to diabetic rats 48 hours after STZ led to significant increases in SM and cartilage growth activity, and insulin therapy 24 hours after STZ prevented the decreases in SM and cartilage growth activity which occurred without insulin."
[C0021641|Insulin – INHIBITS – C0037657|Somatomedins]
[PMID13251208]
Abstract – Taken from publisher's site. "In view of the notable influence of the growth hormone of the pituitary body on the growth of cartilage knowledge of the effects of this principle on the metabolism of cartilage becomes essential."

[C0037663|Somatotropin – AFFECTS – C0018270|Growth]
[PMID142677]
Abstract – Taken from publisher's site. "Observations of a growth hormone-resistant decrease in somatomedin activity associated with conditions of insulinopenia suggest that somatomedin and growth might be related to metabolic control in diabetes mellitus."
[C0037657|Somatomedins – AFFECTS – C1513158|Metabolic Control] [C0037657|Somatomedins – ASSOCIATED_WITH – C0011849|Diabetes Mellitus]
Abstract – Taken from publisher's site. "We examined this relationship by comparing measures of insulin effect with serum somatomedin activity (porcine cartilage bioassay), cartilage growth activity (SO4 uptake in vitro), and change in body weight in streptozotocin-diabetic rats."
[C0037657|Somatomedins – INTERACTS_WITH – C0021641|Insulin]
Abstract – Taken from publisher's site. "These studies demonstrate a close relationship between insulin efficacy, serum somatomedin activity, cartilage growth activity, and weight gain and support the hypothesis that through somatomedin, insulin may contribute to growth."
[C0021641|Insulin – AFFECTS – C0018270|Growth]
[PMID20276173]
Title "The rôle of arginine in growth with some observations on the effects of argininic acid."
[C0003765|Arginine – AFFECTS – C0018270|Growth]
[PMID2408950]
Abstract "Islet cell IGF was identified as predominantly IGF-I or a closely related species and not IGF-II"
[C0021665|Insulin-Like Growth Factor I – ISA – C0037657|Somatomedins]

Abstract “Growth hormone did not consistently increase IGF-I synthesis, suggesting that the previously described effects of growth hormone on islet cell replication do not result from stimulation of IGF-I synthesis by islet cells.”
[C0037663|Somatotropin – AFFECTS – C0007590|Cell division] [C0037663|Somatotropin – STIMULATES – C0021665|Insulin-Like Growth Factor I]
Abstract “Thus, although the IGF-I synthesized by islet cells may be a physiologically relevant growth factor for these cells, the mitogenic effects of growth hormone in islet cells appear to be independent and not mediated by IGF-I.”
[C0018284|Growth Factor – INTERACTS_WITH – C0037663|Somatotropin] [C0021665|Insulin-Like Growth Factor I – INTERACTS_WITH – C0037663|Somatotropin] [C0021665|Insulin-Like Growth Factor I – ISA – C0018284|Growth Factor]
[PMID2413420]
Abstract “Insulin promotes the growth of these cells by binding, with low affinity, to the type I insulin-like growth factor (IGF) receptor, not through the high affinity insulin receptor.”
[C0021641|Insulin – AUGMENTS – C0018270|Growth]
Abstract “In other cell types, such as hepatocytes, embryonal carcinoma cells, or mammary tumor cells, the type I IGF receptor is virtually absent, and insulin stimulates the growth of these cells at physiological concentrations by binding to the high affinity insulin receptor.”
[C0021641|Insulin – AUGMENTS – C0018270|Growth] [C0140079|Insulin-Like Growth Factor Receptor – AUGMENTS – C0018270|Growth]
Abstract “Poor growth follows impaired insulin secretion in diabetes mellitus.”
[C0018270|Growth – ASSOCIATED_WITH – C0011849|Diabetes Mellitus]
Abstract “This is associated with reduced circulating levels of IGF’s which may be partly responsible for the growth failure.”

[C0037657|Somatomedins – CAUSES – C0878787|Growth failure]
Abstract “Diabetic children, treated conventionally, have normal circulating IGF levels, but both growth rate and serum IGF concentration may increase dramatically when diabetic control is optimized”
[C0037657|Somatomedins – AFFECTS – C0241863|Diabetic]
[PMID2899993]
Abstract “Growth hormone stimulates protein synthesis during hypocaloric parenteral nutrition.”
[C0037663|Somatotropin – AUGMENTS – C0597295|Protein Biosynthesis]
Abstract “The influence of growth hormone (GH) on protein metabolism and fuel utilization was investigated in eight paired studies of normal volunteers.”
[C0037663|Somatotropin – AFFECTS – C0597299|Protein Metabolism]
Abstract “GH was associated with an increase in insulin and insulin-like growth factor-I concentrations (IGF-I, 9.1 +/- 0.6 IU/ml vs. 3.3 +/- 0.5, p less than 0.001).”
[C0037663|Somatotropin – STIMULATES – C0021641|Insulin] [C0037663|Somatotropin – STIMULATES – C0021665|Insulin-Like Growth Factor I]
[PMID3099250]
Title “Serum insulin-like growth factors I and II concentrations and growth hormone and insulin responses to arginine infusion in children with protein-energy malnutrition before and after nutritional rehabilitation.”
[C0003765|Arginine – TREATS(INFER) – C0033677|Protein-Energy Malnutrition]
Abstract “GH responses to arginine were depressed in the three malnourished groups and improved significantly in marasmic-kwashiorkor and marasmic children after nutritional rehabilitation.”

[C0003765|Arginine – TREATS – C0342914|Marasmic kwashiorkor]
Abstract "Insulin responses to arginine were impaired in kwashiorkor, and marasmic-kwashiorkor children and improved significantly after refeeding."
[C0003765|Arginine – DISRUPTS – C0022806|Kwashiorkor] [C0003765|Arginine – DISRUPTS – C0342914|Marasmic kwashiorkor]
[PMID3259505]
Abstract "It was found that GH does not show any effect on the recovery of SRBC receptor, but it may act through the increase of SM level."
[C0037663|Somatotropin – STIMULATES – C0037657|Somatomedins]
[PMID3322823]
Abstract "Levels of circulating IGF I and IGF II are affected by growth hormone, but the former peptide is the more sensitive to growth hormone."
[C0037663|Somatotropin – INTERACTS_WITH – C0021665|Insulin-Like Growth Factor I]
Abstract "There is good evidence that the reduction in levels of circulating IGF I is related to decreased secretion of growth hormone that accompanies aging."
[C0021665|Insulin-Like Growth Factor - INHIBITS- C0037663|Somatotropin]
[PMID3380847]
Title "Somatomedins: do they play a pivotal role in wound healing?"
[C0037657|Somatomedins – AFFECTS – C0043240|Wound Healing]
Abstract "An understanding of the role of somatomedins in normal wound healing might be able to help us understand abnormalities of the repair process such as keloid formation."

[C0037657|Somatomedins – AFFECTS – C0043240|Wound Healing]
[PMID3690631]
Abstract "It is concluded that IGF-I is likely to be involved in the repair of the intima in injured arteries."
[C0021665|Insulin-Like Growth Factor I – AFFECTS – C0043240|Wound Healing]
[PMID3803997]
Abstract "These results show that old rats respond to GH in a mechanism involving somatomedin and that significant increases in somatic growth can be obtained, even at advanced age."
[C0037663|Somatotropin – INTERACTS_WITH – C0037657|Somatomedins]
[PMID3898005]
Abstract "Recent reports indicate that children with normal growth hormone responses who have very low integrated concentration of growth hormone may have the potential to improve their growth with growth hormone therapy."
[C0037663|Somatotropin – AFFECTS – C0018270|Growth]
[PMID3923266]
Abstract "The experiments indicate that high arginine levels in IVH solutions improve wound healing and thymic immune function following injury."
[C0003765|Arginine – AFFECTS – C0043240|Wound Healing]
[PMID412936]
Abstract "This study was undertaken to define the interrelationships of somatomedin, growth

hormone, and an inhibitor of SM in protein-calorie malnutrition.”
[C0037657|Somatomedins – ASSOCIATED_WITH – C0033677|Protein-Energy Malnutrition] [C0037663|Somatotropin – ASSOCIATED_WITH – C0033677|Protein-Energy Malnutrition]
[PMID4201417]
Abstract – Taken from publisher’s site. “Arginine infusion elicited an associated release in PRL and GH with a dissociated time course of approximately 15 min earlier for PRL than for GH.”
[C0003765|Arginine – COEXISTS_WITH – C0037663|Somatotropin]
[PMID430225]
Abstract – Taken from publisher’s site. “The diet supplemented with arginine and glycine improved growth before and after trauma, and nitrogen retention after trauma.”
[C0003765|Arginine – AFFECTS – C0018270|Growth]
[PMID430251]
Title “Inhibitory effect of dietary arginine on growth of Ehrlich ascites tumor cells in mice.”
[C0003765|Arginine – DISRUPTS – C0018270|Growth]
Abstract “The effect of dietary L-arginine on the growth and development of transplantable Ehrlich Ascites tumor cells was examined.”
[C0003765|Arginine – AFFECTS – C0018270|Growth]
Abstract “Supplemental dietary arginine at 3 or 5% did not significantly affect the growth of non-tumor bearing mice.”

[C0003765|Arginine – AFFECTS – C0018270|Growth]
Abstract “Dietary arginine may play a critical role in growth of normal as well as neoplastic tissue.”
[C0003765|Arginine – AFFECTS – C0018270|Growth]
[PMID4430939]
Title “Effect of arginine deficiency on growth and intermediary metabolism in rats.”
[C0003765|Arginine – AFFECTS – C0018270|Growth]
[PMID4506104]
Abstract “The action of growth hormone on skeletal tissue is mediated through somatomedin, a low molecular weight peptide found in serum.”
[C0037663|Somatotropin – INTERACTS_WITH – C0037657|Somatomedins]
Abstract “Since at least some of the metabolic effects of insulin on target cells are initiated by a highly specific interaction with receptors on cell membranes, this study was undertaken to determine whether somatomedin might interact with the same binding sites.”
[C0037657|Somatomedins – INTERACTS_WITH – C0021641|Insulin]
[PMID4538721]
Abstract – Taken from publisher’s site. “The significance of pituitary growth hormone (GH) for the regulation of skeletal growth is uniformly recognized.”
[C0037663|Somatotropin – AFFECTS – C0018270|Growth]
[PMID4903729]

Abstract – Taken from publisher's site. "All individuals responded to arginine infusion with an increase in the IRI concentrations."
[C0003765|Arginine – STIMULATES – C0021641|Insulin]
Abstract – Taken from publisher's site. "Four of the five females and two of the five males responded to the initial infusion of arginine with a significant increase in plasma GH concentration."
[C0003765|Arginine – STIMULATES – C0037663|Somatotropin]
Abstract – Taken from publisher's site. "It is concluded (1) that saline is not a stimulus to GH release, (2) that arginine consistently increases IRI concentration, (3) that plasma glucose levels usually, but not invariably, increase with arginine infusion and (4) that some males have a variable GH response to arginine, which is independent of estrogen replacement."
[C0003765|Arginine – STIMULATES – C0021641|Insulin]
[PMID521015]
Title "Prolactin nonresponsiveness to arginine in diabetes mellitus."
[C0003765|Arginine – ASSOCIATED_WITH – C0011849|Diabetes Mellitus]
[PMID5361695]
Abstract – Taken from publisher's site. "The apparently normal production of growth hormone and the impaired insulin, which was persistently demonstrated by these infants, along with the poor growth in cell mass of those under 11 months of age, suggest that gains in cell mass in the younger infant may be primarily dependent on insulin and not on growth hormone."
[C0018270|Growth – PRODUCES – C0021641|Insulin] [C0018270|Growth – PRODUCES – C0037663|Somatotropin]
[PMID5917038]
Abstract – Taken from publisher's site. "During the course of experience with subjects who were given infusions of arginine,

we observed that female subjects showed a greater rise in serum insulin and in serum HGH than did males of a comparable age.”
[C0003765|Arginine – COEXISTS_WITH – C0021641|Insulin] [C0003765|Arginine – COEXISTS_WITH – C0169964|Somatropin]
[PMID6027286]
Title “Effect of arginine on serum levels of insulin and growth hormone in obese subjects.”
[C0037663|Somatotropin – INTERACTS_WITH – C0003765|Arginine] [C0021641|Insulin – INTERACTS_WITH – C0003765|Arginine]
[PMID6153391]
Abstract “This peptide has been classified as a somatomedin on the basis of its biological actions and growth hormone dependence.”
[C0037657|Somatomedins – INTERACTS_WITH – C0037663|Somatotropin] [C0037663|Somatotropin – INTERACTS_WITH – C0037657|Somatomedins]
Abstract “We have generated an antibody to the synthetic 12 amino acid C-peptide region of insulin-like growth factor-I and used it to compare three somatomedin preparations to insulin-like growth factor-I.”
[C0037657|Somatomedins – compared_with – C0021665|Insulin-Like Growth Factor I]
Abstract “We also compared these somatomedin preparations to insulin-like growth factor-I using the standard SM-C RIA.”
[C0037657|Somatomedins – compared_with – C0021665|Insulin-Like Growth Factor I]
[PMID6181237]
Abstract “The possible mechanism(s) by which arginine can modify mammary growth are discussed.”
[C0003765|Arginine – AFFECTS – C0018270|Growth]

[PMID6243390]
Abstract – Taken from publisher's site. "It is now over 20 years since Salmon and Daughaday hypothesized that stimulation of growth by growth hormone (somatotropin) was mediated by a circulating 'sulfation factor'."
[C0037663|Somatotropin – AUGMENTS – C0018270|Growth]
[PMID6338399]
Abstract – Taken from publisher's site. "Plasma IGF-I levels reflect circulating concentrations of pituitary growth hormone (GH), and exogenous IGF-I substitutes for GH in inducing somatic and skeletal growth in GH-deficient rats2."
[C0021665|Insulin-Like Growth Factor I – COEXISTS_WITH – C0037663|Somatotropin]
Abstract – Taken from publisher's site. "By contrast, both GH and PL stimulate IGF-I synthesis in fibroblasts from older rats with no effect on IGF-II synthesis."
[C0037663|Somatotropin – STIMULATES – C0021665|Insulin-Like Growth Factor I]
[PMID6368579]
Abstract "During simultaneous incubation with PDGF and SM-C-deficient PPP, however, hydrocortisone, T4, EGF, and insulin produced concentration-dependent increases in IR-SM production."
[C0021641|Insulin – STIMULATES – C0037657|Somatomedins]
[PMID6385593]
Abstract "There was also a significantly enhanced increase of SM-C/IGF I in the presence of GH during culture in comparison to the controls."
[C0037663|Somatotropin – STIMULATES – C0021665|Insulin-Like Growth Factor I]

Abstract “Evidently IGF I is produced locally in bone and mediates the effect of GH on bone formation.”
[C0021665|Insulin-Like Growth Factor I – INTERACTS_WITH – C0037663|Somatotropin]
[PMID6639868]
Abstract “The acceleration resulting from growth hormone and thyroxine administration may be due to an increased production of somatomedins locally or systemically or by direct action on connective tissue.”
[C0037663|Somatotropin – PRODUCES – C0037657|Somatomedins]
[PMID6751804]
Abstract “The addition of human growth hormone (25 micrograms/ml) and/or insulin (1 mU/ml) to the perfusates caused a significant increase in somatomedin after 120 minutes of recirculation.”
[C0021641|Insulin – COEXISTS_WITH – C0037657|Somatomedins]
Abstract “The effect of insulin on the release of somatomedin was greater in hypox rats than that observed in normal rats.”
[C0037657|Somatomedins – INTERACTS_WITH- C0021641|Insulin]
[PMID6754563]
Abstract “These data suggest that glucose intolerance during arginine deficiency is related to decreased insulin release immediately following glucose administration and possibly a mild insulin resistance.”
[C0003765|Arginine – COEXISTS_WITH – C0021641|Insulin]
[PMID6986740]

Abstract – Taken from publisher's site. "The addition of GH (1 mU/ml) or insulin (0.25-1 mU/ml) to the culture dishes significantly increased the SM activity of the medium and a dose-dependent effect was observed with insulin."
[C0021641|Insulin – STIMULATES – C0037657|Somatomedins]
"These findings suggest that: 1) SM is synthesized by rat liver in organ culture; 2) insulin plays an important role in SM generation, as it itself is capable of both stimulating the release of SM and amplifying the stimulatory effect of GH."
[C0021641|Insulin – INTERACTS_WITH – C0037657|Somatomedins]
[PMID702022]
Title "Production of somatomedin activity in vitro in the presence of growth hormone and cycloheximide."
[C0037663|Somatotropin – PRODUCES – C0037657|Somatomedins]
[PMID7424474]
Abstract "The effect of growth hormone on somatomedin generation was abolished in hypophysectomized rats fed with low-protein diet."
[C0037663|Somatotropin – INTERACTS_WITH – C0037657|Somatomedins]
[PMID807324]
Title "Inhibitory effect of L-arginine on growth of rat mammary tumors induced by 7,12-dimethylbenz(a)anthracene."
[C0003765|Arginine – DISRUPTS – C0018270|Growth]
[PMID862560]
Abstract

"The roles of growth hormone and somatomedin in stimulating muscle cell proliferation were investigated in a series of experiments on myoblast growth in culture."
[C0037657|Somatomedins – AUGMENTS – C0018270|Growth] [C0037663|Somatotropin – AUGMENTS – C0018270|Growth]
[PMID895522]
Abstract – Taken from publisher's site. "Although pituitary hormones, particularly growth hormone (GH), are known to influence skeletal growth, there is no evidence for a direct effect of GH or GH-dependent factors (somatomedins) on bone as opposed to cartilage."
[C0037663|Somatotropin – AFFECTS – C0018270|Growth]
[PMID96201]
Title "Sulphation factor (somatomedin activity) in experimental protein malnutrition in the rat."
[C0037657|Somatomedins – ASSOCIATED_WITH – C0022806|Kwashiorkor]
Abstract "In a rat model of protein malnutrition in which the failure of growth is a major feature, a low level of bioassayable sulphation factor activity was present in the serum, associated with normal levels of growth hormone and low insulin in the plasma."
[C0021641|Insulin – ASSOCIATED_WITH – C0022806|Kwashiorkor] [C0037663|Somatotropin – ASSOCIATED_WITH – C0022806|Kwashiorkor]
Abstract "The administration of pharmacological doses of human or bovine growth hormone did not increase the amount of sulphation factor activity in the serum or the width of the tibial epiphyses in the protein-malnourished animals."
[C0520986|growth hormone, bovine – NEG_STIMULATES – C0037657|Somatomedins]
[PMID967246]
Abstract – Taken from publisher's site. "SOMATOMEDINS are growth hormone-dependent polypeptides that have been

proposed as the mediators of the peripheral actions of growth hormone on skeletal tissue.”
[C0037657|Somatomedins – STIMULATES – C0037663|Somatotropin]

Table C-3. Precision and recall using citations for Somatomedins study

Result	New Intermediate Terms	Relevancy (Topic)	Precision	Recall
d0	Growth	Body Weight	1.0	0.2
d1	Somatotropin	Growth Hormone	1.0	0.4
d2	Insulin	Diabetes	1.0	0.6
d3	Glucose	Diabetes	1.0	0.6
d4	Wound Healing	Wound Healing	1.0	0.8
d5	Ornithine	Not Relevant	0.8	0.8
d6	Testosterone	Not Relevant	0.667	0.8
d7	Protein Biosynthesis	Growth Hormone	0.667	0.8
d8	Diabetes Mellitus	Diabetes	0.667	0.8
d9	Somatropin	Growth Hormone	0.667	0.8
d10	Somatostatin	Growth Hormone	0.667	0.8
d11	Exocytosis	Not Relevant	0.571	0.8
d12	Leucine	Not Relevant	0.5	0.8
d13	Androgens	Not Relevant	0.444	0.8
d14	Kwashiorkor	Malnutrition	0.5	1.0
d15	Protein-Energy Malnutrition	Malnutrition	0.5	1.0

Table C-4. Precision and recall not using citations for Somatomedins study

Result	New Intermediate Terms	Relevancy (Topic)	Precision	Recall
d0	Growth	Body Weight	1.0	0.2
d1	Somatotropin	Growth Hormone	1.0	0.4
d2	Insulin	Diabetes	1.0	0.6
d3	Glucose	Diabetes	1.0	0.6
d4	Wound Healing	Wound Healing	1.0	0.8
d5	Infant	Not Relevant	0.8	0.8
d6	Ornithine	Not Relevant	0.667	0.8
d7	Testosterone	Not Relevant	0.571	0.8
d8	Protein Biosynthesis	Growth Hormone	0.571	0.8
d9	Diabetes Mellitus	Diabetes	0.571	0.8
d10	Somatropin	Growth Hormone	0.571	0.8
d11	Somatostatin	Growth Hormone	0.571	0.8
d12	Exocytosis	Not Relevant	0.5	0.8
d13	Leucine	Not Relevant	0.444	0.8
d14	Glycine	Not Relevant	0.4	0.8
d15	Androgens	Not Relevant	0.364	0.8
d16	Kwashiorkor	Malnutrition	0.417	1.0
d17	Protein-Energy Malnutrition	Malnutrition	0.417	1.0

