Diversity in citations to a single study: A citation context network analysis of how evidence from a prospective cohort study was cited

Rhodri Ivor Leng

The University of Edinburgh, Department of Science, Technology, and Innovation Studies

Keywords: citation analysis, citation context analysis, network analysis, scientometrics, sociology of science, STS

ABSTRACT

Between its origin in the 1950s and its endorsement by a consensus conference in 1984, the diet–heart hypothesis was the subject of intense controversy. Paul et al. (1963) is a highly cited prospective cohort study that reported findings inconvenient for this hypothesis, reporting no association between diet and heart disease; however, many other findings were also reported. By citation context and network analysis of 343 citing papers, I show how Paul et al. was cited in the 20 years after its publication. Generally, different findings were cited by different communities focusing on different risk factors; these communities were established by either research foci titles or via cluster membership as established via modularity maximization. The most frequently cited findings were the significant associations between heart disease and serum cholesterol (n = 85), blood pressure (n = 57), and coffee consumption (n = 54). The lack of association between diet and heart disease was cited in just 41 papers. Yet, no single empirical finding was referred to in more than 25% of the citing papers. This raises questions about the value of inferring impact from citation counts alone and raises problems for studies using such counts to measure citation bias.

1. INTRODUCTION

In a paper with many empirical findings, such as a study examining the association of a range of risk factors with a disease, what findings are actually cited in the scientific literature—what exactly gives that paper its impact? This question has new relevance due to the growing use of citation context analysis to study the semantic content of references to gauge the impact of specific works on understanding (Anderson & Lemken, 2019; Bommann, Wray, & Haunschild, 2020), as well as the biases and errors that can distort that understanding (Greenberg, 2009; Leng, 2018).

As highlighted by Shkurko (2018), much attention in scientometrics has focused on understanding why different papers have different impacts, but less on the varying impacts that different elements within the same paper come to have. This echoes an earlier criticism by Small (1978) of the then absence of work examining the different content authors can use from a cited paper. To rectify this, Small (1978) extended citation context analysis (Moravcsik & Murugesan, 1975) by identifying the claims that papers become associated with in the citing
literature. In a study of 52 highly cited works in chemistry, Small sampled 12 citing papers of each paper, extracted the passages in which these were discussed, and classified the citations by their “symbolic” content. He found that most of these papers tended to have a “uniform and standardized usage and meaning” in the citing literature. However, his sample contained primarily methodological works, and he noted that other kinds of works, such as the empirical papers of a research front (Price, 1965), might be cited in a more diverse manner.

Small argued that it was essential to perform citation context analysis to understand what a paper is being cited for. Citing authors can, and do, attach meanings to a cited document different from those intended by the authors, and the social selection (Gilbert, 1977) of relevant material from a paper can introduce oversimplifications, distortions, and reinterpretations that may spread to become dominant in a literature: “In the extreme case this means that there need not be any similarity between the document and the concept it stands for” (Small, 1978, p. 329).

Since then, a number of studies have focused on what content from particular articles is actually cited (Cozzens, 1985, 1988; Horbach, Aagaard, & Schneider, 2021; Lu, Ding, & Zhang, 2017; Small & Greenlee, 1980; Stang, Jonas, & Poult, 2018), or on tracing what material from influential academic books is used (Anderson, 2006; Bornmann et al., 2020; Crothers, Bornmann, & Haunschild, 2020; Dewey, 2016; McCain & Salvucci, 2006). Recently, Shkurko (2018, p. 84), in a study of three randomly chosen journal articles in the neurosciences, found that each was cited for a diversity of content, and he concluded that “[P]otentially any scientific article can be treated in numerous ways and contain many cognitive elements which can be differently valued by various subpopulations of scientists.” That a given work can be cited for different content by different communities is a consistent finding from such studies (e.g., Bornmann et al., 2020; Cozzens, 1988; McCain & Salvucci, 2006), though it appears that commonly cited papers come to be associated with a particular claim within literature focused on the same research topic (Schneider, 2006; Small, 1986, 2011; Small & Greenlee, 1980). While most studies have not observed the high degree of uniformity in citations observed by Small (1978), there are notable exceptions: most strikingly, Stang et al. (2018) reported that a paper (Stang, 2010) written to detail the major shortcomings of a methodology used in meta-analyses to measure study quality was cited very highly, but almost always wrongly, as though it endorsed the use of that method.

This last case brings me to why citation context analysis has found new relevance. There is currently concern in both the biomedical and social sciences about the reliability of published evidence, and in part this relates to how scientists reference (Harzing, 2002; Leng & Leng, 2020). These concerns include errors in the interpretation of cited works (Jergas & Baethge, 2015; Mogull, 2017), and in some cases such errors have become dominant interpretations in the literature (Leung, MacDonald et al., 2017; Stang et al., 2018). Concerns have been raised over the persistence of uncritical citations to retracted studies (Schneider, Ye et al., 2020); to studies that have failed to replicate (Serra-Garcia & Gneezy, 2021); and to studies that have apparently been refuted (Tatsioni, Bonitis, & Ioannidis, 2007). Concern has also been expressed about the tendency of scientists to reference findings without acknowledging their limitations (Cristea & Naudet, 2018). Finally, there are concerns about citation bias—the tendency to reference studies that support a particular claim while ignoring those that are inconvenient (Gøtzsche, 1987). Greenberg’s (2009) study of the damaging influence that selective citing practices had on his own field has spurred increasing interest. By combining systematic review with citation context and network analysis, Greenberg demonstrated how a claim—that a particular protein, β-amyloid, was abnormally present in the muscle fibers of patients with inclusion body myositis—came to be accepted as a fact in the scientific literature despite an underlying ambiguous evidence base; the relevant research community had
apparently been oblivious to the existence of unsupportive data via an “information cascade,” whereby scientists use the reference lists of other papers to find relevant literature, and, in some cases, copy their interpretations of earlier literature. Other studies of citation bias have done so in the context of particular controversies, and here evidence points to biases becoming prevalent in particular communities advocating particular positions in a debate (Leng, 2018; Trinquart, Johns, & Galea, 2016). Thus, understanding what a work is cited for and why, as well as simple methods for detecting variation in use of scientific papers, is currently of broad interest.

In the study that follows, by a combination of citation context and network analysis, I trace how the findings of the Western Electric Study (Paul, Lepper et al., 1963), a study examining the relationship between lifestyle factors and coronary heart disease (CHD), came to be used over a 21-year period in which it was part of a major debate. This was the first prospective cohort study of the relationship between dietary fats and CHD. In the 1950s, the “diet–heart hypothesis”—the conjectured causal relationship between dietary fat consumption and CHD—had dominated scientific discussion of the etiology of CHD. But this hypothesis was still controversial, with critics attacking what they saw as the circumstantial nature of the evidence base. It was expected that the results of several prospective cohort studies, due to return their findings over the course of the 1960s and 70s, would settle the question. Yet, Paul et al., the first of these to return its dietary findings, found no association between the consumption of dietary fats and CHD.

Paul et al. has been well cited—673 times by 2020, according to the Web of Science (WoS). This seems surprising, because, according to a recent meta-analysis, “negative” studies are generally relatively poorly cited (Duyx, Urlings et al., 2017). Furthermore, previous studies have found evidence of marked citation bias against unsupportive findings in the diet–heart literature (Leng, 2018; Ravnskov, 1992, 1995). These examined how the findings of particular controlled trials had been cited, rather than observational studies. Paul et al. is an early example of the latter, and it produced evidence inconvenient for the diet–heart hypothesis, but its high number of citations suggests that it was not ignored. However, Paul et al. also reported on the relationship between other variables and CHD. Accordingly, it is unclear from the citation count alone what impact the dietary fat findings had. These were a key rationale of the study: They were discussed extensively in the text of Paul et al. and they had a central relevance to a major ongoing debate—but how much attention were they actually given by the citing literature?