Appendix D

Testosterone – Sleep Study

Table D-1. Testosterone – Sleep Core Corpus Medline IDs

1740593	10484567	14687700	17201728	19647784
4321505	10566905	15714228	17253619	19684340
4326799	10628505	15724043	17520786	19738366
6348068	10774867	15728214	18035227	19769952
6405703	10785345	15743338	18248637	20026350
6781027	10938176	15841103	18413429	20171018
6822642	11238497	15914523	18519168	20497841
7089154	11297573	16453977	18710368	20675620
7440701	12055986	16453985	19075717	
9329339	12843160	16890403	19212126	
10344587	12915643	17032928	19602101	

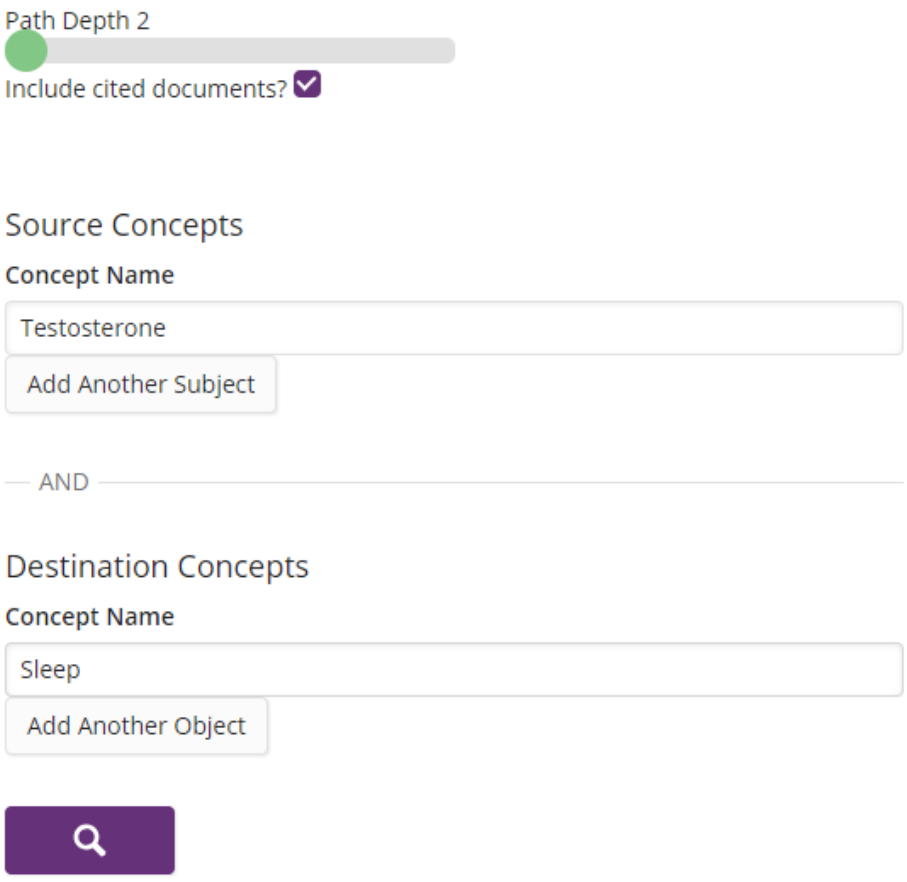


Figure D-1. Testosterone – Sleep search parameters

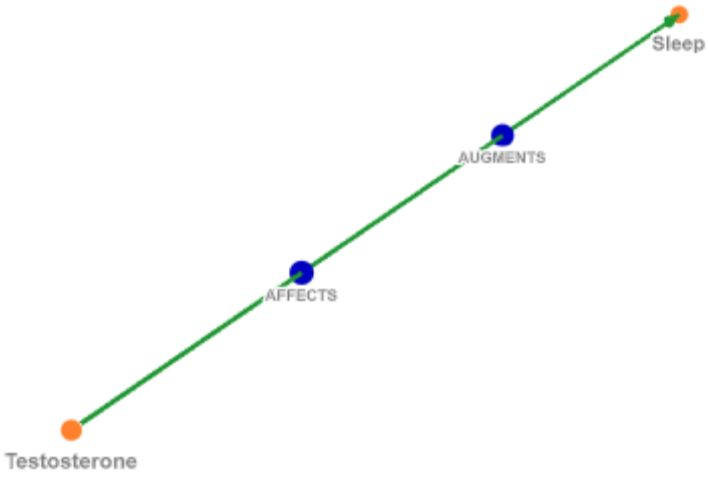


Figure D-2. Testosterone – Sleep result 1

Figure D-3. Testosterone – Sleep result 2

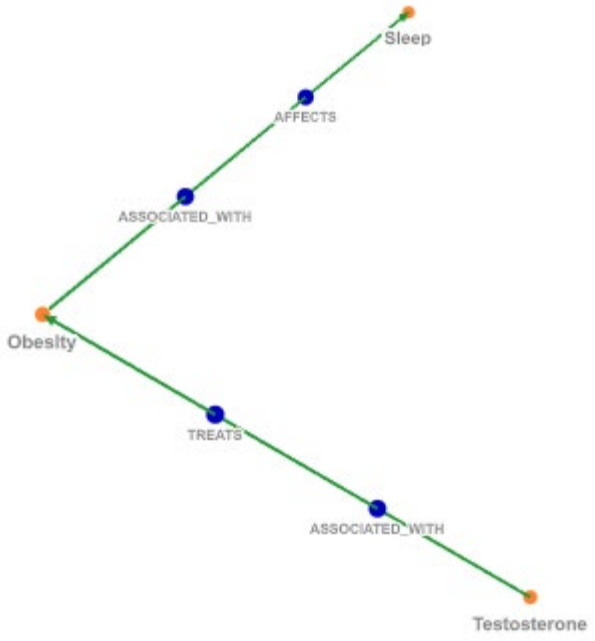


Figure D-4. Testosterone – Sleep result 3

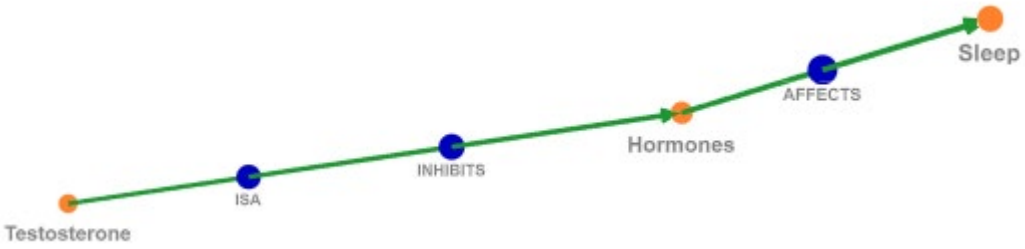


Figure D-5. Testosterone – Sleep result 4

Figure D-6. Testosterone – Sleep result 5

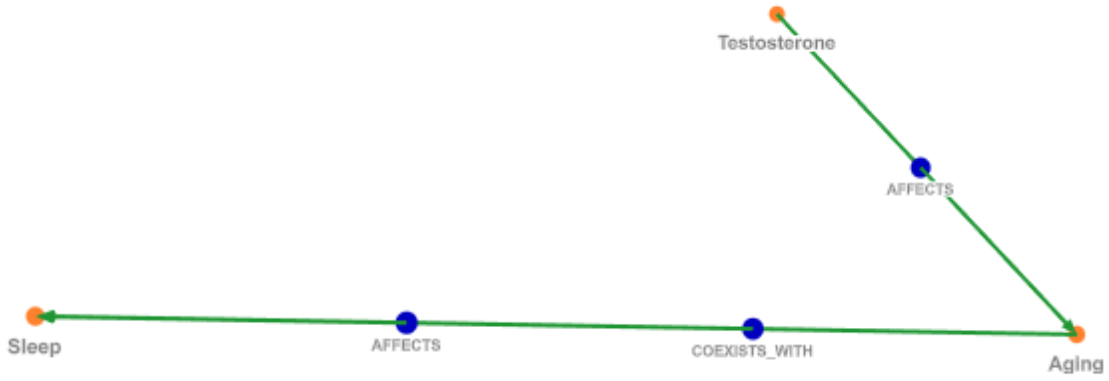


Figure D-7. Testosterone – Sleep result 6

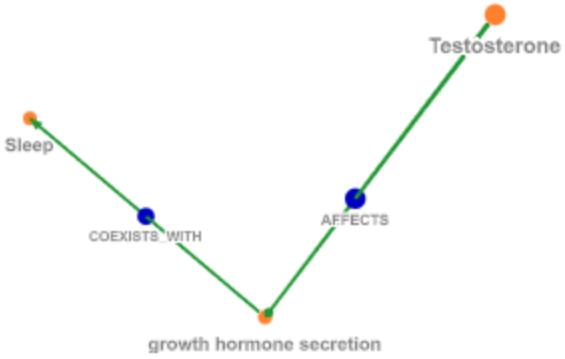


Figure D-8. Testosterone – Sleep result 7

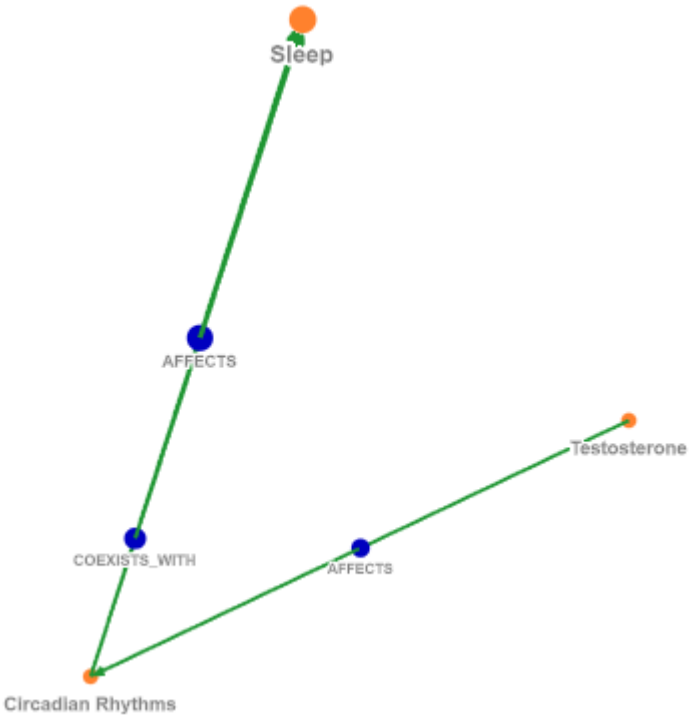


Figure D-9. Testosterone – Sleep result 8

Figure D-10. Testosterone – Sleep result 9

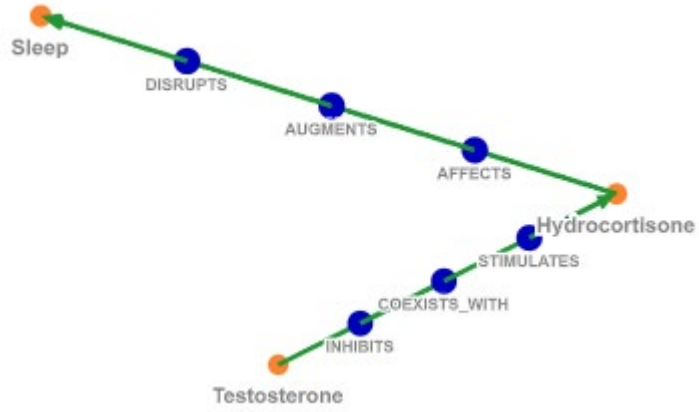


Figure D-11. Testosterone – Sleep result 10

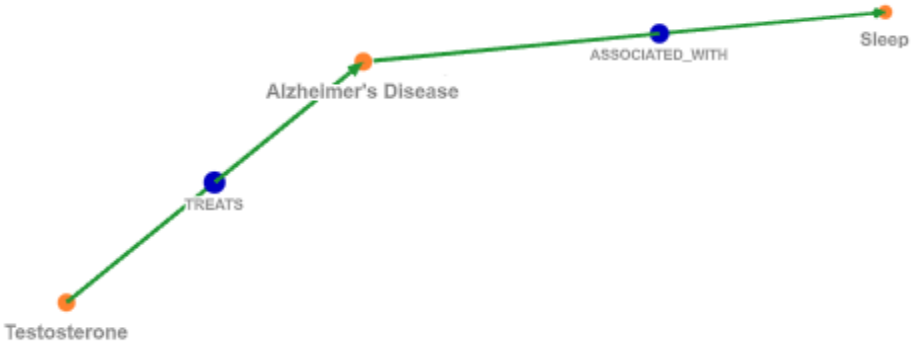


Figure D-12. Testosterone – Sleep result 11

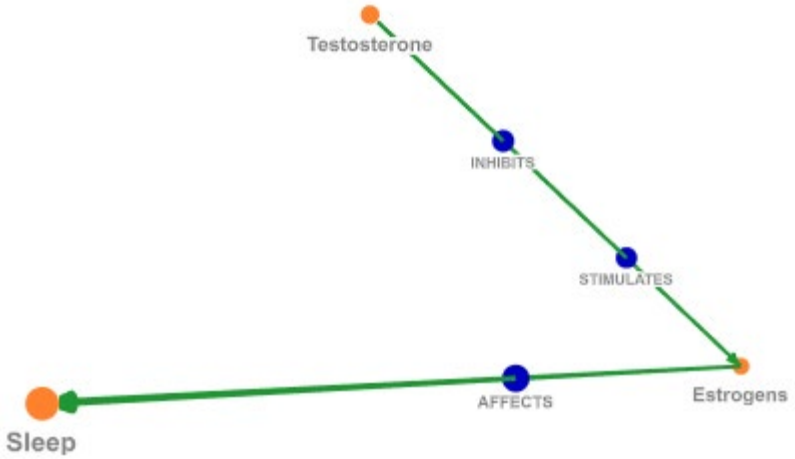


Figure D-13. Testosterone – Sleep result 12

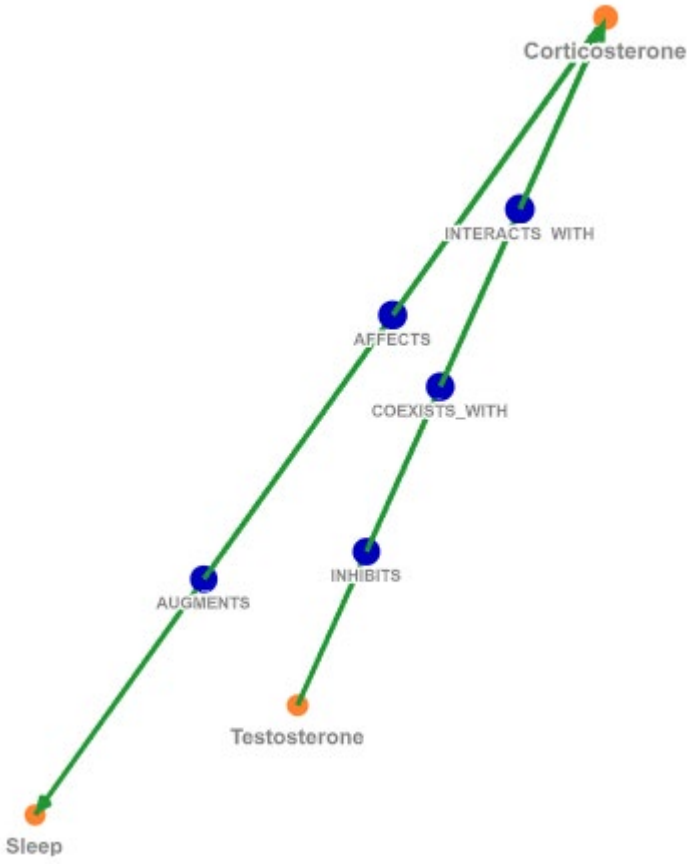


Figure D-14. Testosterone – Sleep result 13

Table D-2. Testosterone – Sleep generated predicates

[PMID10841212]
Abstract "The study compared sleep, cortisol, and sleep-cortisol correlations under baseline and "stress" conditions in men and women."
[C0020268|Hydrocortisone – AFFECTS – C0037313|Sleep]
[PMID10956236]
Title "IL-1 is a mediator of increases in slow-wave sleep induced by CRH receptor blockade."
[C0021753|Interleukin-1 beta – AUGMENTS – C0234451|Sleep, Slow-Wave]
Abstract "IL-1 promotes sleep, and glucocorticoids inhibit IL-1 synthesis."
[C0017710|Glucocorticoids – AUGMENTS – C0037313|Sleep] [C0021753|Interleukin-1 beta – AUGMENTS – C0037313|Sleep]
[PMID12000022]
Title "The role of cytokines in physiological sleep regulation."
[C0079189|cytokine – AFFECTS – C0037313|Sleep]
Abstract "Several growth factors (GFs) are implicated in sleep regulation."
[C0018284|Growth Factor – AFFECTS – C0037313|Sleep]
Abstract "Among these substances, the most extensively studied for their role in sleep regulation are interleukin-1beta (IL-1) and tumor necrosis factor alpha (TNF)."
[C0021753|Interleukin-1 beta – AFFECTS – C0037313|Sleep] [C1456820|Tumor Necrosis Factor-alpha – AFFECTS – C0037313|Sleep]

Abstract “Inhibition of either IL-1 or TNF inhibits spontaneous sleep and the sleep rebound that occurs after sleep deprivation.”
[C0021753|Interleukin-1 beta – DISRUPTS – C0037313|Sleep] [C1456820|Tumor Necrosis Factor-alpha – DISRUPTS – C0037313|Sleep]
Abstract “IL-1 and TNF are part of a complex biochemical cascade regulating sleep.”
[C0021753|Interleukin-1 beta – AFFECTS – C0037313|Sleep] [C1456820|Tumor Necrosis Factor-alpha – AFFECTS – C0037313|Sleep]
[PMID12531038]
Title “Treatment of sleep disturbance in Alzheimer’s disease”
[C0037317|Sleep disturbances – COEXISTS_WITH – C0002395|Alzheimer’s Disease]
Abstract “Nevertheless, the study of sleep in AD is relatively new.”
[C0037313|Sleep – ASSOCIATED_WITH – C0002395|Alzheimer’s Disease]
[PMID15240608]
Abstract “Testosterone has immune-modulating properties, and current in vitro evidence suggests that testosterone may suppress the expression of the proinflammatory cytokines TNFalpha, IL-1beta, and IL-6 and potentiate the expression of the antiinflammatory cytokine IL-10.”
[C0039601|Testosterone – INHIBITS – C0021753|Interleukin-1 beta] [C0039601|Testosterone – INHIBITS – C0021760|Interleukin-6] [C0039601|Testosterone – STIMULATES – C0085295|Interleukin-10]
[PMID15724043]
Abstract “Unlike the norm in the adult, where testosterone production is often inhibited by cortisol, in the fetus there is a positive link between the two.”