To examine this, I build on previous works that have examined the semantic content of citations via citation context analysis. Small (1978) argued that scientists reference particular documents for particular ideas, findings, or methods, and they offer their interpretation of the meaning of that information in the passage accompanying the in-text reference. He argued that highly cited works have a tendency to act as “concept symbols” in the literature; works that are commonly cited for the same content. To establish what a work is most commonly cited for, Small introduced “percent uniformity,” which measures the proportion of citation contexts from a sample of the citing literature that refer to the same content. Small and Greenlee (1980) then showed that co-citation network analysis combined with citation context analysis could be used to construct maps of particular literatures in which papers are annotated by the most commonly cited finding, idea, or method. While this approach has been used effectively to produce cognitive maps of particular research areas (Schneider, 2006; Small, 1986), these studies have highlighted that empirical research tends to be cited more heterogeneously than methodological contributions. Small (1978) found a mean uniformity of 87% in citations to 52 highly cited chemistry papers (mainly methodological contributions), but Schneider (2006)
found a mean uniformity of 52% for 64 papers in periodontology. Cozzens (1988) has argued that an important source of variability of cited content is whether or not a paper is cited by different communities focused on different research questions. While Cozzens (1988) determined the research foci of citing papers via analysis of their titles in her empirical case study, recent studies examining the variation in citation impact have either relied on the broad disciplinary categories offered by bibliometric archives (Bornmann et al., 2020; McCain & Salvucci, 2006) or have ignored the topic of the citing paper (Shkurko, 2018). From this literature, we might expect that a paper with many distinct findings would have a higher probability of being cited in a more varied manner than a work reporting few, particularly if that paper is being cited by literatures focused on different topics. Here, I offer an approach that can track the dissemination of findings from particular papers across the literature, as well as identifying communities of citing papers that focus on specific research topics and the content that they frequently cite from that document. For this, I combine citation context analysis and with citation network analysis, classifying citation contexts by the findings referred to, and classifying citing papers by both research foci title terms and cluster membership as established by modularity maximization via the Leiden algorithm (Traag, Waltman, & van Eck, 2019) on a network composed of the direct citation links between citing papers. Thus, this approach is designed explicitly to track variation in use of specific studies and may be of interest for those pursuing case studies in which particular documents are suspected to be used in different manners in different literatures.

2. BACKGROUND

The “diet–heart hypothesis” proposed that the consumption of saturated fatty acids (SFA), by increasing serum cholesterol levels, contributes to CHD via the development of atherosclerosis (Keys, Anderson, & Grande, 1957a). In the 1950s, this hypothesis dominated scientific discussion of the etiology of CHD. But it was controversial, with critics attacking what they saw as the circumstantial nature of the evidence base (Mann, 1959). In the late 1940s, two prospective cohort studies had been launched to test the link between serum cholesterol and CHD, and in the 1960s these both reported a significant association between serum cholesterol levels and CHD (Kannel, Dawber et al., 1961; Keys, Taylor et al., 1963). This suggested that lowering serum cholesterol might protect against CHD, and three strands of evidence suggested that dietary manipulation might achieve this: (a) Atherosclerosis-like lesions could be produced in the arteries of rabbits by introducing cholesterol into their diet (Anitschkow, 1933); (b) short-term feeding trials in humans had found that SFA elevated cholesterol levels (Bronte-Stewart, Antonis et al., 1956; Keys et al., 1957a); and (c) a reported association between the amount of fat consumed in certain countries and increased CHD mortality measured via national food consumption data and vital statistics (Keys, 1953).

By the end of the 1950s, scientists were contemplating whether population-wide advice to lower the consumption of SFA might help reduce CHD incidence (Keys, Anderson, & Grande, 1957b). However, evidence was needed to confirm that dietary fat was associated with serum cholesterol and CHD in populations. Two observational study designs were used to test this: (a) prospective cohort studies that measured dietary factors and serum cholesterol before the manifestation of CHD within a specific community; and (b) prospective studies that examined diet and serum cholesterol and their relationship to CHD between cohorts.

The best-known example of the latter is the Seven Countries Study (Keys, 1970), which examined diet, serum cholesterol, and CHD in 12,763 men in seven countries. By splitting this sample into 16 cohorts that represented separate communities within each country, the
study tested whether differences in diet could explain the differences between cohorts in serum cholesterol and CHD mortality. The findings indicated that SFA intake was positively associated with both serum cholesterol and CHD.

By contrast, prospective cohort studies that examined the relationship between diet, serum cholesterol, and CHD within communities produced generally unsupportive findings (Ravnskov, 1998). By the end of 1984, 10 such studies had been published (Garcia-Palmieri, Sorlie et al., 1980; Gordon, Kagan et al., 1981; Kannel & Gordon, 1970; Kromhout & de Lezenne Coulander, 1984; McGee, Reed et al., 1984; Medalie, Kahn et al., 1973; Morris, Marr, & Clayton, 1977; Paul et al., 1963; Shekelle, Shryock et al., 1981; Yano, Rhoads et al., 1978). Three of these studies reported that those who developed myocardial infarction or who had died from CHD had significantly higher intakes of SFA when expressed as a percentage of total calories (Gordon et al., 1981 [Honolulu cohort]; McGee et al., 1984; Yano et al., 1978); however, only one of these found a significant relationship after controlling for other known risk-factors (McGee et al., 1984).

Thus SFA intake appeared to predict the occurrence of CHD between communities, but not within communities. This discordance provoked controversy. Stallones (1983, p. 168) observed that “results of research aimed at testing the hypothesis experimentally have been used to support opposed positions.” In 1984, in response to what appeared to be a mounting controversy, the National Institutes for Health (NIH, 1985) gathered a panel of experts to resolve whether the link between diet, raised serum cholesterol, and CHD was causal. They concluded that raised serum cholesterol was a cause of CHD and proposed that a diet containing no more than 10% of total calories derived from SFA ought to be advised for all members of the public over the age of two. Dietary guidelines promoting a reduction in total fat and SFA were adopted in many countries (Kritchevsky, 1998). By the end of 1984, before the NIH had issued its consensus statement, Paul et al. had become the most highly cited of the cohort studies to test the diet–heart hypothesis, attracting 46% of the 998 citations to 10 studies described previously. So what did this study find?

2.1. Findings of the Western Electric Study

Launched in 1957, the Western Electric Study (Paul et al., 1963) examined the diets of 1,989 men aged between 40 and 55 years from a Chicago factory. It reported results after 53 months of follow-up, by which time 88 men had developed CHD. SFA, unsaturated fatty acids, total fat, and dietary cholesterol were not associated with CHD. Men without CHD ate an average of 3,174 calories/day and 59 g/day of SFA; men with CHD ate 3,082 calories/day and 59 g/day of SFA. The authors then split the cohort into sub-groups representing the top and bottom 15% of total fat intake. Men in the high-fat group ate an average (median) of 49% calories from fat compared to 36% in the low-fat group. There were 16 CHD cases in the low-fat group and 14 in the high-fat group. Paul et al. concluded that: “The findings […] would seem to inject a healthy note of caution at least toward attempts to alter the American diet within the modest range of fat intake we have described.” (p. 30)

However, Paul et al. also reported many other findings. They drew attention to the lack of association between CHD and 21 dietary variables (including intakes of fats, carbohydrate, protein, salt, and ascorbic acid; body weight; height; alcohol consumption; blood sugar levels; hemoglobin levels; lipoprotein lipase levels; job type; and job-related physical activity. But they also reported significant relationships between CHD and elevated serum cholesterol; elevated blood pressure; coffee intake; cigarette smoking; age; family history (early death of father); increased skinfold thickness; endomorphic dominance; arteriovenous nicking in the
fundi; electrocardiogram abnormalities; and histories of peptic ulcer, chronic cough, shortness of breath, and “non-cardiac” chest discomfort.