[C0020268|Hydrocortisone – INHIBITS – C0039601|Testosterone]
[PMID15841103]
Title “Testosterone suppression of CRH-stimulated cortisol in men.”
[C0039601|Testosterone – INHIBITS – C0020268|Hydrocortisone]
Abstract “These data demonstrate that testosterone regulates CRH-stimulated HPA axis activity in men, with the divergent effects on ACTH and cortisol suggesting a peripheral (adrenal) locus for the suppressive effects on cortisol.”
[C0039601|Testosterone – STIMULATES – C0001655|Corticotropin] [C0039601|Testosterone – STIMULATES – C0020268|Hydrocortisone]
[PMID15886244]
Abstract “Despite the absence of sex differences in estradiol or testosterone at the time of testing, men showed increased stimulated ACTH (repeated-measures ANOVA for CRH, P < 0.005) and cortisol (repeated-measures ANOVA for exercise, P < 0.05) compared with women.”
[C0014912|Estradiol – STIMULATES – C0010132|Corticotropin-Releasing Hormone] [C0014912|Estradiol - STIMULATES – C0020268|Hydrocortisone] [C0039601|Testosterone – STIMULATES – C0010132|Corticotropin-Releasing Hormone] [C0039601|Testosterone – STIMULATES – C0020268|Hydrocortisone]
[PMID15914523]
Abstract “In conclusion, testosterone increased during sleep and fell during waking, whereas circadian effects seemed marginal.”
[C0039601|Testosterone – AUGMENTS – C0037313|Sleep]
[PMID16344336]

Title “Effects of testosterone on cognition and mood in male patients with mild Alzheimer disease and healthy elderly men.”
[C0039601|Testosterone – TREATS(INFER) – C0002395|Alzheimer’s Disease]
[PMID18519168]
Title “The effects of testosterone on sleep and sleep-disordered breathing in men: its bidirectional interaction with erectile function”
[C0039601|Testosterone – AFFECTS – C0035203|Respiration] [C0039601|Testosterone – AFFECTS – C0037313|Sleep]
[PMID2541159]
Title “Influences of corticotropin-releasing hormone, adrenocorticotropin, and cortisol on sleep in normal man.”
[C0001655|Corticotropin – AFFECTS – C0037313|Sleep] [C0010132|Corticotropin-Releasing Hormone – AFFECTS – C0037313|Sleep] [C0020268|Hydrocortisone – AFFECTS – C0037313|Sleep]
[PMID3661052]
Title “Differential effects of hydrocortisone, fluocortolone, and aldosterone on nocturnal sleep in humans.”
[C0020268|Hydrocortisone – AFFECTS – C0037313|Sleep]
Abstract “The results demonstrate differential effects of synthetic glucocorticoid, cortisol, and aldosterone on sleep in humans, which may be attributed to the heterogeneity of corticosteroid receptors in the brain.”
[C0020268|Hydrocortisone – AFFECTS – C0037313|Sleep]
[PMID3790626]

Abstract “The pattern of nocturnal cortisol secretion appeared to be synchronized with the periodicity of sleep: rapid eye movement (REM) sleep was found to be primarily present when cortisol concentrations were decreasing, indicating a diminished or absent secretory activity of the adrenals at that time; wakefulness and Stage 1 sleep, by contrast, were associated with increasing plasma cortisol concentrations.”
[C0020268|Hydrocortisone – AFFECTS – C0037313|Sleep]
[PMID6822642]
Title “Cortisol secretion is inhibited during sleep in normal man.”
[C0020268|Hydrocortisone – DISRUPTS – C0037313|Sleep]
Abstract “In order to test the hypothesis that cortisol secretion is inhibited during sleep, six healthy young men (ages 18-24) were studied in a 4-day protocol.”
[C0020268|Hydrocortisone – DISRUPTS – C0037313|Sleep]
[PMID8077308]
Title “Corticotropin-releasing hormone-induced adrenocorticotropin and cortisol secretion depends on sleep and wakefulness.”
[C0001655|Corticotropin – AUGMENTS – C0037313|Sleep] [C0001655|Corticotropin – AUGMENTS – C0043012|Wakefulness] [C0020268|Hydrocortisone - AUGMENTS – C0037313|Sleep] [C0020268|Hydrocortisone – AUGMENTS – C0043012|Wakefulness]
[PMID8548511]
Abstract “Cortisol secretion is also related to sleep processes with the 24 hr nadir occurring, like the sleep dependent GH secretory surge, during the first half of nocturnal sleep.”
[C0020268|Hydrocortisone – AUGMENTS – C0037313|Sleep]

Abstract “Age-related changes in sleep-dependent secretion of GH and cortisol correlated significantly (r = .47, r = -.55, respectively; p < .05) with an age-dependent decrease in slow wave sleep.”
[C0020268|Hydrocortisone – AUGMENTS – C0234451|Sleep, Slow-Wave] [C0037663|Somatotropin – AUGMENTS – C0234451|Sleep, Slow-Wave]
Abstract “Both changes in GH and cortisol secretion may act together to reduce anabolic functions of sleep in the aged.”
[C0020268|Hydrocortisone – DISRUPTS – C0037313|Sleep] [C0037663|Somatotropin – DISRUPTS – C0037313|Sleep]
[PMID9329339]
Abstract “The possible role of gonadal steroids in regulating sleep and circadian rhythms in humans has received relatively little attention despite the importance of the topic to several clinical syndromes.”
[C0036884|Gonadal Steroid Hormones – AFFECTS – C0037313|Sleep]
Abstract “These results indicate that testosterone has relatively specific and discrete effects on sleep and hormonal rhythms in men.”
[C0039601|Testosterone – AFFECTS – C0037313|Sleep]
[PMID9530230]
Abstract “IL-1 beta dose dependently increased non-rapid eye movement sleep (NREMS) and suppressed rapid eye movement sleep (REMS) in the controls.”
[C0021753|Interleukin-1 beta – AUGMENTS – C0037322|Sleep, REM] [C0021753|Interleukin-1 beta – AUGMENTS – C0234451|Sleep, Slow-Wave]
Abstract “These results 1) provide further evidence that IL-1 beta is involved in sleep regulation, 2) indicate that the effects of IL-1 beta on sleep are mediated by the type I receptor, and 3) suggest that TNF-alpha is capable of inducing sleep without the involvement of IL-1”

[C0021753|Interleukin-1 beta – AFFECTS – C0037313|Sleep|orgf|orgf|| [C0597357|receptor – INTERACTS_WITH – C0021753|Interleukin-1 beta]
[PMID956348]
Abstract "It was concluded that administration of cortisol leading to plasma levels as seen under treatment with ACTH suppresses testosterone by abolishing or flattening the nocturnal rise."
[C0020268|Hydrocortisone – INHIBITS – C0039601|Testosterone]
[PMID9781819]
Abstract "Much evidence implicates interleukin-1beta (IL-1beta) in sleep regulation."
[C0021753|Interleukin-1 beta – AFFECTS – C0037313|Sleep]

Table D-3. Precision and recall using citations for Testosterone study

Result	New Intermediate Terms	Relevancy (Topic)	Precision	Recall
d0	N/A – Direct Connection			
d1	Gonadal Steroid Hormones	Hormones	1.0	0.333
d2	Obesity		0.5	0.333
d3	Hormones	Hormones	0.5	0.333
d4	Parkinson Disease		0.333	0.333
d5	Aging	Hormones	0.333	0.333
d6	Growth Hormone Secretion	Hormones	0.333	0.333
d7	Circadian Rhythms	Hormones	0.333	0.333
d8	Interleukin-1 beta	IL-1 Beta	0.5	0.667
d9	Hydrocortisone	Hydrocortisone	0.6	1

Table D-4. Precision and recall not using citations for Testosterone study

Result	New Intermediate Terms	Relevancy (Topic)	Precision	Recall
d0	N/A – Direct Connection			
d1	Gonadal Steroid Hormones	Hormones	1.0	0.333
d2	Obesity		0.5	0.333
d3	Hormones	Hormones	0.5	0.333
d4	Parkinson Disease		0.333	0.333
d5	Aging	Hormones	0.333	0.333
d6	Growth Hormone Secretion	Hormones	0.333	0.333
d7	Circadian Rhythms	Hormones	0.333	0.333
d8	Interleukin-1 beta	IL-1 Beta	0.5	0.667
d9	Hydrocortisone	Hydrocortisone	0.6	1

Appendix E

Parkinson's Disease – Crohn's Disease Study

Table E-1. PD – CD Core Corpus Medline IDs

1374941	16773678	18430617	19093628	19416630
7834804	17062028	18456695	19120480	19470958
8739179	17268543	18489401	19221313	19481107
12105943	18034589	18535002	19221317	19524782
15342082	18242711	18586394	19265190	19568370
15370396	18274903	18628756	19289138	19723294
15784530	18308354			

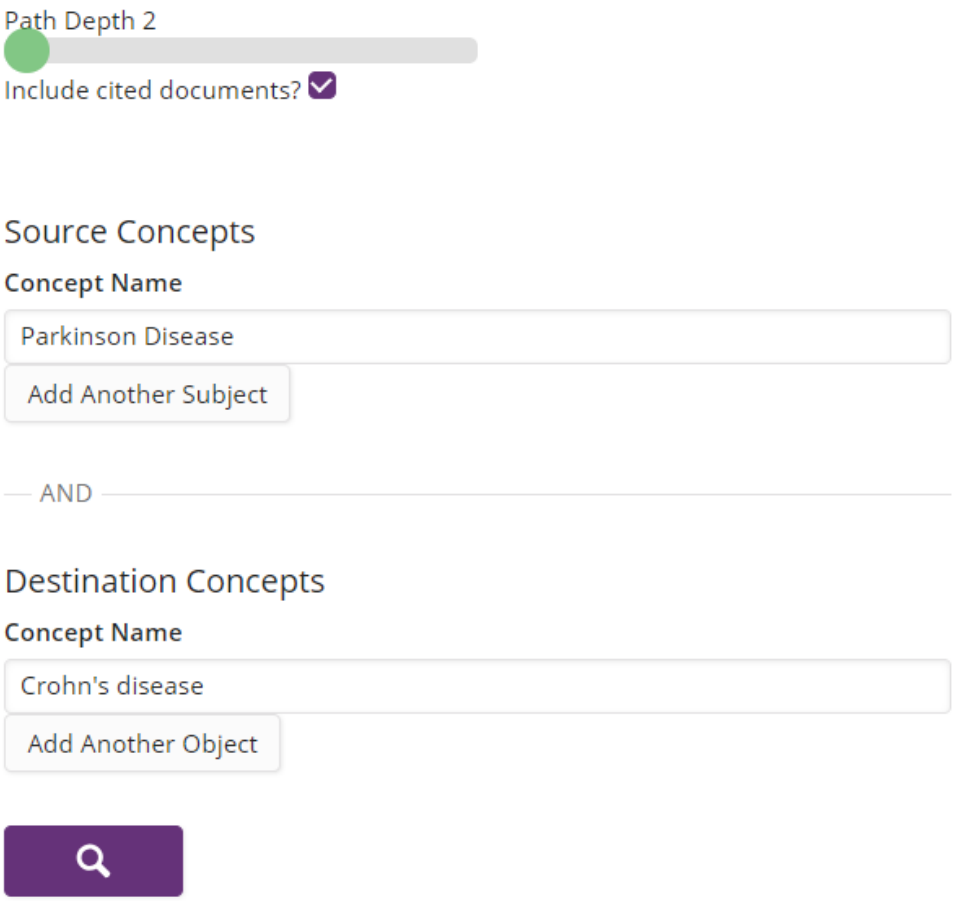


Figure E-1. PD – CD, PL2(path length two) search parameters

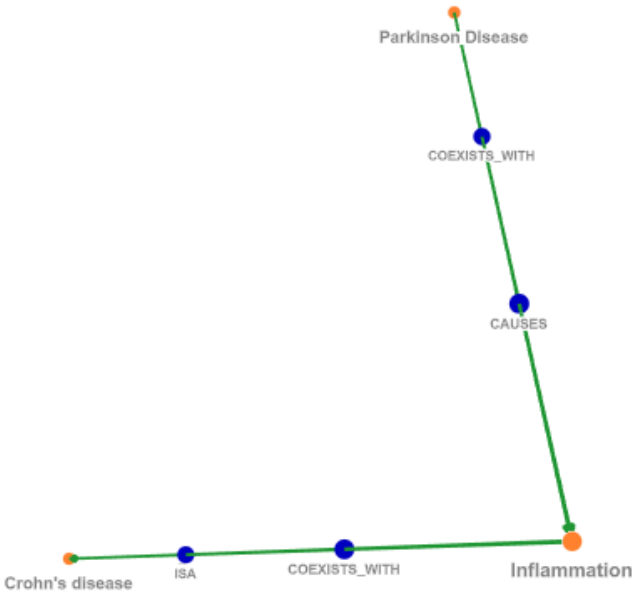


Figure E-2. PD – CD, PL2 result 1

Figure E-3. PD – CD, PL2 result 2

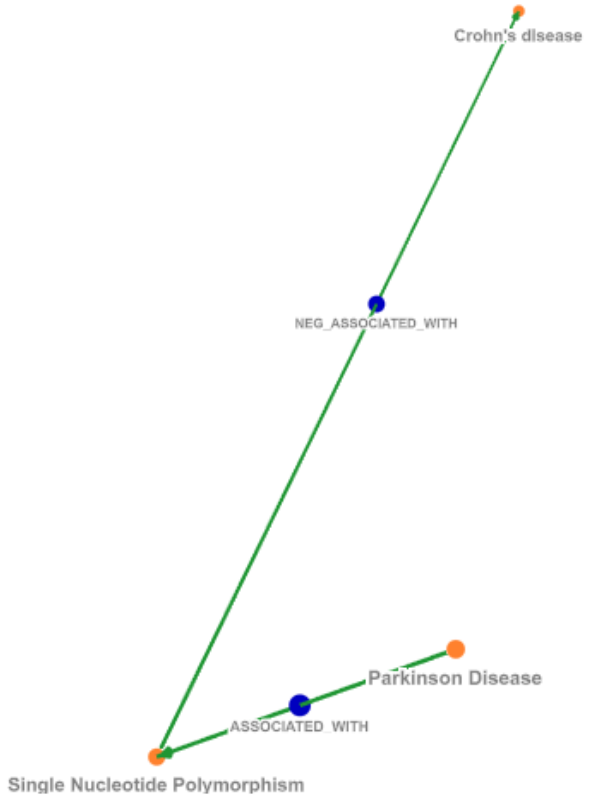


Figure E-4. PD – CD, PL2 result 3

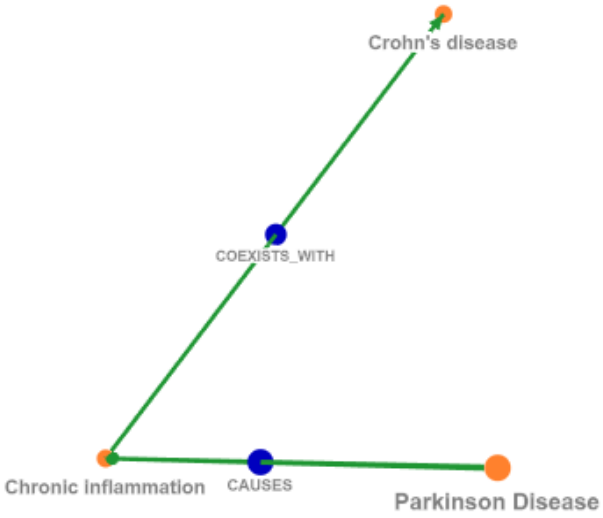


Figure E-5. PD – CD, PL2 result 4

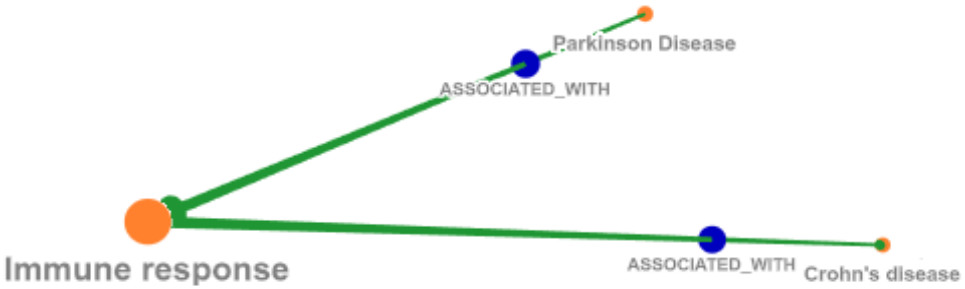


Figure E-6. PD – CD, PL2 result 5

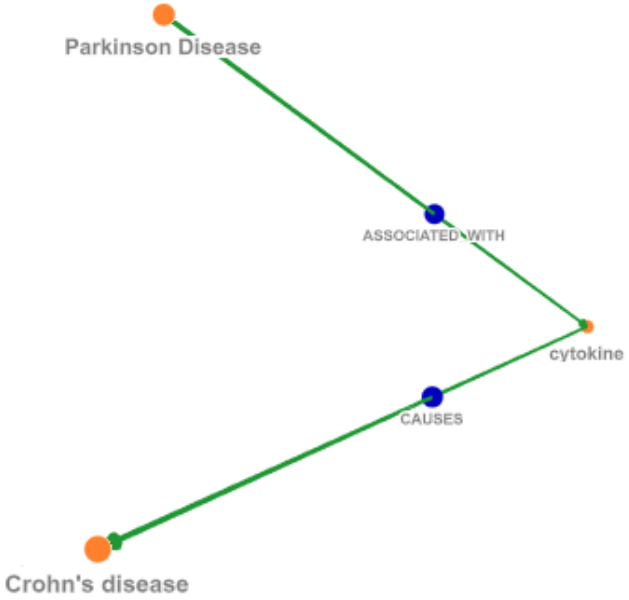


Figure E-7. PD – CD, PL2 result 6

Figure E-8. PD – CD, PL2 result 7

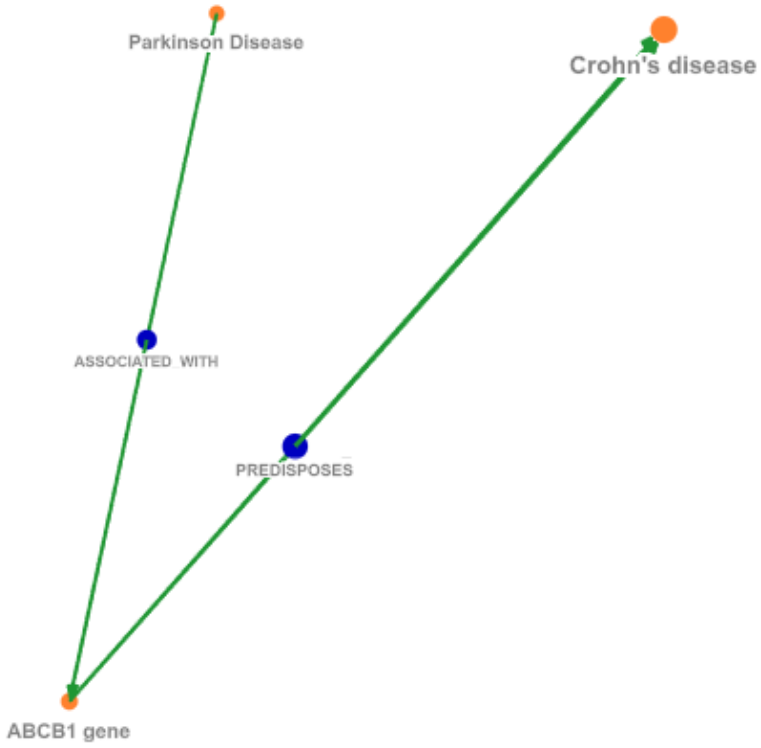


Figure E-9. PD – CD, PL2 result 8

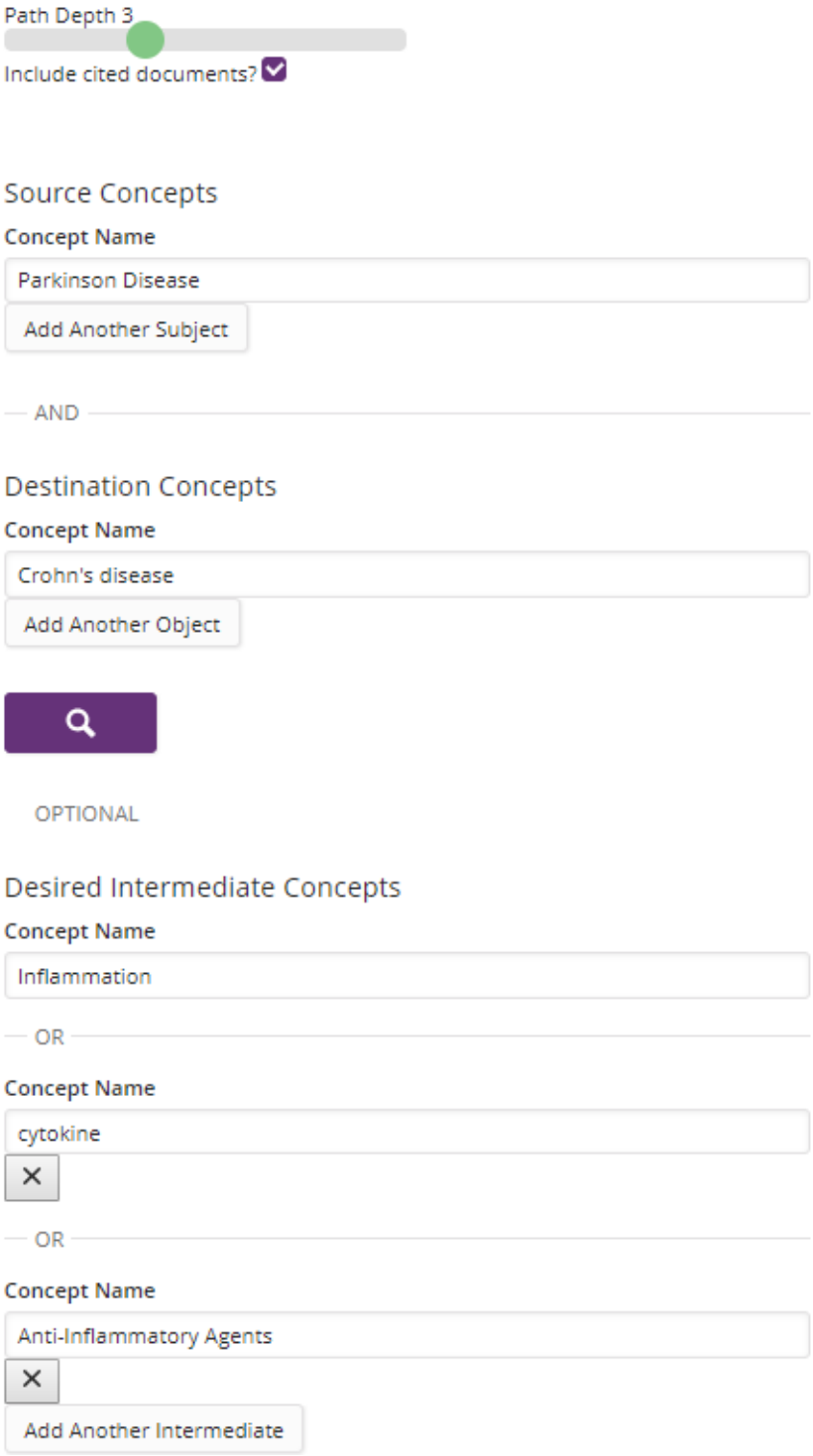


Figure E-10. PD – CD, PL3 Inflammation search parameters

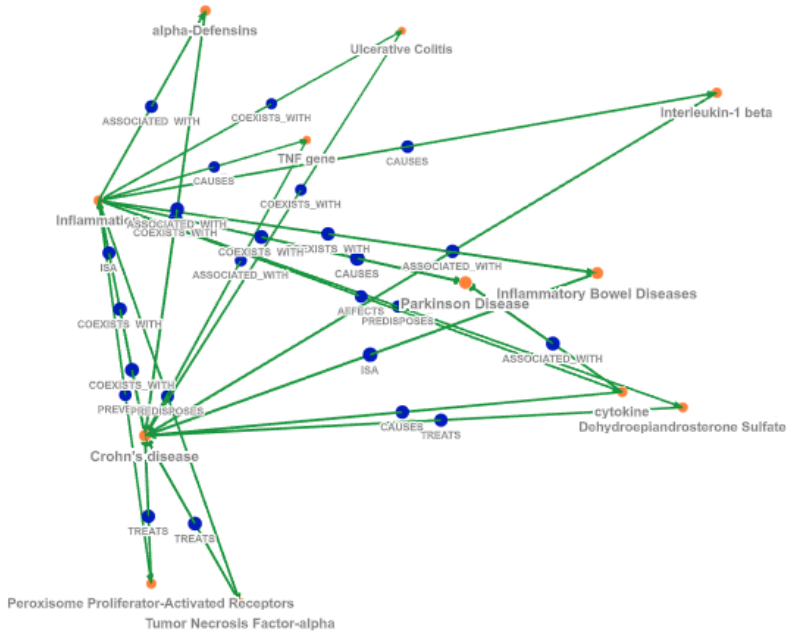


Figure E-11. PD – CD, PL3 Inflammation result 1 (with citations)

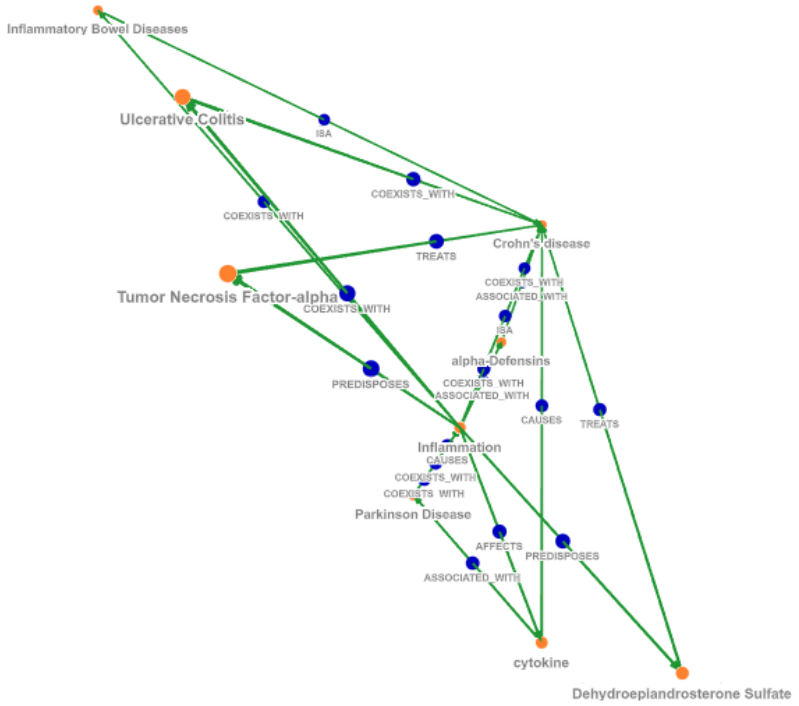


Figure E-12. PD – CD, PL3 Inflammation result 1 (without citations)

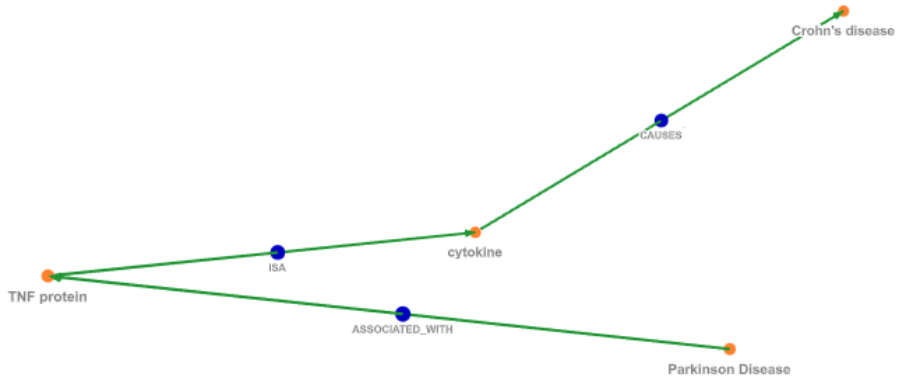


Figure E-13. PD – CD, PL3 Inflammation result 2 (with citations)

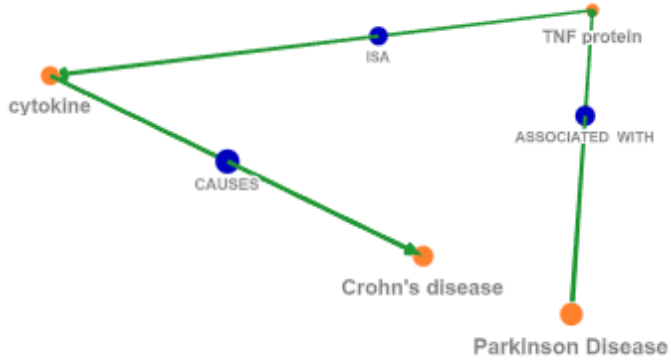


Figure E-14. PD – CD, PL3 Inflammation result 2 (without citations)

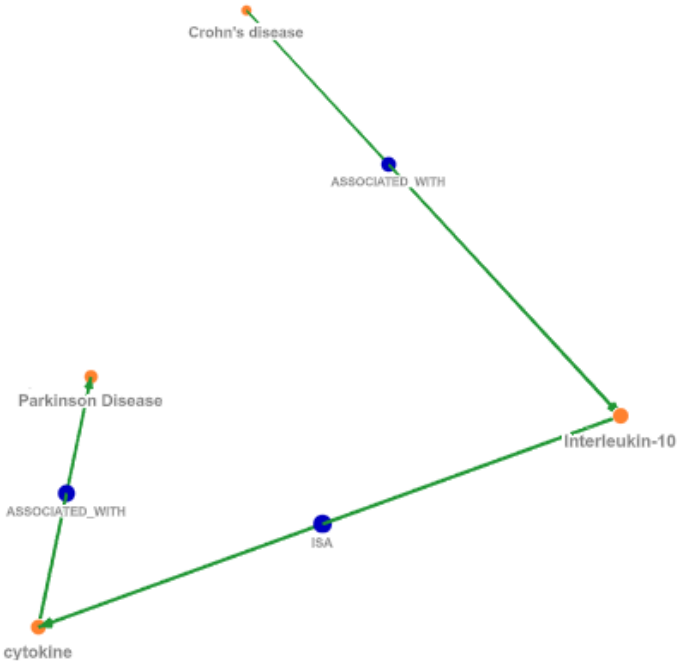


Figure E-15. PD – CD, PL3 Inflammation result 3 (with citations)

<Not Present>

Figure E-16. PD – CD, PL3 Inflammation result 3 (without citations)

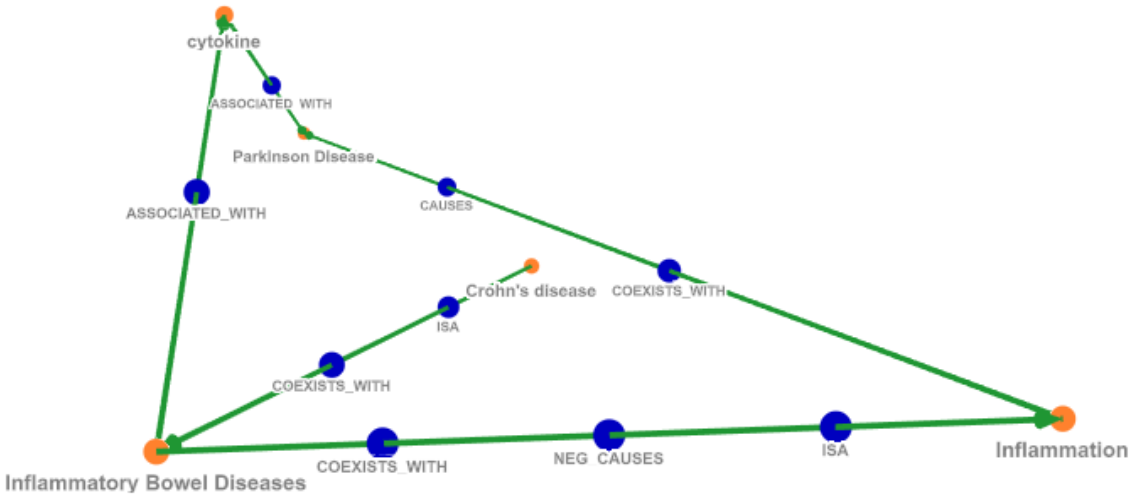


Figure E-17. PD – CD, PL3 Inflammation result 4 (without citations)

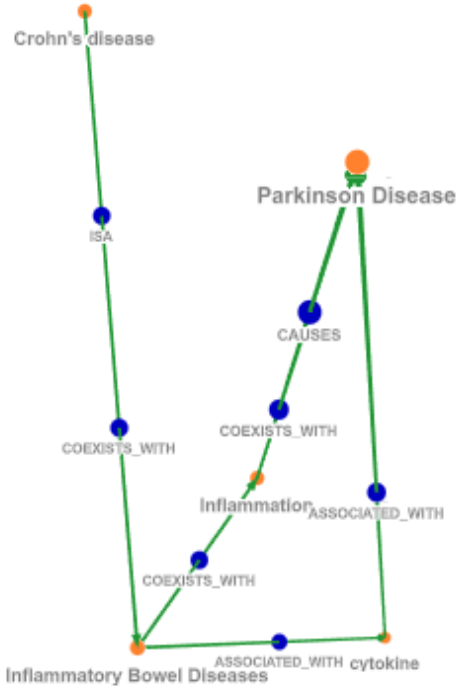


Figure E-18. PD – CD, PL3 Inflammation result 4 (without citations)

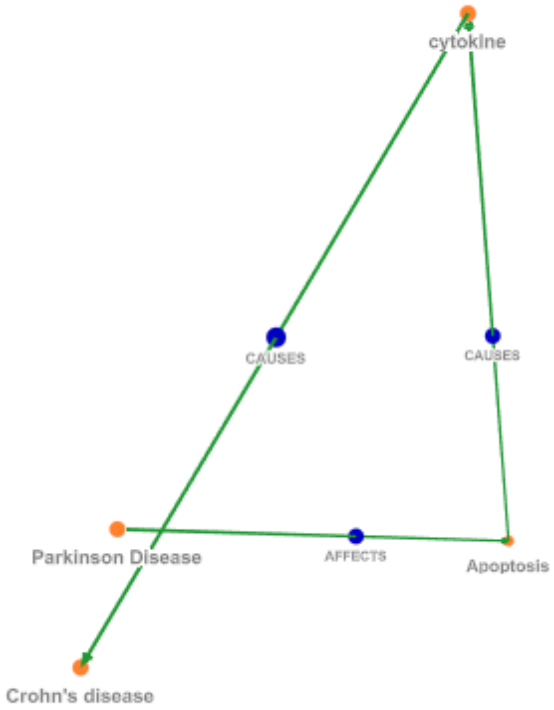


Figure E-19. PD – CD, PL3 Inflammation result 5 (with citations)

<Not Present>

Figure E-20. PD – CD, PL3 Inflammation result 5 (without citations)

Figure E-21. PD – CD, PL3 Inflammation result 6 (with citations)

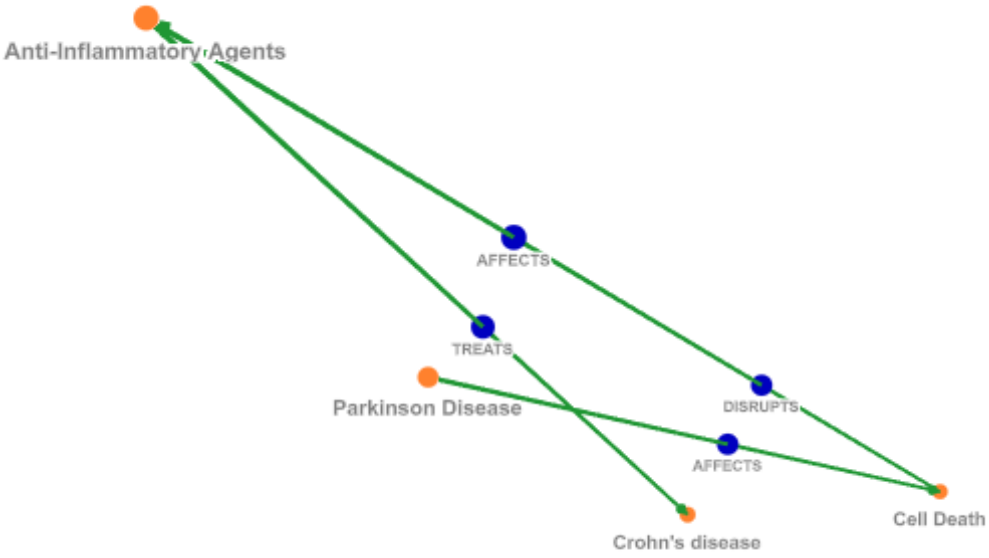


Figure E-22. PD – CD, PL3 Inflammation result 6 (without citations)

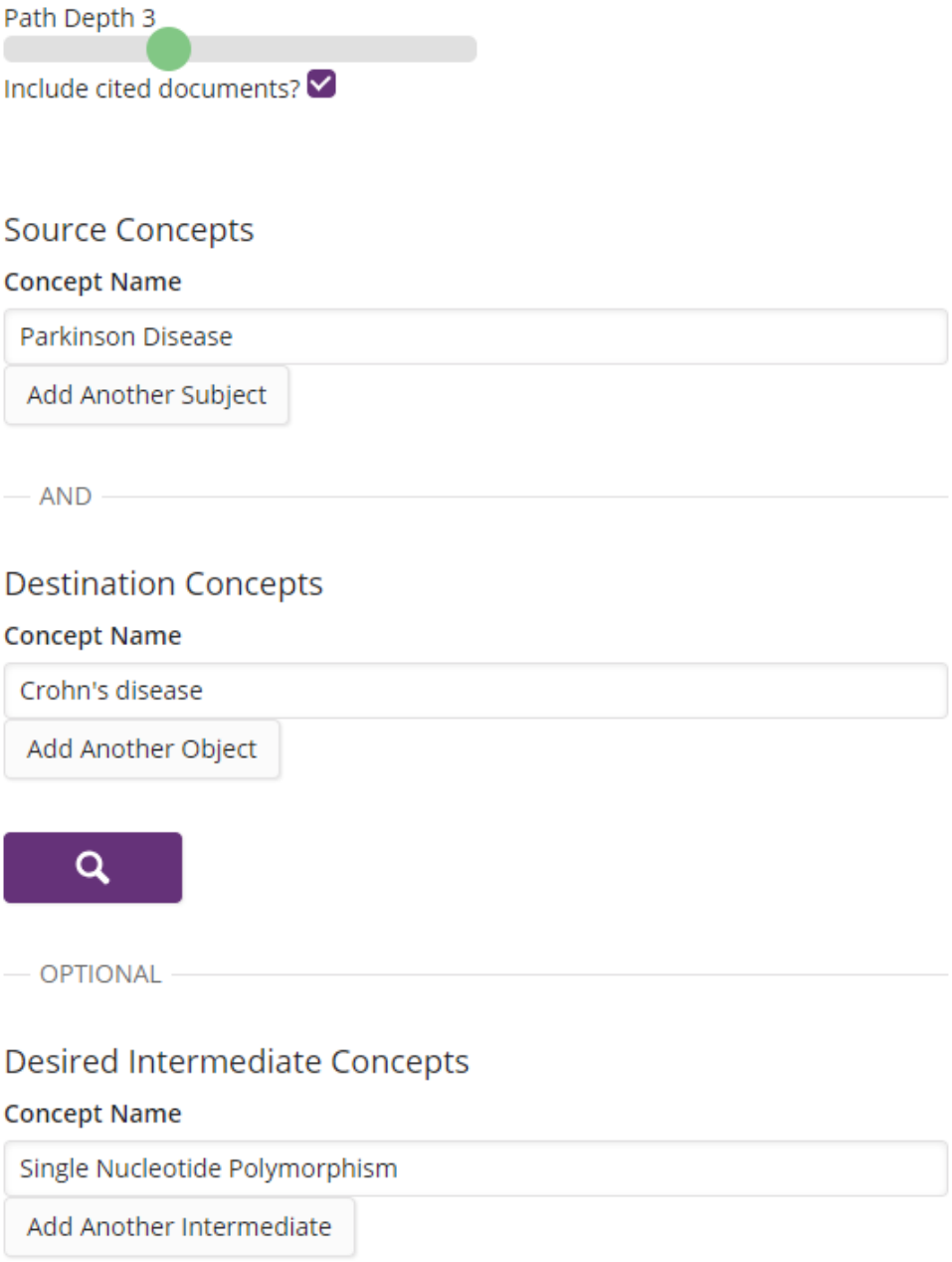


Figure E-23. PD – CD, PL3 SNP search parameters

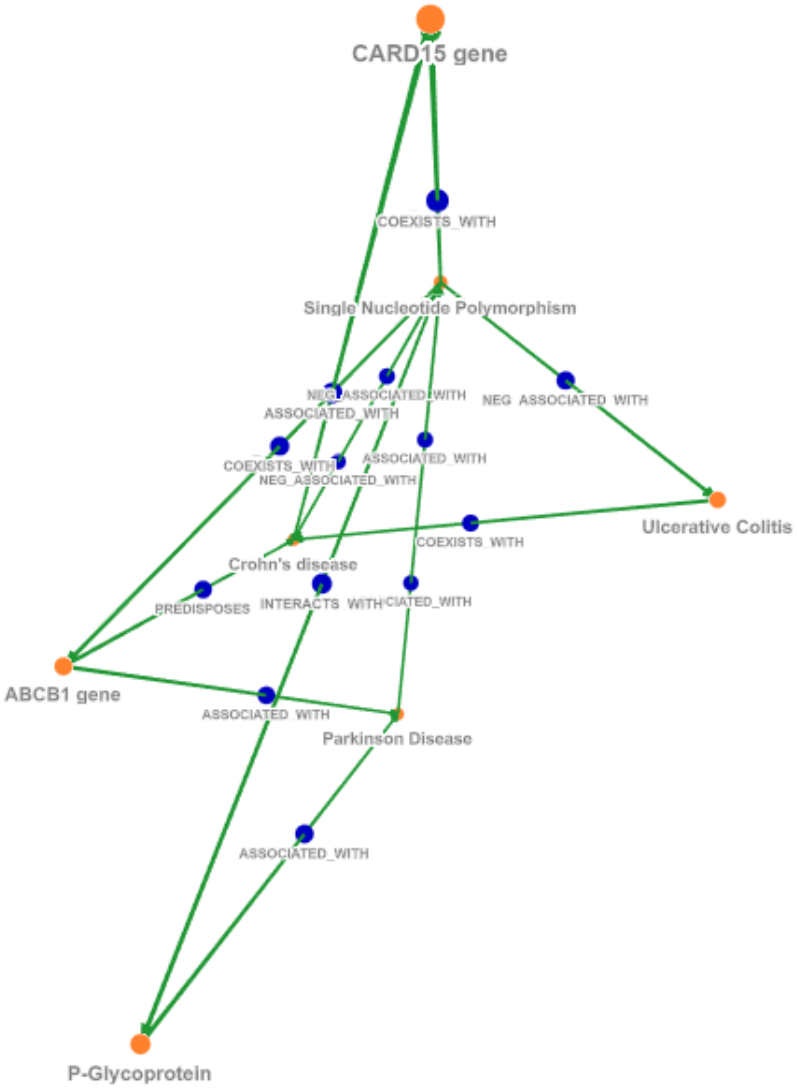


Figure E-24. PD – CD, PL3 SNP result 1 (with citations)