While the sample size of nearly 2,000 men was large for its time, this study ultimately comprised an analysis of associations of many variables with an endpoint that affected only 88 men. In retrospect, it seemed likely to produce abundant false positive results if the analysis did not fully correct for multiple comparisons, or many false negative results where the effect size was small. Paul et al. (1963, p. 20) also did not report the findings from all of the variables examined, but only a “portion” of them, observations that were believed important because of the current theories of CHD. Accordingly, it is possible that all reported associations were due to chance alone.

Despite these limitations, Paul et al. was cited 446 times (WoS, September 2020) between 1963 and 1984. With 673 citations as of 2020, it is within the top 0.2% of the ~1.5 million primary research articles published in the 1960s. But which of its many findings was this study cited for?

3. MATERIALS AND METHODS

3.1. Overview of Study Design

To understand this, I make use of content analysis of citation contexts to establish which findings from Paul et al. (1963) were cited in the period before 1985, and then citation network analysis is performed to understand which findings were circulating in particular communities. Finally, an interpretative account of the history of the impact of Paul et al. contextualizes these findings by tracing the spread of five highly cited findings over time and their changing interpretations. All data associated with the following analysis are available via Leng (2021), including the code book used for classification.

3.2. Cited Reference Search

All documents citing Paul et al. recorded in the WoS Core Collection were identified via a cited reference search in September 2020. This returned 446 citing papers published before 1985 (673 by September 2020). Restricting to journal publications in English reduced the set to 408 papers. In similar searches on other bibliographic databases: Google Scholar returned 375 citing papers before 1985; Semantic Scholar returned 304; Dimensions returned 281; Microsoft Academic returned 279; Scopus returned 163; and Scite returned 36 papers with in-text citation data. Semantic Scholar also captures the text surrounding in-text citations, but these data were only available for eight papers published before 1985. As WoS was the most comprehensive set, this was chosen for the following analysis. All bibliographic data were downloaded, including the reference lists of citing papers.

3.3. Content Analysis of Citation Contexts

Paul et al. was read in full and its findings categorized, producing a list of 34 finding categories. These include the association between CHD and all 21 dietary variables combined into a single “diet” category; other categories included the association between CHD and serum cholesterol, blood pressure, coffee intake, alcohol intake, and cigarette smoking. Two categories were used for citations to Paul et al. that did not discuss reported findings: (a) “Mention” for papers citing Paul et al. for no specific finding or for methods; and (b) “Misattributed” for papers that cited Paul et al. for findings that did not appear in that paper.

As some papers in non-digitalized journals were difficult to access, only 347 full-text documents were retrieved. Two of these referred to Paul et al. in the reference list but not in-text,
one was digitalized without its reference list, and one was a duplicate of another paper despite having a different WoS accession number, title, and author—leaving 343 documents. These papers were read, and quotations that captured the passage in which Paul et al. was cited were copied into an Excel spreadsheet.

A paper was classified as having cited a specific finding category if they reported the finding exactly or used it to support a related claim (see Cozzens (1988), Shkurko (2018), and Small (1978) for similar designs). For example, Leren (1966, p. 5) discusses Paul et al. as follows: “The long-term study at the Western Electric Company in Chicago [Paul et al., 1963] demonstrated a definitive association between the serum cholesterol level and the development of CHD”; this is classified as a citation to the serum cholesterol findings. As papers can cite multiple findings, each cited finding was recorded. For example, Little and Shanoff (1965, p. 184) wrote “Paul et al. showed that, except for coffee, there was no association between diet and later incidence of C.H.D.,” and this is classified as citing both the dietary and coffee findings. If a paper cited Paul et al. several times in different paragraphs, these were copied separately. If separate passages referred to the same category of finding, these were counted as a single citation to that finding. Accordingly, a paper can cite a specific category at most once, but can cite multiple categories.

Dietary findings were then classified into whether the authors specifically discussed dietary fat, carbohydrate, protein, vitamins, or minerals, or whether the lack of association between diet and CHD was highlighted without mentioning specific nutrients. While the above example by Little and Shanoff is an example of highlighting the lack of relationship between diet and CHD, the following passage by Meade (1973, p. 645) is an example of a citation to the dietary fat findings: “Two studies which have been able to link individual dietary fat intake data to subsequent IHD [CHD] incidence have both failed to establish any relationship (Paul et al., 1963, Kannel & Gordon, 1970).”

During analysis it became clear that dietary findings were discussed in a more heterogeneous manner than other findings, and a further classification scheme was required to distinguish between studies that (a) discussed the lack of association between diet or particular nutrients and CHD; (b) discussed Paul et al. to compare dietary intakes of that study population with other populations without discussing the relationship to CHD; and (c) discussed the

Figure 1. Citation history of Paul et al. (1963) as of September 2020 (n = 673 citing papers) as recorded by WoS.
lack of association between diet and its relationship to serum cholesterol. The last of these findings was not reported by Paul et al., except to note the inability to explain seasonal fluctuations in serum cholesterol by dietary intake in this sample. Examining the top 15% and bottom 15% of dietary fat intake, the authors only report that those in the high-fat group had a serum cholesterol of 253.8 mg/100 ml whereas those in the low-fat group had 246.8 mg/100 ml. As the authors do not report the standard deviations, it is not possible to ascertain whether there is a significant difference between these values, nor is it possible to establish whether a relationship exists between these variables via secondary analysis.

3.4. Title Classification: Research Focus

Papers were classified by analysis of their titles. Title words have often been used to classify papers (Cozzens, 1988; Leydesdorff, 1989; Small & Greenlee, 1980), and as no keywords or abstracts are indexed for the period examined here, this was the only feasible method. WoS “research categories” were unsuitably broad (Small & Griffith, 1974). Papers were classified first by any *explanans* (a proposed risk-factor) and second by the *explanandum* (a disease). For example, “Coffee, Alcohol and Risk of Coronary Heart-Disease among Japanese Men Living in Hawaii” was classified as a member of both “Coffee” and “Alcohol” groups in the first level and “CHD” in the second. Papers without a clear *explanans* or *explanandum* were classified as “Unspecified.” For example, “Coronary Heart-Disease - Global Picture” was classified as “Unspecified” in the first level and “CHD” in the second.

3.5. Network Analysis

Paul et al. and 446 citing papers were given identifier codes, and data on authors, paper title, year of publication, WoS accession number were recorded in a node-attribute-list. For all 343 papers with available citation context data, the research focus and title classifications described previously were parsed into this data set. Sci² Tool version 1.3 (Sci² Team, 2009) was used to extract citation links between papers, and these were stored as an “edge-list” in a .csv file. A directed citation network was constructed in Gephi version 0.92 (Bastian, Heymann, & Jacomy, 2009) representing papers as nodes and citation links as directed edges between nodes—with edges going from a citing paper to the cited paper. Duplicate publication records caused by reference variants (Marx, 2011) were merged. Two of the 446 citing papers were removed from analysis due to being duplicates of other publications despite having unique WoS accession numbers. One is explained by the republication of an article in a foreign language, while the other appears to be an indexing error by WoS. Modularity maximization (Newman & Girvan, 2004) was used to partition the network into clusters based on their citation links via the Leiden algorithm (Traag et al., 2019). This algorithm constructs clusters by iteratively moving nodes from one cluster to another to establish clusters in which the fraction of edges between nodes within a cluster is higher than would be expected if edges were distributed randomly. Typically, papers that are densely interconnected together by citation and reference links tend to share a research topic, and modularity maximization has been shown to be very effective for community detection in directed citation networks (Klavans & Boyack, 2017).

3.6. Descriptive Statistics

All research categories and all specific findings mentioned in each citing paper were counted. Following Bornmann et al. (2020), the relationship between cited finding and research category was represented as a contingency table, as was the relationship between publication data of the citing paper and the tendency to cite specific findings. To understand the tendency for
findings to co-occur in citing papers a contingency matrix was used. A similar analysis was performed to understand the relationships between cited finding and cluster membership. As this is an exploratory study in which findings were not presumed to be generalizable to a broader population, no hypothesis testing statistics were used.