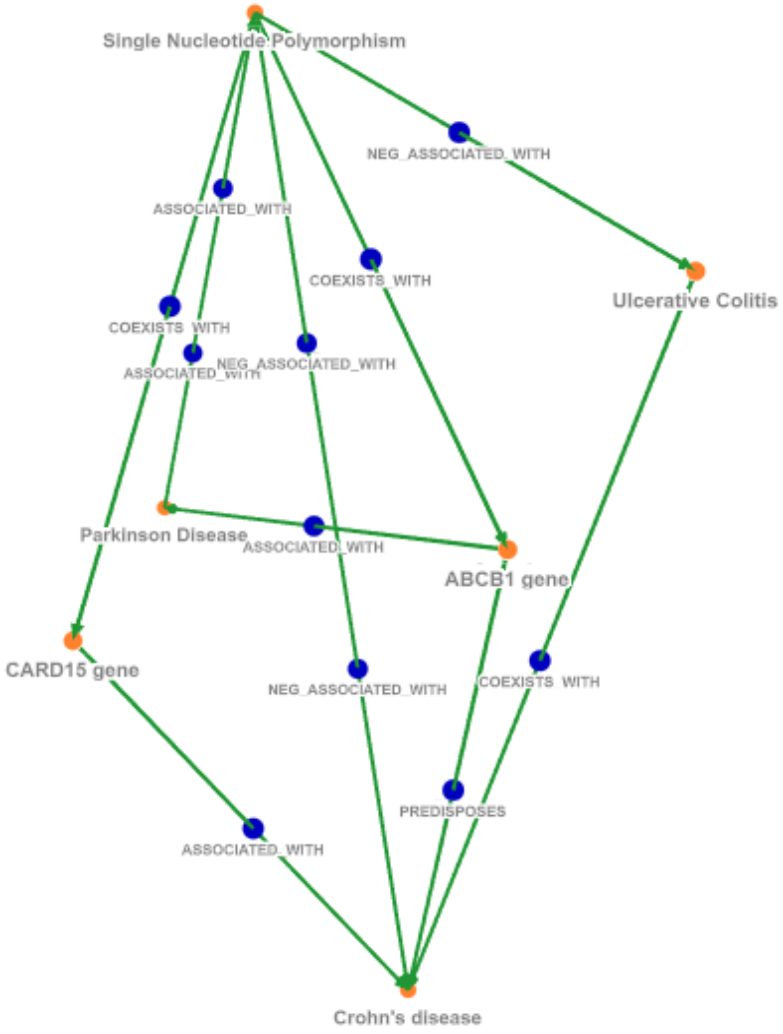


Figure E-25. PD – CD, PL3 SNP result 1 (without citations)

Table E-2. PD – CD generated predicates

[PMID10400088]
Title “Increase in level of tumor necrosis factor (TNF)-alpha in 6-hydroxydopamine-lesioned striatum in rats without influence of systemic L-DOPA on the TNF-alpha induction.”
[C0023570|Levodopa – STIMULATES – C1448177|TNF protein, human]
Abstract “We previously reported that the levels of proinflammatory cytokines such as tumor necrosis factor (TNF)-alpha are increased in the striatum and cerebrospinal fluid from patients with Parkinson’s disease (PD) and in the striatum from 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP)-treated mice, a murine model of PD.”
[C1456820|Tumor Necrosis Factor-alpha – ISA – C0079189|cytokine]
Abstract “Presently we examined the changes in cytokine levels in the nigrostriatal dopaminergic regions in rats treated with an intrastriatal injection of 6-hydroxydopamine (6-OHDA) as a model of slowly progressive neurodegeneration similar to that seen in PD.”
[C0079189|cytokine – ASSOCIATED_WITH – C0030567|Parkinson Disease]
[PMID11052175]
Title “Enteric bacteria, lipopolysaccharides and related cytokines in inflammatory bowel disease: biological and clinical significance.”
[C0023810|Lipopolysaccharides – ASSOCIATED_WITH – C0021390|Inflammatory Bowel Diseases] [C0079189|cytokine – ASSOCIATED_WITH – C0021390|Inflammatory Bowel Diseases]
Abstract “Ulcerative colitis (UC) and Crohn’s disease (CD) [inflammatory bowel disease (IBD)] are both characterized by an exaggerated immune response at the gut associated lymphoreticular tissue level.”
[C0301872|Immune response – ASSOCIATED_WITH – C0009324|Ulcerative Colitis]

[C0301872|Immune response – ASSOCIATED_WITH – C0010346|Crohn's disease]
Abstract "Such an abnormal and dysregulated immune response may be directed against luminal and/or enteric bacterial antigens, as also supported by murine models of inflammatory bowel disease (IBD) caused by organisms such as Citrobacter rodentium and Helicobacter hepaticus."
[C0029235|Organism – CAUSES – C0021390|Inflammatory Bowel Diseases] [C0887836|Citrobacter rodentium – CAUSES(SPEC) – C0021390|Inflammatory Bowel Diseases] [C0887836|Citrobacter rodentium – ISA – C0029235|Organism] [C1003868|Helicobacter hepaticus – CAUSES(SPEC) – C0021390|Inflammatory Bowel Diseases] [C1003868|Helicobacter hepaticus – ISA – C0029235|Organism]
Abstract "On these grounds, in experimental models and in human IBD, anti-cytokine monoclonal antibodies and interleukin receptor antagonists are under investigation for their capacity to neutralize the noxious effects of immune mediators."
[C0206553|Interleukin Receptor – DISRUPTS – C0021390|Inflammatory Bowel Diseases]
[PMID11743591]
Title "Infliximab for the treatment of Crohn's disease: efficacy, safety and pharmacoeconomics."
00000000.tx.1|relation|C0666743|infliximab|aapp,gngm,imft,phsu|aapp|||TREATS|C0010346|Crohn's disease|dsyn|dsyn||
Abstract "Conventional medical treatment of Crohn's disease includes the use of nonspecific anti-inflammatory drugs, immunosuppressives and antibiotics."
00000000.tx.4|relation|C0003209|Anti-Inflammatory Agents|phsu|phsu|||TREATS|C0010346|Crohn's disease|dsyn|dsyn|| 00000000.tx.4|relation|C0003232|Antibiotics|antb|antb|||TREATS|C0010346|Crohn's disease|dsyn|dsyn||
Abstract "Infliximab, a chimeric monoclonal antibody directed toward tumour necrosis factor alpha, is highly effective for the treatment of active Crohn's disease."

00000000.tx.7|relation|C1456820|Tumor Necrosis Factor-alpha|aapp,gngm,imft|aapp|||TREATS|C0010346|Crohn's disease|dsyn|dsyn||
Abstract "Moreover, infliximab is the only medical therapy that has been shown to be effective for the treatment of fistulizing Crohn's disease."
[C0666743|infliximab|aapp,gngm,imft,phsu|aapp|||TREATS(SPEC)|C0010346|Crohn's disease|dsyn|dsyn||
Abstract "Infliximab is recommended for the treatment of active Crohn's disease refractory to conventional drugs, and is the treatment of choice for fistulizing Crohn's disease."
00000000.tx.14|relation|C0087111|Therapeutic procedure|topp|topp|||TREATS|C0010346|Crohn's disease|dsyn|dsyn||
[PMID11922776]
Abstract "Injection with rolipram, an inhibitor of TNF-alpha expression, or use of IL-6 knockout mice was ineffective at impairing thymic apoptosis induction by the toxin cotreatment, suggesting that these cytokines did not mediate LPS potentiation."
[C0079189|cytokine – CAUSES – C0162638|Apoptosis]
[PMID12205053]
Title "Mice deficient in TNF receptors are protected against dopaminergic neurotoxicity: implications for Parkinson's disease."
[C0077503|Tumor Necrosis Factor Receptor – AFFECTS – C0030567|Parkinson Disease]
Abstract "Recent findings suggest that inflammatory processes are associated with several neurodegenerative disorders, including PD."
[C0524851|Neurodegenerative Disorders – AFFECTS – C1512753|Inflammation Process]
Abstract "Enhanced expression of the proinflammatory cytokine, tumor necrosis factor (TNF)-alpha, has been found in association with glial cells in the substantia nigra of patients with PD."

[C1448177|TNF protein, human – ISA – C0079189|cytokine]
Abstract "To determine the potential role for TNF-alpha in PD, we examined the effects of the 1-methyl-4-phenyl-1,2,3,4-tetrahydropyridine (MPTP), a dopaminergic neurotoxin that mimics some of the key features associated with PD, using transgenic mice lacking TNF receptors."
[C1448177|TNF protein, human – ASSOCIATED_WITH – C0030567|Parkinson Disease]
Abstract "The data indicate that the proinflammatory cytokine TNF-alpha is an obligatory component of dopaminergic neurodegeneration."
[C1456820|Tumor Necrosis Factor-alpha – ISA – C0079189|cytokine]
[PMID12474223]
Title "Chemokine expression in IBD."
[C0282554|chemokine – ASSOCIATED_WITH – C0021390|Inflammatory Bowel Diseases]
Abstract "Mucosal chemokine expression is unselectively increased in both ulcerative colitis and Crohn's disease."
[C0282554|chemokine – AUGMENTS – C0009324|Ulcerative Colitis] [C0282554|chemokine – AUGMENTS – C0010346|Crohn's disease]
Abstract "Individual chemokine expression was found to be significantly up-regulated in IBD when patients were compared with the non-diseased group in all areas of the mucosal sections."
[C0282554|chemokine – AUGMENTS – C0021390|Inflammatory Bowel Diseases]
Abstract "It is concluded that human colonic chemokine expression is non-selectively up-regulated in IBD."
[C0282554|chemokine – AUGMENTS – C0021390|Inflammatory Bowel Diseases]

Abstract “The results supported the hypothesis that the degree of local inflammation and tissue damage in UC and CD is dependent on local expression of specific chemokines within IBD tissues.”
[C0010957|Tissue damage – COEXISTS_WITH – C0009324|Ulcerative Colitis] [C0010957|Tissue damage – COEXISTS_WITH – C0010346|Crohn’s disease] [C0021368|Inflammation – COEXISTS_WITH – C0009324|Ulcerative Colitis] [C0021368|Inflammation – COEXISTS_WITH – C0010346|Crohn’s disease]
[PMID12724617]
Title “Polymorphism in the P-glycoprotein drug transporter MDR1 gene: a possible link between environmental and genetic factors in Parkinson’s disease.”
[C0376622|ABCB1 gene - ASSOCIATED_WITH – C0030567|Parkinson Disease]
Abstract “Thus, it appears that mutation of the MDR1 gene predisposes to damaging effects of pesticides, and possibly other toxic xenobiotics transported by P-glycoprotein, leading to Parkinson’s disease.”
[C0242643|P-Glycoprotein – CAUSES – C0030567|Parkinson Disease]
[PMID14724828]
Title “A murine model of chronic inflammation-induced intestinal fibrosis down-regulated by antisense NF-kappa B.”
[C0079904|NF-kappa B – DISRUPTS – C0016059|Fibrosis]
Abstract “To elucidate extracellular matrix (ECM) changes underlying intestinal fibrosis, a frequent complication of inflammatory bowel disease, we developed a murine model of chronic colitis associated with intestinal fibrosis.”
[C0016059|Fibrosis – COEXISTS_WITH – C0267375|Chronic colitis]
Abstract “Colonic expression of collagens (Col1a2, Col3a2), ECM remodeling genes (matrix metalloproteinase [MMP]-1, -3, and tissue inhibitor of matrix metalloproteinase

[TIMP]-1), and inflammation-modulating cytokines (tumor necrosis factor alpha [TNF-alpha], interferon gamma [IFN-gamma], transforming growth factor beta 1 [TGF-beta 1], and insulin-like growth factor 1 [IGF-1]) were assessed by semiquantitative reverse-transcription polymerase chain reaction."
[C1448177|TNF protein, human – ISA – C0079189|cytokine]
[PMID14749689]
Abstract "P-glycoprotein, the encoded product of the human MDR1 (ABCB1) gene, is of particular clinical relevance in that this transporter has broad substrate specificity, including a variety of structurally divergent drugs in clinical use today."
[C0376622|ABCB1 gene – PRODUCES - C0242643|P-Glycoprotein]
Abstract "Recently, a number of single-nucleotide polymorphisms (SNPs) in MDR1 have been identified."
[C0752046|Single Nucleotide Polymorphism - COEXISTS_WITH - C0376622|ABCB1 gene]
Abstract "An increasing number of studies have also implicated certain commonly occurring SNPs in MDR1 in problems including altered drug levels and host susceptibility to diseases such as Parkinson's disease, inflammatory bowel disease, refractory seizures, and CD4 cell recovery during human immunodeficiency virus therapy."
[C0376622|ABCB1 gene – TREATS - C0001175|Acquired Immunodeficiency Syndrome] [C0752046|Single Nucleotide Polymorphism - COEXISTS_WITH – C0376622|ABCB1 gene]
[PMID15058528]
Abstract "Intestinal fibrostenosis is a frequent and debilitating complication of Crohn's disease (CD), not only resulting in small bowel obstruction, but eventually in repeated bowel resection and short bowel syndrome."
[C0010346|Crohn's disease – CAUSES – C0036992|Short Bowel Syndrome] [C0010346|Crohn's disease – CAUSES – C0235329|Small bowel obstruction NOS]

Abstract “Intestinal fibrosis is a consequence of chronic transmural inflammation in CD.”
[C0021376|Chronic inflammation – COEXISTS_WITH – C0010346|Crohn’s disease]
Abstract “Tumor necrosis factor, on the other hand, has antifibrotic bioactivity and pharmacologic inhibition of this cytokine carries a theoretical risk of enhanced stricture formation.”
[C1456820|Tumor Necrosis Factor-alpha – INHIBITS – C0079189|cytokine]
[PMID15109580]
Title “Inflammation and neurodegeneration in Parkinson’s disease.”
[C0021368|Inflammation – COEXISTS_WITH – C0030567|Parkinson Disease] [C0027746|Nerve Degeneration – COEXISTS_WITH – C0030567|Parkinson Disease]
Abstract “Reports in the literature indicate that antiinflammatory agents inhibit dopaminergic cell death in animal models of PD, and there is one epidemiological report that their use significantly diminishes the risk of PD in humans.”
[C0003209|Anti-Inflammatory Agents – DISRUPTS – C0007587|Cell Death]
Abstract “These data support the hypothesis that chronic inflammation may play an important role, if secondary, in the pathogenesis of PD.”
[C0021376|Chronic inflammation – CAUSES – C0030567|Parkinson Disease]
[PMID15256718]
Abstract “In particular, genetic variations of the human ABCB1 (P-glycoprotein/MDR1) gene have been most extensively studied. Hitherto more than fifty single nucleotide polymorphisms (SNPs) and insertion/deletion polymorphisms in the ABCB1 gene have been reported.”
[C0752046|Single Nucleotide Polymorphism – COEXISTS_WITH – C0376622|ABCB1]

[PMID15503194]
Abstract “Accumulating evidence has suggested that inflammation in the brain participates in the pathogenesis of Parkinson’s disease (PD).”
[C0021368|Inflammation – CAUSES – C0030567|Parkinson Disease]
[PMID15580149]
Title “Efficacy and safety of tumor necrosis factor antagonists in Crohn’s disease: overview of randomized clinical studies.”
[C0243076|antagonists – TREATS – C0010346|Crohn’s disease]
Abstract “The past decade has brought forth a series of novel biologic agents targeting tumor necrosis factor (TNF) for the treatment of Crohn’s disease.”
[C1456820|Tumor Necrosis Factor-alpha – TREATS – C0010346|Crohn’s disease]
Abstract “However, the anti-TNF strategies might not have identical efficacy and safety profiles and might differ in dosing compared with therapy for rheumatoid arthritis.”
[C1456820|Tumor Necrosis Factor-alpha – TREATS – C0003873|Rheumatoid Arthritis]
Abstract “Most recently, adalimumab has been approved by the US Food and Drug Administration for the treatment of rheumatoid arthritis and is undergoing evaluation in Crohn’s disease, with promising initial results.”
[C1122087|adalimumab – ASSOCIATED_WITH – C0010346|Crohn’s disease]
Abstract “This review discusses the results of controlled clinical trials of anti-TNF agents for Crohn’s disease.”
[C0040616|Anti-Anxiety Agents – TREATS – C0010346|Crohn’s disease]

[PMID15685540]
Abstract “The G2677T SNP was not associated with UC or CD.”
[C0752046|Single Nucleotide Polymorphism – NEG_ASSOCIATED_WITH – C0009324|Ulcerative Colitis] [C0752046|Single Nucleotide Polymorphism – NEG_ASSOCIATED_WITH – C0010346|Crohn’s disease]
[PMID15778431]
Abstract “Crohn’s disease (CD) is a chronic, relapsing inflammatory bowel disease, characterized by transmural inflammation.”
[C0010346|Crohn’s disease – ISA – C0021390|Inflammatory Bowel Diseases] [C0021368|Inflammation – COEXISTS_WITH – C0021390|Inflammatory Bowel Diseases] [C0021368|Inflammation – COEXISTS_WITH(SPEC) – C0010346|Crohn’s disease]
Abstract “Because transforming growth factor (TGF)-beta1 can mediate both fibrosis and mesenchymal cell proliferation; we studied the effects of delivering adenoviral vectors encoding spontaneously active TGF-beta1 into the colons of mice.”
[C0080222|TGFB1 – ASSOCIATED_WITH – C0016059|Fibrosis]
[PMID15869932]
Title “Oxidative stress and inflammation in Parkinson’s disease: is there a causal link?”
[C0021368|Inflammation – COEXISTS_WITH – C0030567|Parkinson Disease] [C0242606|Oxidative Stress – COEXISTS_WITH – C0030567|Parkinson Disease]
Abstract “Parkinson’s disease (PD) is a neurodegenerative disorder characterized by a dramatic loss of dopaminergic neurons in the substantia nigra (SN).”
[C0030567|Parkinson Disease – ISA – C0524851|Neurodegenerative Disorders]

Abstract “Perhaps not surprisingly, non-steroidal anti-inflammatory drugs (NSAIDs) have been shown to reduce the risk of developing PD.”
[C0003211|Anti-Inflammatory Agents, Non-Steroidal – PREVENTS – C0030567|Parkinson Disease]
[PMID15879017]
Abstract “The extract was then tested for its ability to alter in vitro production of TH1 cytokines (interleukin-2 [IL-2] and gamma interferon) and TH2 cytokines (IL-4 and IL-10) by murine splenocytes.”
[C0021740|Recombinant Interferon-gamma – ISA|C0079189|cytokine] [C0021756|Interleukin-2 – ISA – C0079189|cytokine] [C0021758|Interleukin-4 – ISA – C0079189|cytokine] [C0085295|Interleukin-10 – ISA – C0079189|cytokine]
[PMID16252231]
Abstract “We identified 11 SNPs that were associated with PD (P<.01) in both tier 1 and tier 2 samples and had the same direction of effect.”
[C0752046|Single Nucleotide Polymorphism – ASSOCIATED_WITH-C0030567|Parkinson Disease]
Abstract “The protein encoded by this gene plays an important role in neurogenesis and in neuronal apoptosis, which is consistent with existing hypotheses regarding PD pathogenesis.”
[C0030567|Parkinson Disease – AFFECTS – C0162638|Apoptosis] [C0030567|Parkinson Disease – AFFECTS – C0814002|Neural Development]
[PMID16374256]
Abstract “The multidrug resistance MDR1 gene codes for a membrane transporter associated with inflammatory bowel disease.”
[C0376622|ABCB1 gene – ISA – C0376623|Multidrug Resistance Gene] [C0596902|Membrane Transport Proteins – ASSOCIATED_WITH – C0021390|Inflammatory Bowel Diseases]

Abstract “We studied the association of both polymorphisms in an independent population to reveal the impact of the MDR1 gene on predisposition to inflammatory bowel disease.”
[C0376622|ABCB1 gene – PREDISPOSES – C0021390|Inflammatory Bowel Diseases]
Abstract “Therefore, considering our results and those from others, the MDR1 gene behaves as a common risk factor for both CD and UC.”
[C0376622|ABCB1 gene – PREDISPOSES – C0009324|Ulcerative Colitis] [C0376622|ABCB1 gene – PREDISPOSES – C0010346|Crohn’s disease]
Abstract “We discovered that the C3435 allele conferring susceptibility to CD is different from the described 3435T UC risk allele.”
[C0010346|Crohn’s disease – PREDISPOSES – C0009324|Ulcerative Colitis]
[PMID16461743]
Abstract “These data indicate that the NOD2fs mutation results in a loss-of-function phenotype in human myeloid DC and imply decreased immune regulation by IL-10 as a possible mechanism for this mutation in CD.”
[C0085295|Interleukin-10 – ASSOCIATED_WITH – C0010346|Crohn’s disease]
[PMID16493072]
Abstract “We found that AKBA potentiated the apoptosis induced by TNF and chemotherapeutic agents, suppressed TNF-induced invasion, and inhibited receptor activator of NF-kappaB ligand-induced osteoclastogenesis, all of which are known to require NF-kappaB activation.”
[C0003392|Antineoplastic Agents – CAUSES – C0162638|Apoptosis] [C0812246|TNF gene – CAUSES – C0162638|Apoptosis]
Abstract “Overall, our results indicated that AKBA enhances apoptosis induced by cytokines and chemotherapeutic agents, inhibits invasion, and suppresses osteoclastogenesis through inhibition of NF-kappaB-regulated gene expression.”