4. RESULTS

4.1. What Findings Were Cited?

Of the 343 papers analyzed, 38 mentioned Paul et al. but none of its specific findings. These included follow-up studies that pointed to the original study, studies citing Paul et al. due to use of similar methods, and papers that simply cited it as an example of a prospective cohort study. The other 305 papers referred to up to 18 finding categories, but most (231) referred to just one. The most commonly cited findings were the significant associations between CHD and serum cholesterol (85 papers); blood pressure (57 papers); and coffee consumption (54 papers).

The dietary findings were referred to in 52 papers. Of these, 41 cited the lack of significant associations between diet (or specific dietary variables) and CHD. Of these, 23 specifically mentioned the lack of association between dietary fats and CHD, 15 discussed the lack of association between diet and CHD without specifying nutrients, two discussed the lack of association between dietary cholesterol and CHD, and one discussed the lack of association between salt consumption and CHD. Another 12 papers discussed dietary findings in different contexts; of these, seven referred only to the lack of relationship between diet and serum cholesterol level, while five simply mentioned dietary findings to compare to intakes in their own study participants. Thus, of the 343 papers that cited Paul et al., only 41 (12%) mentioned the lack of association between any dietary variable or the diet in general and CHD. Figure 2 shows all finding categories cited by five or more of the 343 citing papers.

![Figure 2](image-url)  
Citations to specific findings in Paul et al. (1963) from 343 papers published between 1963 and 1984. See Leng (2021) for the code book containing finding category descriptions and supplementary data for counts relevant to 35 categories of cited findings.
4.2. What Findings Were Cited Together?

Table 1 shows the co-occurrence of the above eight most cited findings. A total of 242 papers mentioned the association (or absence of association) between CHD and serum cholesterol, dietary variables, blood pressure, coffee consumption, cigarette smoking, alcohol consumption, physical activity levels, and body fatness. Of these, 182 (75%) cited only one of the above findings, with 60 (25%) citing more than one. The most commonly co-occurring findings were the associations between CHD and elevated serum cholesterol and elevated blood pressure (n = 42). In 24 papers, both of these were cited alongside the association between cigarette smoking and CHD. Only 13 of the 85 papers that discussed the relationship between serum cholesterol and CHD also cited any of the dietary findings.

4.3. When Were Findings Cited?

Table 2 summarizes how the tendency to cite findings varied over time. Most findings attracted most attention in the first decade following publication. Citations tended to peak before 1970, after which citations tended to decrease, a pattern of citation typical at the article level (Larivière, Archambault, & Gingras, 2008; Price, 1965; Parolo, Pan et al., 2015). An exception is the citations to the lack of association between alcohol and CHD; this only begins being cited from 1973, and citations peaked between 1979 and 1981. The delay appears to be explained by the lack of a community focused on the relationship between alcohol and CHD until the 1970s; this emerged only when later studies reported a significant association between alcohol intake and CHD. Paul et al. tended to be cited in discussions about whether alcohol intake really was beneficial, and most authors used the null finding to highlight the then conflicted evidence-base: “Epidemiologic follow-up studies have yielded apparently conflicting data. No association between alcohol consumption and clinical evidence of coronary artery disease was found in the Western Electric Study [7], a negative association in a study of Kaiser Foundation Health Plan subscribers [8], and a positive association in Gothenburg, Sweden [9].” (Stason, Neff et al., 1976, p. 604).

4.4. What Research Foci Were Citing Paul et al?

The most common explanandum in the paper titles is “CHD”, a category including all variant terms for myocardial infarction, angina pectoris, and terms related to coronary atherosclerosis (n = 229). This is followed by “serum cholesterol” (n = 46), which includes terms related to serum

<table>
<thead>
<tr>
<th>Cited finding, number of citing papers</th>
<th>SC</th>
<th>BP</th>
<th>C</th>
<th>D</th>
<th>S</th>
<th>A</th>
<th>PA</th>
<th>BF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Serum cholesterol (SC), n = 85</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blood pressure (BP), n = 57</td>
<td>42</td>
<td></td>
<td>6</td>
<td>9</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Caffeine (C), n = 55</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diet (D), n = 41</td>
<td>13</td>
<td>5</td>
<td>9</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Smoking (S), n = 43</td>
<td>24</td>
<td>24</td>
<td>8</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alcohol (A), n = 22</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physical activity (PA), n = 28</td>
<td>6</td>
<td>5</td>
<td>1</td>
<td>3</td>
<td>5</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Body fatness (BF), n = 34</td>
<td>14</td>
<td>14</td>
<td>3</td>
<td>4</td>
<td>9</td>
<td>1</td>
<td>5</td>
<td></td>
</tr>
</tbody>
</table>

Quantitative Science Studies 10
cholesterol or lipoproteins. "Blood pressure" or “hypertension” appeared as explanandum in 18 papers, “cancer” in eight, and “stroke” or cerebral atherosclerosis in seven, with the remaining papers concerning diverse topics. In 33 papers, the title contained no clear explanandum.

The most common explanans was the diet (n = 64), excluding caffeine and alcohol. Terms indicating a focus on caffeine (or coffee) were in 34 papers; psychosocial factors in 29; alcohol consumption in 25; serum cholesterol in 22; physical activity in 19; weight or obesity in 18; smoking (or nicotine intake) in 16; and blood pressure (or hypertension) in 14. No other explanans was mentioned in more than 10 papers. A total of 87 papers had no clear explanans in the title (i.e., risk-factor term + disease term); most of these included “CHD” as explanandum, and most were addressing multiple risk factors.

As 33 papers contained no explanandum in the title and 87 contained no explanans, the following analysis counted the most frequent research classifications across both levels to establish the 10 most common topics. A total of 244 papers had terms in either explanandum or explanans that indicated a focus on diet, caffeine, alcohol, serum cholesterol, blood pressure, physical activity, or obesity. As most of the citing literature was focused on understanding CHD (n = 229), and because most of the common risk factors that were highlighted in titles occurred together with terms related to CHD, a group of “Unspecified CHD” was created for all papers that did not include terms indicating a focus on the nine most frequently mentioned risk factors (n = 83) but that did have terms related to CHD. By this, 327 (95%) of the 343 citing papers are classified in at least one of the below 10 categories (Figure 3).

To understand the relationship between research foci and cited finding, Table 3 shows the above 10 title categories and the eight most common cited empirical findings concerning the relationship between factors examined and CHD incidence as described in Figure 2.
This analysis indicated (unsurprisingly) that papers focusing on specific risk-factors cited the relevant specific findings from Paul et al. more often than papers of other classifications. Thus, for example, all “caffeine” papers cited the coffee findings, and 84% of “alcohol” papers cited the alcohol findings. However, this relationship was much weaker in papers focusing on diet (38%) and serum cholesterol (35%). One exception to this is the papers on psychosocial factors; in these, the most frequently cited finding was the lack of association between job role and CHD, cited in only five papers. Most of these papers cited Paul et al. because it was the

Table 3. How papers of each title class cited the different classes of finding. The most frequently cited finding in each title classification group is highlighted in bold. Note that row totals can be higher than the total numbers of papers within a title classification group because each paper can reference multiple finding categories.