[C0003392|Antineoplastic Agents – CAUSES – C0162638|Apoptosis] [C0079189|cytokine – CAUSES – C0162638|Apoptosis]
[PMID16717254]
Abstract "This raises the possibility that anti-apoptotic agents might be neuroprotective in PD."
[C0040616|Anti-Anxiety Agents – TREATS – C0030567|Parkinson Disease]
Abstract "Propargylamines have been demonstrated to be potent anti-apoptotic agents in both in vitro and in vivo studies, presumably by maintaining glyceraldehyde-3-phosphate dehydrogenase (GAPDH) as a dimer and thereby preventing its nuclear translocation where it blocks upregulation of anti-apoptotic proteins."
[C0040616|Anti-Anxiety Agents – STIMULATES – C1565114|Apoptosis Inhibiting Proteins]
[PMID17052658]
Title "Lack of replication of thirteen single-nucleotide polymorphisms implicated in Parkinson's disease: a large-scale international study."
[C0752046|Single Nucleotide Polymorphism – ASSOCIATED_WITH – C0030567|Parkinson Disease]
Abstract "A genome-wide association study identified 13 single-nucleotide polymorphisms (SNPs) significantly associated with Parkinson's disease."
[C0752046|Single Nucleotide Polymorphism – ASSOCIATED_WITH – C0030567|Parkinson Disease]
[PMID17156147]
Title "Fibrogenesis in Crohn's disease."
[C0596570|fibrogenesis – ASSOCIATED_WITH – C0010346|Crohn's disease]
Abstract "While the pathogenesis of inflammation in CD has been extensively investigated,

knowledge of stricture pathogenesis remains limited.”
[C0021368|Inflammation – COEXISTS_WITH – C0010346|Crohn’s disease]
[PMID17185560]
Abstract “We report that a synonymous SNP in the Multidrug Resistance 1 (MDR1) gene, part of a haplotype previously linked to altered function of the MDR1 gene product P-glycoprotein (P-gp), nonetheless results in P-gp with altered drug and inhibitor interactions.”
[C0752046|Single Nucleotide Polymorphism - COEXISTS_WITH - C0376622|ABCB1 gene]
[PMID17339843]
Title “Inflammation as a causative factor in the aetiology of Parkinson’s disease.”
[C0021368|Inflammation – CAUSES – C0030567|Parkinson Disease]
Abstract “Parkinson’s disease (PD) is a progressive neurodegenerative disorder affecting mainly the elderly, although a small proportion of PD patients develop the illness at a much younger age.”
[C0030567|Parkinson Disease – ISA – C0524851|Neurodegenerative Disorders]
Abstract “These factors are examined in this review along with a consideration of the possible use of anti-inflammatory drugs in PD.”
[C0003209|Anti-Inflammatory Agents – TREATS – C0030567|Parkinson Disease]
[PMID17470824]
Abstract “Adalimumab, a fully human tumor necrosis factor (TNF) antagonist, is an effective treatment for patients with Crohn disease who are naïve to the chimeric TNF antagonist, infliximab.”

[C0666743|infliximab – TREATS(INFER) – C0010346|Crohn's disease] [C1122087|adalimumab - INHIBITS – C1448177|TNF protein, human]
Abstract "No anti-TNF agent has been evaluated prospectively in patients with Crohn disease who had responded to another anti-TNF agent and then lost that response or were intolerant of the agent."
[C0040616|Anti-Anxiety Agents – TREATS(INFER) – C0010346|Crohn's disease]
Abstract "To determine whether adalimumab induces remissions more frequently than placebo in adult patients with Crohn disease who have symptoms despite infliximab therapy or who cannot take infliximab because of adverse events."
[C1122087|adalimumab – TREATS(INFER) – C0010346|Crohn's disease]
Abstract "Adalimumab induces remissions more frequently than placebo in adult patients with Crohn disease who cannot tolerate infliximab or have symptoms despite receiving infliximab therapy."
[C1122087|adalimumab – TREATS(INFER) – C0010346|Crohn's disease]
[PMID17908040]
Abstract "TNF-alpha is secreted by the brain resident marcophage (the microglial cell) in response to various stimuli. It has been demonstrated to play a major role in central nervous system (CNS) neuroinflammation-mediated cell death in AD, PD and amyotrophic lateral sclerosis (ALS) as well as several other CNS complications."
[C0002395|Alzheimer's Disease – AFFECTS - C0007587|Cell Death] [C0002736|Amyotrophic Lateral Sclerosis – AFFECTS - C0007587|Cell Death] [C0030567|Parkinson Disease - AFFECTS - C0007587|Cell Death]
[PMID17984451]
Title "Nonsteroidal anti-inflammatory drugs may protect against Parkinson disease."
[C0003211|Anti-Inflammatory Agents, Non-Steroidal – PREVENTS – C0030567|Parkinson Disease]

Abstract “Markers of neuroinflammation, including activated microglia and increased levels of circulating proinflammatory cytokines, have been observed in the brains and CSF of patients with Parkinson disease (PD).”
[C0079189|cytokine – ASSOCIATED_WITH(INFER) – C0030567|Parkinson Disease]
Abstract “Our study contributes to the growing body of literature suggesting a protective role for nonsteroidal anti-inflammatory drugs (NSAIDs) in Parkinson disease (PD).”
[C0003211|Anti-Inflammatory Agents, Non-Steroidal – TREATS – C0030567|Parkinson Disease]
Abstract “Yet the link between anti-inflammatory agents and PD in humans remains uncertain, despite indications that neuroinflammation may contribute to cell death in the PD brain and experimental evidence of anti-inflammatory agents such as nonsteroidal anti-inflammatory drugs (NSAIDs) exerting neuroprotective effects in animal models.”
[C0003209|Anti-Inflammatory Agents – AFFECTS – C0007587|Cell Death]
[PMID19221310]
Abstract “Imaging of nondopaminergic targets such as inflammation or alpha-synuclein deposition may provide further insight into the etiology of PD.”
[C0021368|Inflammation – CAUSES – C0030567|Parkinson Disease]
[PMID8015728]
Abstract “Since TNF-alpha is an important signal transducer of the immune system with cytotoxic and stimulator properties, these results suggest that an immune response may occur in the nigrostriatal dopaminergic regions in Parkinson’s disease and that TNF-alpha may be related, at least in part, to the neuronal degeneration.”
[C0301872|Immune response – ASSOCIATED_WITH – C0030567|Parkinson Disease] [C1456820|Tumor Necrosis Factor-alpha – DISRUPTS – C0027746|Nerve Degeneration]

[PMID8196673]
Abstract “An understanding of the molecular basis of the complex I defect in PD and its relationship to other biochemical changes will provide important insight into the potential chain of events that lead to dopaminergic cell death in PD.”
[C0030567|Parkinson Disease – AFFECTS - C0007587|Cell Death]
[PMID9422513]
Abstract “Various molecules such as cytokines and anticancer drugs, as well as factor deprivation, rapidly induce apoptosis (programmed cell death), which is morphologically characterized by cell shrinkage and the blebbing of plasma membranes and by nuclear condensation.”
[C0079189|cytokine – CAUSES - C0162638|Apoptosis]
[PMID9550432]
Title “A role for TNF-alpha and mucosal T helper-1 cytokines in the pathogenesis of Crohn's disease.”
[C0079189|cytokine – CAUSES - C0010346|Crohn's disease] [C1456820|Tumor Necrosis Factor-alpha – CAUSES - C0010346|Crohn's disease]
Abstract “To assess the role of TNF-alpha in mucosal cytokine regulation, the effects of TNF-alpha on lamina propria mononuclear cell (LPMC) Th1 production were determined.”
[C1456820|Tumor Necrosis Factor-alpha - INTERACTS_WITH - C0079189|cytokine]
[PMID9820396]
Abstract “This study aimed to shed more light on the interrelation between DHEAS and cortisol (and humoral markers of inflammation) in chronic inflammatory bowel disease.”
[C0020268|Hydrocortisone - ASSOCIATED_WITH - C0021390|Inflammatory Bowel Diseases]

[C0057277|Dehydroepiandrosterone Sulfate - ASSOCIATED_WITH - C0021390|Inflammatory Bowel Diseases]
Abstract “DHEAS was lower in patients with CD (p < 0.005) and UC (p < 0.005) than in controls, which was, in part, dependent on previous corticosteroid treatment (p < 0.01).”
[C0057277|Dehydroepiandrosterone Sulfate - TREATS(INFER) - C0009324|Ulcerative Colitis] [C0057277|Dehydroepiandrosterone Sulfate - TREATS(INFER) - C0010346|Crohn's disease]
Abstract “DHEAS as a marker of inflammation was low in CD and UC.”
[C0009324|Ulcerative Colitis – ISA - C0021368|Inflammation] [C0010346|Crohn's disease – ISA - C0021368|Inflammation] [C0057277|Dehydroepiandrosterone Sulfate – PREDISPOSES - C0021368|Inflammation] [C0057277|Dehydroepiandrosterone Sulfate - PREDISPOSES(SPEC) - C0009324|Ulcerative Colitis] [C0057277|Dehydroepiandrosterone Sulfate - PREDISPOSES(SPEC) - C0010346|Crohn's disease]

Table E-3. Precision and recall using citations for Parkinson – Crohn study

Result	New Intermediate Terms	Relevancy (Topic)	Precision	Recall
d0	Inflammation	Neuroimmunology	1	0.333
d1	Anti-Apoptotic Agents	Cell Death	1	0.667
d2	Single Nucleotide Polymorphism	Genetics	1	1
d3	Chronic Inflammation	Neuroimmunology	1	1
d4	Immune Response	Neuroimmunology	1	1
d5	Cytokine	Neuroimmunology	1	1
d6	Anti-Inflammatory Agents	Neuroimmunology	1	1
d7	ABCB1 Gene	Genetics	1	1

Table E-4. Precision and recall not using citations for Parkinson – Crohn study

Result	New Intermediate Terms	Relevancy (Topic)	Precision	Recall
d0	Inflammation	Neuroimmunology	1	0.333
d1	Anti-Apoptotic Agents	Cell Death	1	0.667
d2	Single Nucleotide Polymorphism	Genetics	1	1
d3	Chronic Inflammation	Neuroimmunology	1	1
d4	Immune Response	Neuroimmunology	1	1
d5	Cytokine	Neuroimmunology	1	1
d6	Anti-Inflammatory Agents	Neuroimmunology	1	1
d7	ABCB1 Gene	Genetics	1	1

Appendix F

MeSH Details

Table F-1. MeSH Categories

A – Anatomy
B – Organisms
C – Diseases
D – Chemicals and Drugs
E – Analytical, Diagnostic and Therapeutic Techniques, and Equipment
F – Psychiatry and Psychology
G – Phenomena and Processes
H – Disciplines and Occupations
I – Anthropology, Education, Sociology, and Social Phenomena
J – Technology, Industry, and Agriculture
K – Humanities
L – Information Science
M – Named Groupos
N – Health Care
V – Publication Characteristics
Z – Geographicals

References

Aamodt, A. H., Stovner, L. J., Midthjell, K., Hagen, K., & Zwart, J. A. (2007). Headache prevalence related to diabetes mellitus. The Head-HUNT study. *European Journal of Neurology*, 14(7), 738-744.

Abu-Jbara, A., & Radev, D. (2011, June). Coherent citation-based summarization of scientific papers. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1* (pp. 500-509). Association for Computational Linguistics.

Adams, S. O., Nissley, S. P., Handwerger, S., & Rechler, M. M. (1983). Developmental patterns of insulin-like growth factor-I and-II synthesis and regulation in rat fibroblasts. *Nature*, 302(5904), 150.

Ahlers, C. B., Hristovski, D., Kilicoglu, H., & Rindflesch, T. C. (2007). Using the literature-based discovery paradigm to investigate drug mechanisms. In *Proceedings of the American Medical Informatics Association Annual Symposium* (Vol. 2007, p. 6). American Medical Informatics Association.

Allahyari, M., Pouriyeh, S., Assefi, M., Safaei, S., Trippe, E. D., Gutierrez, J. B., & Kochut, K. (2017). Text summarization techniques: a brief survey. arXiv preprint arXiv:1707.02268.

Allen, G. S., Gross, C. J., Henderson, L. M., & Chou, S. N. (1976). Cerebral arterial spasm: Part 4: In vitro effects of temperature, serotonin analogues, large nonphysiological concentrations of serotonin, and extracellular calcium and magnesium on serotonin-induced contractions of the canine basilar artery. *Journal of Neurosurgery*, *44*(5), 585-593.

Altura, B. M. (1970). Influence of magnesium and cysteine on vasopressin-induced contractions in various canine blood vessels. *Experientia*, *26*(10), 1089-1090.

Altura, B. M. (1981). Magnesium ions and contraction of vascular smooth muscles: relationship to some vascular diseases. In *Federation Proceedings* (Vol. 40, pp. 2672-2679).

Altura, B. M., & Turlapaty, P. D. (1982). Withdrawal of magnesium enhances coronary arterial spasms produced by vasoactive agents. *British Journal of Pharmacology*, *77*(4), 649-659.

Altura, B. M., & Altura, B. T. (1983). Pharmacologic inhibition of cerebral vasospasm in ischemia, hallucinogen ingestion, and hypomagnesemia: barbiturates, calcium antagonists, and magnesium. *The American Journal of Emergency Medicine*, *1*(2), 180-190.

Altura, B. M., & Altura, B. T. (1984). Magnesium, electrolyte transport and coronary vascular tone. *Drugs*, *28*(1), 120-142.

Andersen, M. L., & Tufik, S. (2008). The effects of testosterone on sleep and sleep-disordered breathing in men: its bidirectional interaction with erectile function. *Sleep Medicine Reviews*, *12*(5), 365-379.

Anderson, W. W., Lewis, D. V., Swartzwelder, H. S., & Wilson, W. A. (1986). Magnesium-free medium activates seizure-like events in the rat hippocampal slice. *Brain Research*, *398*(1), 215-219.

Andersson, K. E. (1986). Pharmacodynamic profiles of different calcium channel blockers. *Basic & Clinical Pharmacology & Toxicology*, *58*(S2), 31-42.

Anthony, M. (1982). Serotonin and cyclic nucleotides in migraine. *Advances in neurology*, *33*, 45.

Athenikos, S. J., & Han, H. (2010). Biomedical question answering: A survey. *Computer Methods and Programs in Biomedicine*, 99(1), 1-24.

Axelsson, J., Ingre, M., Åkerstedt, T., & Holmbäck, U. (2005). Effects of acutely displaced sleep on testosterone. *The Journal of Clinical Endocrinology & Metabolism*, *90*(8), 4530-4535.

Baeza-Yates, R., & Ribeiro-Neto, B. (1999). *Modern Information Retrieval* (Vol. 463). New York: ACM press.

Bakal, G., Talari, P., Kakani, E. V., & Kavuluru, R. (2018). Exploiting semantic patterns over biomedical knowledge graphs for predicting treatment and causative relations. *Journal of Biomedical Informatics*, 82, 189-199.

Barbul, A., Fishel, R. S., Shimazu, S., Wasserkrug, H. L., Yoshimura, N. N., Tao, R. C., & Efron, G. (1985). Intravenous hyperalimentation with high arginine levels improves wound healing and immune function. *Journal of Surgical Research*, 38(4), 328-334.

Basser, L. S. (1969). The relation of migraine and epilepsy. *Brain*, *92*(2), 285-300.

Belch, J. J. F., Newman, P., Drury, J. K., Capell, H., Leiberman, P., James, W. B., Forbes, C. D., & Prentice, C. R. M. (1981). Successful treatment of Raynaud's syndrome with prostacyclin. *Thrombosis and Haemostasis*, *46*(03), 255-256.

Berge, L. I., Riise, T., Fasmer, O. B., Hundal, Ø., Oedegaard, K. J., Midthjell, K., & Lund, A. (2013). Does diabetes have a protective effect on migraine?. *Epidemiology*, 129-134.

Binoux, M., Hossenlopp, P., Lassarre, C., & Seurin, D. (1980). Somatomedin production by rat liver in organ culture. I. Validity of the technique. Influence of the released material on cartilage sulphation. Effects of growth hormone and insulin. *Acta Endocrinologica*, 93(1), 73.

Borman, A., Wood, T. R., Black, H. C., Anderson, E. G., Oestekling, M., Womack, M., & Rose, W. C. (1946). The role of arginine in growth, with some observations on the effects of argininic acid. *Journal of Biological Chemistry*, 166, 585-594.

Brandt, L., Andersson, K. E., Edvinsson, L., & Ljunggren, B. (1981). Effects of extracellular calcium and of calcium antagonists on the contractile responses of isolated human pial and mesenteric arteries. *Journal of Cerebral Blood Flow & Metabolism*, *1*(3), 339-347.

Briel, R. C., Lippert, T. H., & Zahradnik, H. P. (1985). Action of magnesium sulfate on platelet prostacyclin interaction and prostacyclin of blood vessels. *American Journal of Obstetrics & Gynecology*, *153*(2), 232.

Briel, R. C., Lippert, T. H., & Zahradnik, H. P. (1987). Changes in blood coagulation, thrombocyte function and vascular prostacyclin synthesis caused by magnesium sulfate. *Geburtshilfe und Frauenheilkunde*, *47*(5), 332-336.

Bücking, H., & Baumgartner, G. (1974). Clinical and pathophysiological considerations on initial neurological symptoms of focal migraine (author's transl). *Archiv fur Psychiatrie und Nervenkrankheiten*, *219*(1), 37.

Burke, J. P., Mulsow, J. J., O'keane, C., Docherty, N. G., Watson, R. W. G., & O'connell, P. R. (2007). Fibrogenesis in Crohn's disease. *The American Journal of Gastroenterology*, *102*(2), 439.

Cameron, D., Bodenreider, O., Yalamanchili, H., Danh, T., Vallabhaneni, S., Thirunarayan, K., Sheth, A. P., & Rindflesch, T. C. (2013). A graph-based recovery and decomposition of Swanson's hypothesis using semantic predications. *Journal of Biomedical Informatics*, *46*(2), 238-251.

Cameron, D. H. (2014). *A Context-driven Subgraph Model for Literature-based Discovery* (Doctoral dissertation, Wright State University).

Cameron, D., Kavuluru, R., Rindflesch, T. C., Sheth, A. P., Thirunarayan, K., & Bodenreider, O. (2015). Context-driven automatic subgraph creation for literature-based discovery. *Journal of Biomedical Informatics*, *54*, 141-157.