<table>
<thead>
<tr>
<th>Title classification, number of papers</th>
<th>Cited finding classification</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SC</td>
</tr>
<tr>
<td>Serum cholesterol (SC), n = 68</td>
<td>24</td>
</tr>
<tr>
<td>Blood pressure (BP), n = 32</td>
<td>8</td>
</tr>
<tr>
<td>Caffeine (C), n = 34</td>
<td>2</td>
</tr>
<tr>
<td>Diet (D), n = 66</td>
<td>14</td>
</tr>
<tr>
<td>Smoking (S), n = 16</td>
<td>1</td>
</tr>
<tr>
<td>Alcohol (A), n = 25</td>
<td>0</td>
</tr>
<tr>
<td>Physical activity (PA), n = 19</td>
<td>2</td>
</tr>
<tr>
<td>Body fatness (BF), n = 23</td>
<td>2</td>
</tr>
<tr>
<td>Psychosocial, n = 30</td>
<td>4</td>
</tr>
<tr>
<td>Unspecified CHD, n = 83</td>
<td>31</td>
</tr>
</tbody>
</table>

Figure 3. Distribution of research foci from analysis of 343 titles of papers citing Paul et al., showing only research foci terms appearing in more than 10 titles as either explanandum or explanans. See Leng (2021) for the code book containing full research foci description and supplementary data on distribution across other title classifications.
original cohort from which later studies that examined psychosocial variables were derived (e.g., Ostfeld, Lebovits et al., 1964).

4.5. Network Analysis

I then asked whether papers that share research foci title terms and findings tend to cite other papers that share these attributes. I constructed a directed citation network composed of all papers that had cited Paul et al. by the end of 1984, and the interactions between them (Figure 4).

To understand community structure, as revealed by interactions between citing papers, Paul et al. was then removed from the network, as this node expresses no important relational information. This left a large weakly connected component containing 399 papers (316 with full title and citation context data); 41 isolates (papers that do not cite any other citing papers); and two pairs of papers connected by a citation link, but unconnected to any other papers. On average, papers in this component connect to 3.7 other citing papers. I then applied the Leiden algorithm (Resolution 1; Iterations 100; Restarts 100; Random Seed 1); this resolved the network into nine clusters with $Q = 0.55$ (Figure 5). These clusters were then classified by terms in the titles of papers (Table 4). For clusters 2, 3, 5, 6, 7, and 8, this identified communities

Figure 4. Citation network representing citations between papers referencing Paul et al. ($n = 445$; $m = 1,912$). Layout: Lin-log ForceAtlas 2, Scaling 4.0. Paul et al. is the large black node (sized to draw attention to it), the blue nodes are papers with full title and finding classifications, the red nodes are papers removed from the analysis due to inaccessibility, and yellow nodes are papers removed from analysis due to being written in a language other than English.
Figure 5. Directed citation network showing citations between 399 papers citing Paul et al. \( n = 399; m = 1,466 \). Nodes are sized relative to their within-network citations (in-degree) and are colored by cluster membership as determined via modularity maximization \( (Q = 0.55; 9 \text{ clusters}) \). The edges between nodes are colored by the color of the source node (the citing paper). Cluster 0 = Blue; Cluster 1 = Light blue; Cluster 2 = Green; Cluster 3 = Red; Cluster 4 = Yellow; Cluster 5 = Pink; Cluster 6 = Orange; Cluster 7 = Black; Cluster 8 = Brown.

Table 4. Cluster number, number of citing papers and number of papers with full citation content data, highest percentage of papers with a specific title classification, and cluster classification. For papers in which less than 50% of full data papers had the same research foci classification, the top two terms were used to classify the cluster.

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Color</th>
<th>#Nodes/#full data papers</th>
<th>% of cluster papers with common title category</th>
<th>Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Blue</td>
<td>70/52</td>
<td>40% Unspecified CHD 37% Serum cholesterol</td>
<td>Serum cholesterol &amp; Unspecified CHD</td>
</tr>
<tr>
<td>1</td>
<td>Light blue</td>
<td>66/49</td>
<td>45% Unspecified CHD 22% Blood pressure</td>
<td>Blood pressure &amp; Unspecified CHD</td>
</tr>
<tr>
<td>2</td>
<td>Green</td>
<td>64/48</td>
<td>69% Caffeine</td>
<td>Caffeine</td>
</tr>
<tr>
<td>3</td>
<td>Red</td>
<td>53/47</td>
<td>77% Diet</td>
<td>Diet</td>
</tr>
<tr>
<td>4</td>
<td>Yellow</td>
<td>39/36</td>
<td>39% Unspecified CHD 28% Body fatness</td>
<td>Body fatness &amp; Unspecified CHD</td>
</tr>
<tr>
<td>5</td>
<td>Pink</td>
<td>34/28</td>
<td>68% Physical activity</td>
<td>Physical activity</td>
</tr>
<tr>
<td>6</td>
<td>Orange</td>
<td>32/27</td>
<td>93% Alcohol</td>
<td>Alcohol</td>
</tr>
<tr>
<td>7</td>
<td>Black</td>
<td>30/21</td>
<td>71% Psychosocial</td>
<td>Psychosocial</td>
</tr>
<tr>
<td>8</td>
<td>Brown</td>
<td>11/8</td>
<td>88% Smoking</td>
<td>Smoking</td>
</tr>
</tbody>
</table>
focusing on specific risk-factors, with between 68% and 93% of papers in a cluster sharing the same research foci classification. Clusters 0, 1, and 4 contained <50% of papers with any one risk-factor, and for these clusters the two most common research foci classification were used to classify these.

The same analysis as for Table 3 was repeated, but exchanging cluster classification for title classification (Table 5). All citations to these eight findings categories from all papers \( (n = 27) \) that were not contained in the large component were also included under “Detached nodes,” which allows us to obtain the total number of citing papers for each finding.

Clusters cited specific findings differentially. For example, cluster 8, consisting primarily of papers with smoking in the title, cited Paul et al. only for the positive relationship between cigarette smoking and CHD. In cluster 6, composed primarily of papers with alcohol in the title, 78% of papers cited the lack of a significant association between alcohol and CHD, and only one paper outside this cluster mentioned that finding. Thus Paul et al. was cited for different findings by different communities. Some findings circulated primarily in communities focused on a specific risk factor, but some other findings were cited by several communities.

To understand the interaction between papers in clusters and between clusters, the contingency matrix in Table 6 includes the total number of citations from one cluster to another. The papers of Cluster 2 direct \(~88\%\) of their within-network references to other papers in this cluster. By contrast, cluster 4 is the least inward-looking cluster, with papers in this cluster directing only \(~55\%\) of their references other papers in this cluster.

### Table 5

The rows show the number of full data papers by cluster classification. The columns classify papers by the cited finding class. The most frequently cited finding per cluster category is highlighted in bold. SC = Serum cholesterol; BP = Blood pressure; C = Caffeine; D = Diet; S = Smoking; A = Alcohol; PA = Physical activity; BF = Body fatness. Note that row totals can be higher than the total numbers of papers within a cluster because each paper can reference multiple finding categories.

| Cluster number, classification, and number of papers with citation context data | Cited finding classification |
|---|---|---|---|---|---|---|---|---|---|
| | SC | BP | C | D | S | A | PA | BF |
| C0 – Serum cholesterol & CHD, \( n = 52 \) | 34 | 18 | 1 | 6 | 11 | 0 | 6 | 9 |
| C1 – Blood pressure & CHD, \( n = 49 \) | 19 | 17 | 3 | 4 | 9 | 0 | 1 | 6 |
| C2 – Caffeine, \( n = 48 \) | 6 | 4 | 42 | 9 | 5 | 1 | 1 | 3 |
| C3 – Diet, \( n = 47 \) | 13 | 2 | 5 | 15 | 3 | 0 | 1 | 0 |
| C8 – Smoking, \( n = 8 \) | 0 | 0 | 0 | 0 | 8 | 0 | 0 | 0 |
| C6 – Alcohol, \( n = 27 \) | 0 | 0 | 2 | 0 | 0 | 21 | 0 | 0 |
| C5 – Physical activity, \( n = 28 \) | 3 | 4 | 0 | 0 | 3 | 0 | 15 | 2 |
| C4 – Body fatness & CHD, \( n = 36 \) | 3 | 5 | 0 | 2 | 0 | 0 | 1 | 7 |
| C7 – Psychosocial, \( n = 21 \) | 3 | 2 | 1 | 1 | 0 | 0 | 2 | 0 |
| Detached nodes, \( n = 27 \) | 4 | 5 | 0 | 4 | 4 | 0 | 1 | 2 |
| Total | 85 | 57 | 54 | 41 | 43 | 22 | 28 | 29 |
5. INTERPRETIVE ANALYSIS OF RESULTS

The preceding analyses show that different findings tended to be cited by communities of papers focusing on different risk factors. However, these quantitative analyses do not reveal how particular findings were cited. In what follows, I examine how five of the categories of findings were used: the dietary findings, and the statistically significant associations between CHD and serum cholesterol, blood pressure, cigarette smoking, and coffee consumption.

5.1. On Serum Cholesterol, Blood Pressure, and Smoking

Of the 343 papers for which citation content analysis was performed, 100 discussed Paul et al.’s findings on the relationships between CHD and serum cholesterol (n = 85) and blood pressure (n = 57), with 42 referencing both findings. Neither finding was novel; both factors had long been suspected to be important in the development of atherosclerosis and CHD (Anitschkow, 1933; Gubner & Ungerleider, 1949). The Framingham Study, a larger prospective cohort study, had already reported on the association between these factors and CHD in a longer follow-up study (Kannel et al., 1961).

Table 6. Citation interaction between 399 papers in nine different clusters established via modularity maximization via the Leiden algorithm. The Pearson residuals corresponding to the raw counts (Sharpe, 2015) are between 16.1 and 29.3 for all of the green cells and between −6.1 and 0.1 for all other cells.

<table>
<thead>
<tr>
<th>Cluster</th>
<th>C0</th>
<th>C1</th>
<th>C2</th>
<th>C3</th>
<th>C4</th>
<th>C5</th>
<th>C6</th>
<th>C7</th>
<th>C8</th>
<th>Total references</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson residual</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cluster 0 (C0)</td>
<td>179</td>
<td>29</td>
<td>2</td>
<td>18</td>
<td>19</td>
<td>15</td>
<td>0</td>
<td>10</td>
<td>1</td>
<td>273</td>
</tr>
<tr>
<td>Cluster 1 (C1)</td>
<td>29</td>
<td>120</td>
<td>7</td>
<td>15</td>
<td>2</td>
<td>14</td>
<td>0</td>
<td>9</td>
<td>0</td>
<td>196</td>
</tr>
<tr>
<td>Cluster 2 (C2)</td>
<td>7</td>
<td>6</td>
<td>212</td>
<td>9</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>241</td>
</tr>
<tr>
<td>Cluster 3 (C3)</td>
<td>35</td>
<td>12</td>
<td>28</td>
<td>142</td>
<td>12</td>
<td>3</td>
<td>7</td>
<td>1</td>
<td>0</td>
<td>240</td>
</tr>
<tr>
<td>Cluster 4 (C4)</td>
<td>29</td>
<td>19</td>
<td>3</td>
<td>5</td>
<td>86</td>
<td>7</td>
<td>3</td>
<td>4</td>
<td>0</td>
<td>156</td>
</tr>
<tr>
<td>Cluster 5 (C5)</td>
<td>18</td>
<td>6</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>68</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>99</td>
</tr>
<tr>
<td>Cluster 6 (C6)</td>
<td>1</td>
<td>2</td>
<td>7</td>
<td>4</td>
<td>9</td>
<td>10</td>
<td>135</td>
<td>2</td>
<td>0</td>
<td>170</td>
</tr>
<tr>
<td>Cluster 7 (C7)</td>
<td>9</td>
<td>5</td>
<td>0</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>56</td>
<td>0</td>
<td>75</td>
</tr>
<tr>
<td>Cluster 8 (C8)</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>12</td>
<td>16</td>
</tr>
<tr>
<td>Total citations</td>
<td>307</td>
<td>200</td>
<td>260</td>
<td>198</td>
<td>133</td>
<td>120</td>
<td>147</td>
<td>86</td>
<td>15</td>
<td>1466</td>
</tr>
</tbody>
</table>
The vast majority of papers citing Paul et al. for these findings appeared to be using them to reinforce the importance of these factors in CHD. For example, Strisower, Adamson, and Strisower (1968, p. 498) state: “The Framingham and other studies [17–20] have clearly established the strongly positive correlation between blood cholesterol concentration and the risk of newly developing ischemic heart disease.” Here, Paul et al. is reference 19, and it was common to find several studies cited together in support of the claim that elevated serum cholesterol was linked to CHD. A decade later, the claim was typically asserted more strongly; Truswell (1978, p. 977) in citing Paul et al. among other studies stated: “It is firmly established that plasma total cholesterol is positively related to the risk of subsequent CHD. This has been found in 21/21 prospective studies [4–24].”

No papers challenged these findings, at least in the text that surrounds the in-text citations. I detected only one major error of interpretation: Matsumoto (1970, p. 11) stated that “Dietary studies by Paul […] and others reveal no significant association of intake of dietary fat or of serum cholesterol level with the probability of developing coronary heart disease.” As Paul et al. did find a significant association between serum cholesterol and CHD, Matsumoto’s statement is wrong.

Paul et al.’s findings regarding smoking were cited in 43 papers. While the association between CHD and smoking had not been well studied by previous prospective studies, a combined analysis of the Albany and Framingham cohorts published the year before had found a significant association between smoking and myocardial infarction (Doyle, Dawber et al., 1962). These findings were, like the serum cholesterol and blood pressure findings, cited as demonstrations of an accepted “fact.” For example, Hass, Landerholm, and Hemmens (1966, p. 751) wrote: “Most population studies have led to the conclusion that there is an association between smoking and a high incidence of coronary artery disease in man.”

A total of 24 papers citing the smoking papers also cited either the serum cholesterol or blood pressure findings. Thus, in total, 119 papers cited Paul et al. for at least one of the findings of these three findings. They were cited as demonstrations of a relationship that was accepted as existing between these factors and CHD, and often alongside other prospective studies that had demonstrated these associations before or after the publication of Paul et al. However, this is in marked contrast to the way in which other frequently cited findings were used.

5.2. Diet and Heart Disease

Paul et al. was also cited in 41 papers that specifically discussed the lack of association between diet variables and CHD. One of these focused on dietary salt intake, which was hypothesized to increase CHD risk via increasing blood pressure. In 13 of these papers, the significant association between serum cholesterol and CHD was also discussed. As the diet–heart hypothesis was the leading explanation for why different populations had varying levels of serum cholesterol and why different countries experienced different rates of CHD, we might have expected any discussion of either finding to reflect on both of these findings. So how were these dietary findings used?

From an analysis of citation contexts, 23 citing papers specifically noted the lack of association between dietary fats and CHD, 15 highlighted the lack of association between diet and CHD without specifying nutrients, two cited the lack of association between dietary cholesterol and CHD, and one focused on salt intake and CHD.

The first paper to cite the lack of association between diet and CHD was Yudkin and Roddy (1964, p. 7). Yudkin was a long-time critic of the diet–heart hypothesis and had proposed a
rival hypothesis linking sugar intake with an increased risk of CHD. He interpreted the findings as follows:

Paul et al. [...] mention the lack of association between any of the dietary constituents that they considered, and the development of I.H.D. (CHD) in 88 of their subjects. They considered calories, protein, total carbohydrate, fat, animal fat, total saturated and total unsaturated fatty acids, individual unsaturated fatty acids, and a number of mineral elements and vitamins ... However they do not appear to have calculated the intake of sugar, or its relation to the development of disease.