Canelas, H. M., De Assis, L. M., & De Jorge, F. B. (1965). Disorders of magnesium metabolism in epilepsy. *Journal of Neurology, Neurosurgery, and Psychiatry*, *28*(4), 378.

Carlson, L. A., Ekelund, L. G., & Orö, L. (1968). Clinical and metabolic effects of different doses of prostaglandin E1 in man. *Journal of Internal Medicine*, *183*(1-6), 423 - 430.

Chang, F. C., & Opp, M. R. (2000). IL-1 is a mediator of increases in slow-wave sleep induced by CRH receptor blockade. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, *279*(3), R793-R802.

Chen, C., Chen, Y., Horowitz, M., Hou, H., Liu, Z., & Pellegrino, D. (2009). Towards an explanatory and computational theory of scientific discovery. *Journal of Informetrics*, *3*(3), 191-209.

Chevrier, M. R., Ryan, A. E., Lee, D. Y. W., Zhongze, M., Wu-Yan, Z., & Via, C. S. (2005). Boswellia carterii extract inhibits TH1 cytokines and promotes TH2 cytokines in vitro. *Clinical and diagnostic laboratory immunology*, *12*(5), 575-580.

Daughaday, W. H., Phillips, L. S., & Herington, A. C. (1975). [6] Measurement of somatomedin by cartilage in Vitro. In *Methods in Enzymology* (Vol. 37, pp. 93-109). Academic Press.

DiGiacomo, R. A., Kremer, J. M., & Shah, D. M. (1989). Fish-oil dietary supplementation in patients with Raynaud's phenomenon: a double-blind, controlled, prospective study. *The American Journal of Medicine*, *86*(2), 158-164.

Divoli, A., Nakov, P., & Hearst, M. A. (2012). Do peers see more in a paper than its authors? *Advances in Bioinformatics*, *2012*.

Doerr, P., & Pirke, K. M. (1976). Cortisol-induced suppression of plasma testosterone in normal adult males. *The Journal of Clinical Endocrinology & Metabolism*, *43*(3), 622-629.

Domzal, T. (1975). Electroencephalographic studies in migraine. *Neurologia I Neurochirurgia Polska*, *9*(3), 337-341.

Edmundson, H. P. (1969). New methods in automatic extracting. *Journal of the ACM* (JACM), 16(2), 264-285.

Elbaz, A., Nelson, L. M., Payami, H., Ioannidis, J. P., Fiske, B. K., Annesi, G., Carmine Belin, A., Factor, S.A., Ferrarese, C., Hadjigeorgiou, G.M., Higgins, D.S., Kawakami, H., Krüger, R., Marder, K.S., Mayeux, R.P., Mellick, G.D., Nutt, J.G., Ritz, B., Samii, A., Tanner, C.M., Van Broeckhoven, C., Van Den Eeden, S.K., Wirdefeldt, K., Zabetian, C.P., Dehem, M., Montimurro, J.S., Southwick, A., Myers, R.M., & Trikalinos, T.A. (2006). Lack of replication of thirteen single-nucleotide polymorphisms implicated in Parkinson's disease: a large-scale international study. *The Lancet Neurology*, *5*(11), 917-923.

Elkiss, A., Shen, S., Fader, A., Erkan, G., States, D., & Radev, D. (2008). Blind men and elephants: What do citation summaries tell us about a research article? *Journal of the Association for Information Science and Technology*, *59*(1), 51-62.

Espiritu, J. R. D. (2008). Aging-related sleep changes. *Clinics in Geriatric Medicine*, *24*(1), 1-14.

Fagherazzi, G., El Fatouhi, D., Fournier, A., Gusto, G., Mancini, F., Balkau, B., Boutron-Ruault, M., Kurth, T., & Bonnet, F. (2019). Associations Between Migraine and Type 2 Diabetes in Women. *Jama Neurology*.

Fanciullacci, M., Pietrini, U., Geppetti, P., Nicolodi, M., Curradi, C., & Sicuteri, F. (1988). Substance P in the human iris: possible involvement in echothiophate-induced miosis in cluster headache. *Cephalalgia*, *8*(1), 49-53.

Fang, J., Wang, Y., & Krueger, J. M. (1998). Effects of interleukin-1β on sleep are mediated by the type I receptor. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, *274*(3), R655-R660.

Feldman, R., & Sanger, J. (2007). *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. Cambridge University Press.

Feller, I. & Stern, P. C. (2007). *A Strategy for Assessing Science: Behavioral and Social Research on Aging*. National Academies Press.

Fewtrell, C. M. S., Foreman, J. C., Jordan, C. C., Oehme, P., Renner, H., & Stewart, J. M. (1982). The effects of substance P on histamine and 5-hydroxytryptamine release in the rat. *The Journal of Physiology*, *330*(1), 393-411.

Firth, J. R. (1957). A synopsis of linguistic theory, 1930-1955. *Studies in linguistic analysis*.

Fiszman, M., Rindflesch, T. C., & Kilicoglu, H. (2004, May). Abstraction summarization for Managing the Biomedical Research Literature. In *Proceedings of the HLT-NAACL Workshop on Computational Lexical Semantics* (pp. 76-83). Association for Computational Linguistics.

Flink, E. B., Shane, S. R., Scobbo, R. R., Blehschmidt, N. G., & McDowell, P. (1979). Relationship of free fatty acids and magnesium in ethanol withdrawal in dogs. *Metabolism-Clinical and Experimental*, *28*(8), 858-865.

Francis, M. J. O., & Hill, D. J. (1975). Prolactin-stimulated production of somatomedin by rat liver. *Nature*, 255(5504), 167.

Friedman, A. P., Von Storch, T. J., & Merritt, H. H. (1954). Migraine and tension headaches: clinical study of two thousand cases. *Neurology*.

Frijters, R., Van Vugt, M., Smeets, R., Van Schaik, R., De Vlieg, J., & Alkema, W. (2010). Literature mining for the discovery of hidden connections between drugs, genes and diseases. *PLoS Comput Biol*, *6*(9), e1000943.

Fujii, S., Takemura, T., Wada, M., Akai, T., & Okuda, K. (1982). Magnesium levels in plasma, erythrocyte and urine in patients with diabetes mellitus. *Hormone and Metabolic Research*, *14*(03), 161-162.

Ganiz, M. C., Pottenger, W. M., & Janneck, C. D. (2005). Recent advances in literature based discovery. *Journal of the American Society for Information Science and Technology, JASIST (Submitted)*.

Gitau, R., Adams, D., Fisk, N. M., & Glover, V. (2005). Fetal plasma testosterone correlates positively with cortisol. *Archives of Disease in Childhood-Fetal and Neonatal Edition*, *90*(2), F166-F169.

Gong, Y., & Liu, X. (2001, September). Generic text summarization using relevance measure and latent semantic analysis. In *Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 19-25).

Gordon, M. D., & Lindsay, R. K. (1996). Toward discovery support systems: A replication, re-examination, and extension of Swanson's work on literature-based discovery of a connection between Raynaud's and fish oil. *Journal of the American Society for Information Science*, *47*(2), 116-128.

Gordon, M. D., & Dumais, S. (1998). Using latent semantic indexing for literature based discovery. *Journal of the American Society for Information Science*, 49(8), 674-685.

Gordon, M., Lindsay, R. K., & Fan, W. (2002). Literature-based discovery on the World Wide Web. *ACM Transactions on Internet Technology (TOIT)*, *2*(4), 261-275.

Hachinski, V. C., Olesen, J., Norris, J. A., Larsen, B., Enevoldsen, E., & Lassen, N. A. (1977). Cerebral hemodynamics in migraine. *Canadian Journal of Neurological Sciences*, *4*(4), 245-249.

Haghighi, F. S., Rahmanian, M., Namiranian, N., Arzaghi, S. M., Dehghan, F., Chavoshzade, F., & Sepehri, F. (2015). Migraine and type 2 diabetes; is there any association?. *Journal of Diabetes & Metabolic Disorders*, 15(1), 37.

Hall K 1972 Human somatomedin: determination, occurrence, biological activity and purification. *Acta Endocrinol* [Suppl] (Copenh) 163:1-52

Hammerman, M. R. (1987). Insulin-like growth factors and aging. *Endocrinology and Metabolism Clinics of North America*, 16(4), 995-1011.

Hanauer, S. B. (2004). Efficacy and safety of tumor necrosis factor antagonists in Crohn's disease: overview of randomized clinical studies. *Reviews in Gastroenterological Disorders*, *4*, S18-24.

Hanington, E., Jones, R. J., Amess, J. A. L., & Wachowicz, B. (1981). Migraine: a platelet disorder. *The Lancet*, *318*(8249), 720-723.

Hanington, E. (1986). The platelet and migraine. *Headache: The Journal of Head and Face Pain*, *26*(8), 411-415.

Hansen, A. J., & Zeuthen, T. (1981). Extracellular ion concentrations during spreading depression and ischemia in the rat brain cortex. *Acta Physiologica*, *113*(4), 437-445.

Hansson, H. A., Jennische, E., & Skottner, A. (1987). Regenerating endothelial cells express insulin-like growth factor-I immunoreactivity after arterial injury. *Cell and Tissue Research*, 250(3), 499-505.

Healy, M. D., Ward, J. A., & Wang, P. I. W. G. (1984). Inhibition of calcium by magnesium in the contraction of rat aortic smooth muscle. I. Mathematical model of calcium and magnesium binding. *Magnesium*, *3*(2), 63-72.

Hearst, M. A. (1999, June). Untangling text data mining. In *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics on Computational Linguistics* (pp. 3-10). Association for Computational Linguistics.

Hill, D. J., & Milner, R. D. G. (1985). Insulin as a growth factor. *Pediatric Research*, 19(9), 879.

Hintz, R. L., Suskind, R., Amatayakul, K., Thanangkul, O., & Olson, R. (1978). Plasma somatomedin and growth hormone values in children with protein-calorie malnutrition. *The Journal of Pediatrics*, 92(1), 153-156.

Hintz, R. L., Liu, F., Marshall, L. B., & Chang, D. (1980). Interaction of somatomedin-C with an antibody directed against the synthetic C-peptide region of insulin-like growth factor-I. *The Journal of Clinical Endocrinology & Metabolism*, 50(2), 405-407.

Ho, G. T., Nimmo, E. R., Tenesa, A., Fennell, J., Drummond, H., Mowat, C., Arnott, I.D., & Satsangi, J. (2005). Allelic variations of the multidrug resistance gene determine susceptibility and disease behavior in ulcerative colitis. *Gastroenterology*, *128*(2), 288-296.

Horrobin, D. F. (1977). Prostaglandins and migraine. *Headache: The Journal of Head and Face Pain*, *17*(3), 113-117.

Hristovski, D., Friedman, C., Rindflesch, T. C., & Peterlin, B. (2006). Exploiting semantic relations for literature-based discovery. In *Proceedings of American Medical Medical Association Annual Symposium* (Vol. 2006, p. 349). American Medical Informatics Association.

Hsu, L. K. G., Kalucy, R. S., Crisp, A. H., Koval, J., Chen, C. N., Carruthers, M., & Zilkha, K. J. (1977). Early morning migraine: nocturnal plasma levels of catecholamines, tryptophan, glucose, and free fatty acids and sleep encephalographs. *The Lancet*, *309*(8009), 447-451.

Hu, X., Li, G., Yoo, I., Zhang, X., & Xu, X. (2005, July). A semantic-based approach for mining undiscovered public knowledge from biomedical literature. In *Proceedings of IEEE International Conference on Granular Computing*, 2005 IEEE International Conference on (Vol. 1, pp. 22-27).

Hu, X., Yoo, I., Song, M., Zhang, Y., & Song, I. Y. (2005, October). Mining undiscovered public knowledge from complementary and non-interactive biomedical literature through semantic pruning. In *Proceedings of the 14th ACM International Conference on Information and Knowledge Management* (pp. 249-250). ACM.

Ishikawa, T., Onishi, Y., Hirano, H., Oosumi, K., Nagakura, M., & Tarui, S. (2004). Pharmacogenomics of drug transporters: a new approach to functional analysis of the genetic polymorphisms of ABCB1 (P-glycoprotein/MDR1). *Biological and Pharmaceutical Bulletin*, *27*(7), 939-948.

Jackson, C. E., & Meier, D. W. (1968). Routine serum magnesium analysis: correlation with clinical state in 5,100 patients. *Annals of Internal Medicine*, *69*(4), 743-748.

Jiang, J. J., & Conrath, D. W. (1997). Semantic similarity based on corpus statistics and lexical taxonomy. *arXiv preprint cmp-lg/9709008*.

Jinha, A. E. (2010). Article 50 million: an estimate of the number of scholarly articles in existence. *Learned Publishing*, *23*(3), 258-263.

Joon Ho Lee, Myoung Ho Kim, and Yoon Joon Lee. (1993) Information retrieval based on conceptual distance in IS-A hierarchies. *Journal of Documentation*, *49*(2), 188-207.

Joseph, R., & Welch, K. M. A. (1987). The platelet and migraine: a nonspecific association. *Headache: The Journal of Head and Face Pain*, *27*(7), 375-380.

Kan, M. Y., McKeown, K. R., & Klavans, J. L. (2001). Domain-specific informative and indicative summarization for information retrieval. In *Proc. of the Document Understanding Conference (DUC)* (pp. 19-26).

Keller, J., Kaltenecker, A., Schricker, K. T., Neidhardt, B., & Hornstein, O. P. (1985). Inhibition of platelet aggregation by a new stable prostacyclin introduced in therapy of patients with progressive scleroderma. *Archives of Dermatological Research*, *277*(4), 323-325.

Kern, W., Dodt, C., Born, J., & Fehm, H. L. (1996). Changes in cortisol and growth hormone secretion during nocturnal sleep in the course of aging. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, *51*(1), M3-M9.

Kilicoglu, H., Fiszman, M., Rodriguez, A., Shin, D., Ripple, A., & Rindflesch, T. C. (2008, September). Semantic MEDLINE: a web application for managing the results of PubMed Searches. In *Proceedings of the Third International Symposium for Semantic Mining in Biomedicine* (Vol. 2008, pp. 69-76).

Kilicoglu, H., Rosemblat, G., & Rindflesch, T. C. (2017). Assigning factuality values to semantic relations extracted from biomedical research literature. *PloS one*, *12*(7), e0179926.

Kogawa, M., TAKANO, K., HIZUKA, N., ASAKAWA, K., & SHIZUME, K. (1982). Effect of GH and insulin on the generation of somatomedin by perfused rat liver. *Endocrinologia Japonica*, 29(2), 141-147.

Kostoff, R. N. (2008). Literature-related discovery (LRD): Potential treatments for cataracts. *Technological Forecasting and Social Change*, *75*(2), 215-225.

Kostoff, R. N., Block, J. A., Solka, J. L., Briggs, M. B., Rushenberg, R. L., Stump, J. A., Johnson, D., Lyons, T. J., & Wyatt, J. R. (2009). Literature-related discovery. *Annual Review of Information Science and Technology*, *43*(1), 1-71.

Kostoff, R. N. (2014). Literature-related discovery: common factors for Parkinson's Disease and Crohn's Disease. *Scientometrics*, *100*(3), 623-657.

Krallinger, M., Rabal, O., Lourenço, A., Oyarzabal, J., & Valencia, A. (2017). Information Retrieval and Text Mining Technologies for Chemistry. *Chemical Reviews*.

Kramer, M., Netea, M. G., De Jong, D. J., Kullberg, B. J., & Adema, G. J. (2006). Impaired dendritic cell function in Crohn's disease patients with NOD2 3020insC mutation. *Journal of Leukocyte Biology*, *79*(4), 860-866.

Krueger, J. M., Obál, F., Fang, J., Kubota, T., & Taishi, P. (2001). The role of cytokines in physiological sleep regulation. *Annals of the New York Academy of Sciences*, *933*(1), 211-221.

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes*, 25(2-3), 259-284.

Le Roith, D., Shapiro, M., Jabotinsky, K., & Spitz, I. M. (1979). Prolactin nonresponsiveness to arginine in diabetes mellitus. *Hormone and Metabolic Research*, 11(10), 583-584.

Leibenluft, E., Schmidt, P. J., Turner, E. H., Danaceau, M. A., Ashman, S. B., Wehr, T. A., & Rubinow, D. R. (1997). Effects of leuprolide-induced hypogonadism and testosterone replacement on sleep, melatonin, and prolactin secretion in men. *The Journal of Clinical Endocrinology & Metabolism*, *82*(10), 3203-3207.

Li, F. Q., Wang, T., Pei, Z., Liu, B., & Hong, J. S. (2005). Inhibition of microglial activation by the herbal flavonoid baicalein attenuates inflammation-mediated degeneration of dopaminergic neurons. *Journal of Neural Transmission*, *112*(3), 331-347.

Li, J., Zhu, X., & Chen, J. Y. (2010). Discovering breast cancer drug candidates from biomedical literature. *International Journal of Data Mining and Bioinformatics*, *4*(3), 241-255.

Liddy, E. D. (2001). *Natural language processing*. Encyclopedia of Library and Information Science. Marcel Decker.

Lin, D. (1998, July). An information-theoretic definition of similarity. In *Proceedings of the 15th International Conference on Machine Learning,* (Vol. 98, No. 1998, pp. 296-304).

Lindsay, R. K., & Gordon, M. D. (1999). Literature-based discovery by lexical statistics. *Journal of the Association for Information Science and Technology*, *50*(7), 574.

Liu, Y., McInnes, B. T., Pedersen, T., Melton-Meaux, G., & Pakhomov, S. (2012, January). Semantic relatedness study using second order co-occurrence vectors computed from biomedical corpora, UMLS and WordNet. In *Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium* (pp. 363-372).

Lu, P. H., Masterman, D. A., Mulnard, R., Cotman, C., Miller, B., Yaffe, K., Reback, E., Porter, V., Swerdloff, R., & Cummings, J. L. (2006). Effects of testosterone on cognition and mood in male patients with mild Alzheimer disease and healthy elderly men. *Archives of Neurology*, *63*(2), 177-185.

López-de-Andrés, A., del Barrio, J. L., Hernández-Barrera, V., de Miguel-Díez, J., Jimenez-Trujillo, I., Martinez-Huedo, M. A., & Jimenez-García, R. (2018). Migraine in adults with diabetes; is there an association? Results of a population-based study. *Diabetes, Metabolic Syndrome and Obesity: Targets and Therapy*, 11, 367.

Lussier, Y., Borlawsky, T., Rappaport, D., Liu, Y., & Friedman, C. (2006). PhenoGO: assigning phenotypic context to gene ontology annotations with natural language processing. In *Proceedings of Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing* (p. 64). NIH Public Access.

Malkin, C. J., Pugh, P. J., Jones, R. D., Kapoor, D., Channer, K. S., & Jones, T. H. (2004). The effect of testosterone replacement on endogenous inflammatory cytokines and lipid profiles in hypogonadal men. *The Journal of Clinical Endocrinology & Metabolism*, *89*(7), 3313-3318.

Mani, I. (1999). *Advances in Automatic Text Summarization*. MIT press.

Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to information retrieval* (Vol. 39). Cambridge University Press.

Manson, J. M., Smith, R. J., & Wilmore, D. W. (1988). Growth hormone stimulates protein synthesis during hypocaloric parenteral nutrition. Role of hormonal-substrate environment. *Annals of Surgery*, 208(2), 136.

Maraganore, D. M., De Andrade, M., Lesnick, T. G., Strain, K. J., Farrer, M. J., Rocca, W. A., Pant, P.V., Frazer, K.A., Cox, D.R., & Ballinger, D. G. (2005). High-resolution whole-genome association study of Parkinson disease. *The American Journal of Human Genetics*, *77*(5), 685-693.

Markley, H. G., Cheronis, J. C., & Piepho, R. W. (1984). Verapamil in prophylactic therapy of migraine. *Neurology*, *34*(7), 973-973.

Martin, M. F., Dowd, P. M., Ring, E. F., Cooke, E. D., Dieppe, P. A., & Kirby, J. D. (1981). Prostaglandin E1 infusions for vascular insufficiency in progressive systemic sclerosis. *Annals of the Rheumatic Diseases*, *40*(4), 350.

Marzolini, C., Paus, E., Buclin, T., & Kim, R. B. (2004). Polymorphisms in human MDR1 (P-glycoprotein): recent advances and clinical relevance. *Clinical Pharmacology & Therapeutics*, *75*(1), 13-33.

McCarron, D. A. (1983). Calcium and magnesium nutrition in human hypertension. *Annals of Internal Medicine*, *98*(5_Part_2), 800-805.

McCurry, S. M., Reynolds III, C. F., Ancoli-Israel, S., Teri, L., & Vitiello, M. V. (2000). Treatment of sleep disturbance in Alzheimer's disease. *Sleep Medicine Reviews*, *4*(6), 603-628.

McGeer, P. L., & McGeer, E. G. (2004). Inflammation and neurodegeneration in Parkinson's disease. *Parkinsonism & Related Disorders*, *10*, S3-S7.

McNair, P., Christensen, M. S., Christiansen, C., Madsbad, S., & Transbøl, I. B. (1982). Renal hypomagnesaemia in human diabetes mellitus: its relation to glucose homeostasis. *European Journal of Clinical Investigation*, *12*(1), 81-85.

Merimee, T. J., Burgess, J. A., & Rabinowitz, D. (1966). Sex-determined variation in serum insulin and growth hormone response to amino acid stimulation. *The Journal of Clinical Endocrinology & Metabolism*, 26(7), 791-793.

Meyer, J. S., & Hardenberg, J. (1983). Clinical effectiveness of calcium entry blockers in prophylactic treatment of migraine and cluster headaches. *Headache: The Journal of Head and Face Pain*, *23*(6), 266-277.

Miller, C. M., Rindflesch, T. C., Fiszman, M., Hristovski, D., Shin, D., Rosemblat, G, Zhang, H., & Strohl, K. P. (2012). A closed literature-based discovery technique finds a mechanistic link between hypogonadism and diminished sleep quality in aging men. *Sleep*, *35*(2), 279-285.

Milner, J. A., Wakeling, A. E., & Visek, W. J. (1974). Effect of arginine deficiency on growth and intermediary metabolism in rats. *The Journal of Nutrition*, 104(12), 1681-1689.

Milner, J. A., & Stepanovich, L. V. (1979). Inhibitory effect of dietary arginine on growth of Ehrlich ascites tumor cells in mice. *The Journal of Nutrition*, 109(3), 489-494.

Milojević, S. (2012). How are academic age, productivity and collaboration related to citing behavior of researchers?. *PloS one*, *7*(11), e49176.

Mishra, R., Bian, J., Fiszman, M., Weir, C. R., Jonnalagadda, S., Mostafa, J., & Del Fiol, G. (2014). Text summarization in the biomedical domain: a systematic review of recent research. *Journal of Biomedical Informatics*, 52, 457-467.

Mogi, M., Togari, A., Tanaka, K. I., Ogawa, N., Ichinose, H., & Nagatsu, T. (1999). Increase in level of tumor necrosis factor (TNF)-α in 6-hydroxydopamine-lesioned striatum in rats without influence of systemic L-DOPA on the TNF-α induction. *Neuroscience Letters*, *268*(2), 101-104.

Moses, A. C., NISSLEY, S. P., Cohen, K. L., & RECHLER, M. M. (1976). Specific binding of a somatomedin-like polypeptide in rat serum depends on growth hormone. *Nature*, 263(5573), 137.

Moskowitz, M., Romero, J., Reinhard JR, J., Melamed, E., & Pettibone, D. (1979). Neurotransmitters and the fifth cranial nerve: is there a relation to the headache phase of migraine?. *The Lancet*, *314*(8148), 883-885.

Mtabaji, J. P., Manku, M. S., & Horrobin, D. F. (1976). Vascular actions of furosemide and bumetanide on the rat superior mesenteric vascular bed: interactions with prolactin and prostaglandins. *Canadian Journal of Physiology and Pharmacology*, *54*(3), 357-366.

Nakov, P. I., Schwartz, A. S., & Hearst, M. (2004). Citances: Citation sentences for semantic analysis of bioscience text. In *Proceedings of the Special Interest Group on Information Retrieval* (Vol. 4, pp. 81-88).

Nicolaisen, J. (2007). Citation analysis. *Annual Review of Information Science and Technology*, *41*(1), 609-641.

Okun, M. S., Fernandez, H. H., Rodriguez, R. L., Romrell, J., Suelter, M., Munson, S., Louid E.D., Mulligan T., Foster P.S., Shenel B.V., Armaghani, S. J., Jacobson C., Wu S., & Crucian G. (2006). Testosterone therapy in men with Parkinson disease: results of the TEST-PD Study. *Archives of Neurology*, *63*(5), 729-735.

Olanow, C. W. (2006). Rationale for considering that propargylamines might be neuroprotective in Parkinson's disease. *Neurology*, *66*(10 suppl 4), S69-S79.

Parantainen, J., Vapaatalo, H., & Hokkanen, E. (1985). Relevance of prostaglandins in migraine. *Cephalalgia*, *5*(2_suppl), 93-97.

Peatfield, R. C., Gawel, M. J., & Rose, F. C. (1981). The effect of infused prostacyclin in migraine and cluster headache. *Headache: The Journal of Head and Face Pain*, *21*(5), 190-195.

Peroutka, S. J. (1983). The Pharmacology of Calcium Channel Antagonists: a Novel Class of Anti-migraine Agents?. *Headache: The Journal of Head and Face Pain*, *23*(6), 278-283.

Pertseva, M. N., & Soltitskaia, L. P. (1987). Properties of the adenylate cyclase system in the muscle tissue of the mollusk Anodonta cygnea. *Ukrainskii biokhimicheskii zhurnal (1978)*, *59*(4), 14-20.

Peters, J. A. (1959). *Classic Papers in Genetics*. Prentice Hall.

Phillips, L. S., & Orawski, A. T. (1977). Nutrition and Somatomedin: III. Diabetic Control, Somatomedin, and Growth in Rats. *Diabetes*, 26(9), 864-869.

Phillips, L. S., Vassilopoulou-Sellin, R., & Reichard, L. A. (1979). Nutrition and Somatomedin. VIM. The "Somatomedin Inhibitor" in Diabetic Rat Serum is a General Inhibitor of Growing Cartilage. *Diabetes*, 28(10), 919-924.

Plaza, L., Díaz, A., & Gervás, P. (2011). A semantic graph-based approach to biomedical summarisation. *Artificial Intelligence in Medicine*, 53(1), 1-14.

Pratt, W., & Yetisgen-Yildiz, M. (2003, October). LitLinker: capturing connections across the biomedical literature. In *Proceedings of the 2nd International Conference on Knowledge Capture* (pp. 105-112). ACM.

Prinz, P. N., Bailey, S. L., & Woods, D. L. (2000). Sleep impairments in healthy seniors: roles of stress, cortisol, and interleukin-1 beta. *Chronobiology international*, *17*(3), 391-404.

Pui, Y. M. L., & Fisher, H. (1979). Factorial supplementation with arginine and glycine on nitrogen retention and body weight gain in the traumatized rat. *The Journal of Nutrition*, 109(2), 240-246.

Quintana, L., Konda, R., Ishibashi, Y., Yoshimoto, T., & Suzuki, J. (1982). The effect of prostacyclin on cerebral vasospasm an experimental study. *Acta Neurochirurgica*, *62*(3-4), 187-193.

Rada, R., Mili, H., Bicknell, E., & Blettner, M. (1989). Development and application of a metric on semantic nets. *IEEE Transactions on Systems, Man, and Cybernetics*, *19*(1), 17-30.

Rayssiguier, Y. (1984). Role of magnesium and potassium in the pathogenesis of arteriosclerosis. *Magnesium*, *3*(4-6), 226-238.

Reeves, R. D., Dickinson, L., Lee, J., Kilgore, B., Branham, B., & Elders, M. J. (1979). Effects of dietary composition on somatomedin activity in growing rats. *The Journal of Nutrition*, 109(4), 613-620.

Resnik, P. (1995). Using information content to evaluate semantic similarity in a taxonomy. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence.* Vol. 1, 448-453.

Rindflesch, T. C., & Fiszman, M. (2003). The interaction of domain knowledge and linguistic structure in natural language processing: interpreting hypernymic propositions in biomedical text. *Journal of Biomedical Informatics*, *36*(6), 462-477.

Rosell, S., Olgart, L., Gazelius, B., Panopoulos, P., Folkers, K., & Hörig, J. (1981). Inhibition of antidromic and substance P-induced vasodilatation by a substance P antagonist. *Acta Physiologica*, *111*(3), 381-382.

Ross, S. (2014). *A First Course in Probability*. Pearson.

Sagel, J., Colwell, J. A., Crook, L., & Laimins, M. (1975). Increased platelet aggregation in early diabetus mellitus. *Annals of Internal Medicine*, *82*(6), 733-738.

Salmon, W. D., & Daughaday, W. H. (1957). A hormonally controlled serum factor which stimulates sulfate incorporation by cartilage in vitro. *The Journal of Laboratory and Clinical Medicine*, 49(6), 825-836.

Sandborn, W. J., Rutgeerts, P., Enns, R., Hanauer, S. B., Colombel, J. F., Panaccione, R., D'Haens, G., Li, J., Rosenfeld, M.R., & Pollack, P. F. (2007). Adalimumab induction therapy for Crohn disease previously treated with infliximab: a randomized trial. *Annals of Internal Medicine*, *146*(12), 829-838.

Shi, B., & Weninger, T. (2016). Discriminative predicate path mining for fact checking in knowledge graphs. *Knowledge-Based Systems*, *104*, 123-133.

Shine, K. I. (1979). Myocardial effects of magnesium. *American Journal of Physiology-Heart and Circulatory Physiology*, *237*(4), H413-H423.

Sicuteri, F., Rainò, L., & Geppetti, P. (1983). Substance P and endogenous opioids: how and where they could play a role in cluster headache. *Cephalalgia*, *3*(1_suppl), 143-145.

Sicuteri, F., Renzi, D., & Geppetti, P. (1986). Substance P and enkephalins: a creditable tandem in the pathophysiology of cluster headache and migraine. *Advances in experimental Medicine and Biology*, *198*, 145-152.

Sjögren, A., & Edvinsson, L. (1986). Vasomotor effects of magnesium: a comparison with nifedipine and verapamil of in vitro reactivity in feline cerebral and peripheral arteries. *Magnesium*, *5*(2), 66-75.

Skinhoj, E., & Paulson, O. B. (1969). Regional blood flow in internal carotid distribution during migraine attack. *British Medical Journal*, *3*(5670), 569.

Smalheiser, N. R., & Swanson, D. R. (1996). Linking estrogen to Alzheimer's disease: an informatics approach. *Neurology*, *47*(3), 809-810.

Smalheiser, N. R., & Torvik, V. I. (2008). The place of literature-based discovery in contemporary scientific practice. *Literature-Based Discovery*, 13-22.

Smalheiser, N. R., Torvik, V. I., & Zhou, W. (2009). Arrowsmith two-node search interface: a tutorial on finding meaningful links between two disparate sets of articles in MEDLINE. *Computer Methods and Programs in Biomedicine*, 94(2), 190-197.

Smalheiser, N. R. (2012). Literature-based discovery: Beyond the ABCs. *Journal of the Association for Information Science and Technology*, *63*(2), 218-224.

Soliman, A. T., Abd El Hadi, I. H., Aref, M. K., Hintz, R. L., Rosenfeld, R. G., & Rogol, A. D. (1986). Serum insulin-like growth factors I and II concentrations and growth hormone and insulin responses to arginine infusion in children with protein-energy malnutrition before and after nutritional rehabilitation. *Pediatric Research*, 20(11), 1122.

Soma, M., Cunnane, S. C., Horrobin, D. F., Manku, M. S., Honda, M., & Hatano, M. (1988). Effects of low magnesium diet on the vascular prostaglandin and fatty acid metabolism in rats. *Prostaglandins*, *36*(4), 431-441.

Song, M., Heo, G. E., & Ding, Y. (2015). SemPathFinder: Semantic path analysis for discovering publicly unknown knowledge. *Journal of Informetrics*, *9*(4), 686-703.

Sparck Jones, K. (1972). A statistical interpretation of term specificity and its application in retrieval. *Journal of Documentation*, *28*(1), 11-21.

Spencer, E. M., Skover, G., & Hunt, T. K. (1988). Somatomedins: do they play a pivotal role in wound healing?. *Progress in Clinical and Biological Research*, 266, 103-116.
Srinivasan, P. (2004). Text mining: generating hypotheses from MEDLINE. *Journal of the American Society for Information Science and Technology*, *55*(5), 396-413.

Sriram, K., Matheson, J. M., Benkovic, S. A., Miller, D. B., Luster, M. I., & O'Callaghan, J. P. (2002). Mice deficient in TNF receptors are protected against dopaminergic neurotoxicity: implications for Parkinson's disease. *The Federation of American Societies for Experimental Biology Journal*, *16*(11), 1474-1476.

Steiner, T. J., Joseph, R., & Rose, F. C. (1985). Migraine is not a platelet disorder. *Headache: The Journal of Head and Face Pain*, *25*(8), 434-440.

Stracke, H., Schulz, A., Moeller, D., Rossol, S., & Schatz, H. (1984). Effect of growth hormone on osteoblasts and demonstration of somatomedin-C/IGF I in bone organ culture. *Acta Endocrinologica*, 107(1), 16-24.

Sutton, W. S. 1902. On the morphology of the chromosome group in Brachystola magna. *Biological Bulletin*, 4:24-39.

Swanson DR. Fish oil, Raynaud's syndrome, and undiscovered public knowledge. *Perspectives in Biology and Medicine* 1986;30(1):7–18.

Swanson, D. R. (1988). Migraine and magnesium: eleven neglected connections. *Perspectives in Biology and Medicine*, *31*(4), 526-557.

Swanson, D. R. (1989). Online search for logically-related noninteractive medical literatures: a systematic trial-and-error strategy. *Journal of the American Society for Information Science (1986-1998)*, *40*(5), 356.

Swanson, D. R. (1990). Somatomedin C and arginine: implicit connections between mutually isolated literatures. *Perspectives in Biology and Medicine*, *33*(2), 157-186.

Swanson, D. R., Smalheiser, N. R., & Torvik, V. I. (2006). Ranking indirect connections in literature-based discovery: The role of medical subject headings. *Journal of the American Society for Information Science and Technology*, *57*(11), 1427-1439.

Swanson, D. R. (2008). Literature-based discovery? The very idea. *Literature-Based Discovery*, 3-11.

Taishi, P., Chen, Z., OBÁL JR, F. E. R. E. N. C., Hansen, M. K., Zhang, J., Fang, J., & Krueger, J. M. (1998). Sleep-associated changes in interleukin-1β mRNA in the brain. *Journal of Interferon & Cytokine Research*, *18*(9), 793-798.

Terano, T., Hirai, A., Hamazaki, T., Kobayashi, S., Fujita, T., Tamura, Y., & Kumagai, A. (1983). Effect of oral administration of highly purified eicosapentaenoic acid on platelet function, blood viscosity and red cell deformability in healthy human subjects. *Atherosclerosis*, *46*(3), 321-331.

Trenell, M. I., Marshall, N. S., & Rogers, N. L. (2007). Sleep and metabolic control: waking to a problem? *Clinical and Experimental Pharmacology and Physiology*, *34*(1-2), 1-9.

Turlapaty, P. D., Weiner, R., & Altura, B. M. (1981). Interactions of magnesium and verapamil on tone and contractility of vascular smooth muscle. *European Journal of Pharmacology*, *74*(4), 263-272.

Tweedie, D., Sambamurti, K., & Greig, N. H. (2007). TNF-α inhibition as a treatment strategy for neurodegenerative disorders: new drug candidates and targets. *Current Alzheimer Research*, *4*(4), 378-385.

Urcelay, E., Mendoza, J. L., Martín, C. M., Mas, A., Martínez, A., Taxonera, C., Fernandez-Arquero, M., Diaz-Rubio, M., & de la Concha, E. G. (2006). MDR1 gene: susceptibility in Spanish Crohn's disease and ulcerative colitis patients. *Inflammatory Bowel Diseases*, *12*(1), 33-37.

Vallance, B. A., Gunawan, M. I., Hewlett, B., Bercik, P., Van Kampen, C., Galeazzi, F., Sime, P.J., Gauldie, J., & Collins, S. M. (2005). TGF-β1 gene transfer to the mouse colon leads to intestinal fibrosis. *American Journal of Physiology-Gastrointestinal and Liver Physiology*, *289*(1), G116-G128.

Van Harreveld, A. (1978). Two mechanisms for spreading depression in the chicken retina. *Developmental Neurobiology*, *9*(6), 419-431.

Van Harreveld, A. (1984). The nature of the chick's magnesium-sensitive retinal spreading depression. *Developmental Neurobiology*, *15*(5), 333-343.

Wahner, A. D., Bronstein, J. M., Bordelon, Y. M., & Ritz, B. (2007). Nonsteroidal anti-inflammatory drugs may protect against Parkinson disease. *Neurology*, *69*(19), 1836-1842.

Walther, H., Lambert, J. D. C., Jones, R. S. G., Heinemann, U., & Hamon, B. (1986). Epileptiform activity in combined slices of the hippocampus, subiculum and entorhinal cortex during perfusion with low magnesium medium. *Neuroscience Letters*, *69*(2), 156-161.

Weeber, M., Klein, H., de Jong-van den Berg, L., & Vos, R. (2001). Using concepts in literature-based discovery: Simulating Swanson's Raynaud –fish oil and migraine–magnesium discoveries. *Journal of the Association for Information Science and Technology*, *52*(7), 548-557.

Weeber, M., Vos, R., Klein, H., Aronson, A. R., & Molema, G. (2003). Generating hypotheses by discovering implicit associations in the literature: a case report of a search for new potential therapeutic uses for thalidomide. *Journal of the American Medical Informatics Association*, *10*(3), 252-259.

Weeber, M., Kors, J. A., & Mons, B. (2005). Online tools to support literature-based discovery in the life sciences. *Briefings in Bioinformatics*, *6*(3), 277-286.

Welch, K. M. A., Chabi, E., Nell, J. H., Bartosh, K., Chee, A. N. C., Mathew, N. T., & Achar, V. S. (1976). Biochemical comparison of migraine and stroke. *Headache: The Journal of Head and Face Pain*, *16*(4), 160-167.

Wilkowski, B., Fiszman, M., Miller, C. M., Hristovski, D., Arabandi, S., Rosemblat, G., & Rindflesch, T. C. (2011). Graph-based methods for discovery browsing with semantic predications. In *Proceedings of the American Medical Informatics Association Annual Symposium* (Vol. 2011, p. 1514). American Medical Informatics Association.

Wren, J. D. (2004). Extending the mutual information measure to rank inferred literature relationships. *BMC Bioinformatics*, *5*(1), 145.

Wren, J. D., Bekeredjian, R., Stewart, J. A., Shohet, R. V., & Garner, H. R. (2004). Knowledge discovery by automated identification and ranking of implicit relationships. *Bioinformatics*, 20(3), 389-398.

Wren, J. D. (2008). The 'open discovery' challenge. *Literature-Based Discovery*, 39-55.

Xun, G., Jha, K., Gopalakrishnan, V., Li, Y., & Zhang, A. (2017, November). Generating Medical Hypotheses Based on Evolutionary Medical Concepts. In *Data Mining (ICDM), 2017 IEEE International Conference on Data Mining* (pp. 535-544). IEEE.

Yetisgen-Yildiz, M., & Pratt, W. (2006). Using statistical and knowledge-based approaches for literature-based discovery. *Journal of Biomedical Informatics*, *39*(6), 600-611.

Yetisgen-Yildiz, M., & Pratt, W. (2008). Evaluation of literature-based discovery systems. In *Literature-Based Discovery* (pp. 101-113). Springer, Berlin, Heidelberg.

Yogan, J. K., Goh, O. S., Halizah, B., Ngo, H. C., & Puspalata, C. (2016). A review on automatic text summarization approaches. *Journal of Computer Science*, 12(4), 178-190.

Youn, H., Strumsky, D., Bettencourt, L. M., & Lobo, J. (2015). Invention as a combinatorial process: evidence from US patents. *Journal of The Royal Society Interface*, *12*(106), 20150272.

Zhang, G., Ding, Y., & Milojević, S. (2013). Citation content analysis (cca): A framework for syntactic and semantic analysis of citation content. *Journal of the Association for Information Science and Technology*, *64*(7), 1490-1503.

Zhang, H., Fiszman, M., Shin, D., Miller, C. M., Rosemblat, G., & Rindflesch, T. C. (2011). Degree centrality for semantic abstraction summarization of therapeutic studies. *Journal of Biomedical Informatics*, *44*(5), 830-838.

Zhang, H., Fiszman, M., Shin, D., Wilkowski, B., & Rindflesch, T. C. (2013). Clustering cliques for graph-based summarization of the biomedical research literature. *BMC Bioinformatics*, *14*(1), 182.

Zipf, G. K. (1949). *Human Behaviour and the Principle of Least-Effort*.