The findings of Paul et al. thus appear to play two roles for Yudkin and Roddy: (a) an attempt to discredit the diet–heart hypothesis; (b) a call to pay greater attention to the role of sugar in CHD. Over the next few years, many papers discussing the lack of association between diet and CHD also discussed Paul et al.’s failure to test whether sugar intake was correlated with CHD. Many of these either mirrored Yudkin’s interpretation or simply highlighted the negative results. For example, Marquis and Oliver (1964, p. 874) berated the editor of the British Medical Journal in a letter that cited both the dietary findings and Yudkin’s paper:

It is interesting that, at this time, you can still include in your leading article of 19 September the recommendation that persons with ischaemic heart disease should reduce their intake of fat, and exclude any suggestion that they should take less sugar. There has never been any direct evidence for the hypothesis that fat consumption has anything to do with causing the disease; indeed, recent evidence points increasingly against it [...] Results have recently been published of a prospective study in American men, in which no difference in fat intake was found between those who developed ischaemic heart disease and those who did not.

However, Epstein (1965, p. 762), a dietary epidemiologist and early advocate of the diet–heart hypothesis, proposed a different interpretation of the evidence:

A series of investigations have suggested that serum cholesterol and dietary fat intake are not correlated in Western countries. This is well illustrated in the study by Paul and his associates who found no difference in coronary disease incidence among men in the highest and lowest quintile of fat intake; however, even the men in the lowest quintile consumed 36 per cent of their calories as fat and their average serum cholesterol level was as high as 247 mg/100 ml so that, as the authors indicate, minor alterations in the so-called ‘average’ American diet are not likely to achieve the desired goal.

For Epstein, Paul et al. failed to find an association because the population studied were all eating too much fat. This interpretation was mirrored by other advocates of the diet–hypothesis. In Keys, Aravanis et al. (1967, pp. 13–14) the dietary findings are discussed in a way that suggests that Epstein’s interpretation was endorsed. Reflecting on several large prospective studies, they state:

These American population samples are characterized by a relatively high degree of homogeneity in certain respects of mode of life so that they throw little light on the effect of differences in mode of life [...] Universals in the disease will be clear only when populations of differing habits and cultures are included in broader studies.
Other researchers voiced their uncertainty over what inferences could be made from the evidence (or thereof) linking diet and CHD, despite accepting the evidence linking serum cholesterol to CHD. For example, Ostrander, Francis et al. (1965, pp. 1195–96):

Much of the current therapy or prophylaxis of vascular disease depends upon efforts to lower serum cholesterol levels. Proponents of fat restriction support their position with the impressive statistical evidence of a direct relationship between the level of the serum cholesterol and the incidence of coronary heart disease in a number of other epidemiological studies [...] As yet, however, there is no proof that levels of cholesterol attainable with a tolerable diet prevent or delay the progress of atherosclerosis.

Other citing papers distorted the findings. For example, Shillingford claims Paul et al. was a controlled trial rather than an observational study:

There is an interesting study in progress in Chicago at the Hawthorne Works of the Western Electric Company by Paul and his associates [...]. Some 2,000 men at the factory between the ages of 40 and 55 have been followed for four years or more: 296 men were put on a high-fat diet and the same number on a low-fat. Fourteen on the high-fat developed coronary disease as against 16 on the low-fat diet. (Shillingford in Hughes, 1964, p. 365)

Parkes, Benjamin, and Fitzgerald (1969, p. 743), by rolling together multiple references to different studies that found different findings regarding the association of different risk factors and CHD to highlight the apparent importance of several risk factors, ended up referencing studies that contradicted some of those claims:

It is possible that emotional stress acts by altering the consumption of fats, sugar, coffee, or tobacco, all of which have been shown to be statistically related to the mortality from coronary artery disease [Hammond & Horn, 1958; Paul et al., 1963; Yudkin, 1967].

This citation is misleading as Paul et al. found no association between dietary fats and CHD. Other papers that advocated dietary change simply ignored the evidence. For example, Leren (1966), a randomized controlled trial that produced results supportive of the diet–heart hypothesis, references only the significant association between serum cholesterol and CHD in Paul et al., ignoring the dietary findings. Mayer (1967), who stated that there was “no doubt in my mind that profound changes in our diet, cutting down on saturated fats and increasing polyunsaturated fats, are long overdue” (p. 1030), quoted the significant association with serum cholesterol but ignored the inconvenient dietary findings:

It is probably unnecessary to review in any detail the evidence that a diet high in saturated fat is conducive to hypercholesterolemia and that, in turn, an increase in the level of serum cholesterol in a population is correlated with an increase in coronary disease [...] The long term 1963 study at the Western Electric Company at Chicago also showed a good correlation between serum cholesterol level and the development of coronary heart disease. (Mayer, 1967, p. 1029)

By the late 1970s, there remained major differences in how scientists were interpreting these findings. Some continued to suggest that the findings undermined the diet–heart hypothesis: thus, Armstrong, Mann et al. (1975, p. 467) stated that “[D]irect epidemiological studies have not supported the reported association between either sugar or saturated fat intake and
myocardial infarction […].” On the other hand, Yano et al. (1978, p. 1270), following Epstein’s interpretation, claimed that “Epidemiological studies within homogeneous Caucasian populations so far have consistently failed to show statistically significant associations between individual nutrient intakes and either the level of blood lipids or the risk of CHD within populations.” Others expressed their uncertainty:

Two studies which have been able to link individual dietary fat intake data to subsequent IHD incidence have both failed to establish any relationship [Paul et al., 1963, Kannel & Gordon, 1970]. All we are really entitled to conclude from these data (the best currently available) is that affluence and IHD are associated; we cannot conclude that dietary fat (or any other nutrient) is a direct, major cause of IHD – or that it is not such a cause; the case is ‘not proven’” (Meade, 1973, p. 645)

However, in 1981, a year follow-up of the Western Electric Studies cohort (Shekelle et al., 1981) re-analyzed data from Paul et al. (1963) on diet and serum cholesterol and examined how these were related to CHD mortality over the following 19 years. Paul et al. had measured diet and serum cholesterol levels in both the first and second years of the study, but had not examined the relationship between the variables. Shekelle et al. analyzed the first-year results for a relationship between serum cholesterol levels and the dietary intake of SFA, polyunsaturated fats (PUFA), and dietary cholesterol (as a percentage of caloric intake), finding significant positive associations with SFA intake and dietary cholesterol, but no significant association with PUFA intake. The authors then evaluated the mean dietary scores for the initial examination via equations that combined the intakes of SFA, PUFA, and dietary cholesterol and estimated the likely effects of a diet on serum cholesterol. By this, the authors reported a significant, positive association between diet scores and changes in serum cholesterol.

The authors then noted that in the second year of observation the mean intakes of SFA and dietary cholesterol were lower and the intakes of PUFA moderately higher, and the mean values for serum cholesterol were lower too for 1,900 men for which data were available. They then examined how serum cholesterol levels changed from the first- and second-year examinations and found that this was significantly associated with changes in intake of SFA and dietary cholesterol, and with calculated dietary scores.

Finally, they examined whether these dietary factors were related to the development of CHD via an analysis of hospital records. Splitting the cohort into three groups reflecting the bottom, middle, and top intakes of consumption of each variable, they reported that SFA intake was not significantly associated with CHD. However, dietary cholesterol was positively associated with CHD—10.9% died of CHD in the lowest tertile, 9.5% in the middle tertile, and 13.6% in the highest, while PUFA intake was reported to be negatively associated with CHD—13.5% in the lowest tertile died from CHD, 10.4% in the middle tertile, and 10.1% in the highest tertile. Finally, both dietary scores were reported to be positively associated with incidence of CHD.

Shekelle et al. (1981, p. 69) concluded: “[T]he present observations support the conclusion that the lipid composition of the diet affects the level of serum cholesterol and the long-term risk of death from CHD in middle-aged American men.” With this conclusion, any analysis of how the original study by Paul et al. was interpreted needed to take into account these findings too. In the few citing papers published after this that discussed both papers and the dietary fat and CHD findings, a difference of opinion was still evident:

One of the reasons for the controversy about the aetiological role of dietary fat has been the failure in population studies to show a relationship between an individual’s dietary
intake and his subsequent risk of coronary heart disease. Critics of the dietary fat theory have pointed to this as important negative evidence ... Among the negative studies was the Western Electric, begun in Chicago in 1957 [...]. Now, in a later report from that study, Shekelle and colleagues trace the mortality experience of the men in the original sample over 19 years and show a positive result [...] (Hetzel & Dwyer, 1981, p. 597)

But not all accepted the findings:

However, epidemiologic studies done within single countries usually show no association between dietary fat and heart disease risk [16–25, Paul et al., 1961] [...] Another large epidemiologic study [Shekelle et al., 1981] revealed an association between dietary lipids and serum cholesterol and between dietary lipids and risk. The very improbable association [...] was so small as to be considered ‘of no biological importance’ [...] and ‘not supporting’ the lipid hypothesis’ [50]. These conclusions were not those of the authors, however. (Klevay, 1983, p. 247)

5.3. Coffee and Heart Disease

The significant association between coffee consumption and CHD was cited in 54 papers. Unlike the previous findings discussed, no one expected a relationship between coffee consumption and CHD. Most citing authors simply stated that a study had found a significant association. Many of these studies were directly exploring whether coffee consumption could be found to be linked to CHD in their own studies and were looking for plausible mechanisms that might explain why coffee consumption could cause CHD. Thus Bellet, Kershbaum, and Aspe, (1965, p. 752) examined the effect of coffee consumption on serum free fatty acids, citing Paul et al. to contextualize their own findings:

A recent epidemiological study has shown a positive association between coffee intake and coronary heart disease. The FFA [free-fatty acids] effect after caffeine observed in the present study may be related to this clinical observation.

Similarly, Little, Shanoff et al. (1966, p. 732) cited Paul et al. (1963) to contextualize their own research:

We found no significant correlations between various nutrients and serum-cholesterol levels in healthy Canadians and patients with coronary heart-disease ... The single exception was a positive correlation between daily cups of coffee and serum-cholesterol in the coronary group ($r = 0.31; P < 0.01$). Paul et al. (1963) reported a correlation between coffee and the later development of coronary heart-disease ($P < 0.025$).

However, as highlighted previously, the discussion of these coffee findings has an interesting relationship to the dietary fat findings. Yudkin and Roddy (1964, p. 8), after discussing the dietary fat finding, write:

[Subjects who developed ischaemic heart-disease drank significantly more cups of coffee than the control subjects. Though they make no comment on this observation, it is likely that the amount of coffee drunk was an indication of a higher intake of sugar, just as we ourselves have found that our English patients with arterial disease drink more cups, mostly of tea.
This provoked a debate between those who believed coffee consumption was a true association and those who saw it as confounded by sugar intake. Little and Shanoff (1965, pp. 184–185) called for caution with regard to Yudkin’s interpretation of the Paul et al. study:

Such a study by Paul et al. showed that, except for coffee, there was no association between diet and later incidence of C.H.D. Unfortunately, sucrose as such was not specifically mentioned. Possibly their data could be reanalysed for dietary sucrose, and we are suggesting this to Dr. Paul […] [However] until results from prospective studies are available, it should not be accepted that C.H.D. patients ate more sucrose than normal subjects … The observation by Professor Yudkin that the consumption of sugar in our civilisation has greatly increased in recent time may be very significant [but] … replacing these speculative facts will require much more work.

In a follow-up study, Paul, MacMillan et al. (1968) tested both whether dietary sugar was related to CHD, and whether the relationship between coffee consumption and CHD persisted after controlling for cigarette smoking. The results were not what either those interested in the relationship between coffee or sugar and CHD wished for. Coffee consumption was strongly associated with smoking. The consumption of sugar was higher in those who developed CHD, but this association failed to reach statistical significance after controlling for smoking and coffee consumption. What was initially taken as a promising risk factor and topic of research was now very uncertain. Nine of the 33 papers that cited the coffee findings also cited this follow-up, and all did so to cast doubt on the relationship (e.g., Heyden, 1969). While others noted the discrepancy between the original findings of Paul et al. and other prospective studies that later tested this link. The last paper to cite this finding did so in the following manner:

Paul and associates […] in 1963 reported a positive association between coffee consumption and ischemic heart disease […] Paul’s group […] later showed that this apparent association between coffee consumption and ischemic heart disease could be accounted for by cigarette smoking. (Curatolo & Robertson, 1983, p. 646)

6. DISCUSSION

This study has shown that the findings of Paul et al. were cited in many different ways that appeared to reflect the diverse interests of particular communities. Paul et al. was, in part, a test of the diet–heart hypothesis that predicted associations between dietary SFA intake and (a) serum cholesterol levels and (b) CHD, and (c) an association between serum cholesterol and CHD. This hypothesis was key to the argument that the incidence of CHD in the population would be lowered if only people could be persuaded to eat less fat. However, while Paul et al. found an association between serum cholesterol and CHD, they found no association between dietary intake and CHD.

Of the 343 sample papers that cited Paul et al., most (~89%) cited at least one of its empirical findings—but most did so for just one of the many findings—most commonly, for the significant associations of serum cholesterol ($n = 85$), blood pressure ($n = 57$), and coffee consumption ($n = 54$) with CHD. But the lack of association between dietary intake and CHD was discussed in just 41 papers, and only 23 of these discussed the lack of association between dietary fats and CHD. The lack of association between diet and CHD was not mentioned in 72 of the 85 papers that cited the association between serum cholesterol and CHD. Of the 13 papers that did mention this, only five specifically mentioned the lack of association between
dietary fats and CHD, six discussed the lack of association between diet and CHD without specifying nutrients, and two mentioned the lack of association between dietary cholesterol intake and CHD only.

The dietary fat findings, at least initially, made the biggest impact not on those interested in the diet–heart hypothesis, but in a community advocating a rival hypothesis—the sugar–heart hypothesis. The most highly cited paper in this group (Yudkin & Roddy, 1964) took the dietary fat finding as undermining the diet–heart hypothesis, and the authors did so to argue for more research into the relationship between sugar consumption and CHD. This paper also interpreted the significant association with coffee consumption as a potentially supportive finding towards the sugar hypothesis because these authors believed it to be a confounded relationship that concealed the true culprit—not the coffee being drunk by the men in this population, but by the sugar that they likely added. Importantly, others offered very different explanations of these findings. For Epstein (1965), the lack of association between diet and CHD was further evidence that Americans generally consume too much fat, concealing the true relationship between dietary fat intake and CHD, while Bellet et al. (1965) and Little et al. (1966) seemed to see the caffeine finding as hypothesis-generating finding, and used the finding to help contextualize their own studies examining the relationship between caffeine and CHD.

Thus, scientists appeared to use Paul et al. to support their contentions and to justify their particular research agendas. We might understand these findings by an appeal to the constructivist view that scientific texts are not simply objective technical accounts of experiments or findings, but tools used to advance particular arguments and perspectives (Cozzens, 1989; Gilbert, 1977; Latour, 1986; Latour & Woolgar, 1979; Luukkonen, 1997). For Latour and Woolgar (1979), scientists are professional “writers and readers in the business of being convinced and convincing others” (p. 88). Scientists devote both their time in the laboratory and in writing their papers to converting particular claims into statements of higher or lower credibility. Scientists, through citation, support and attack statements contained in other scientific publications that help them to alter the level of credibility of those scientific assertions.

By this account, the act of referencing is primarily a rhetorical tool, to convince an audience of the validity and importance of an author’s claim. How else can we understand the emphasis placed on reporting the lack of nearly all dietary variables by Yudkin and Roddy (1964) when other papers devoted less time and emphasis to these results or simply ignored them? How else can we understand the differing interpretations of the meaning of the lack of association of dietary fat between these authors and Epstein (1965) or the different interpretations of the meaning of the caffeine association? Why would the serum cholesterol findings be reported but not the lack of association between diet and CHD? Why else would single empirical findings be cited rather than a more holistic evaluation of Paul et al.?

Sociologists have long been interested in how interpretative factors can lead to scientific disagreement (Collins, 1981; Jacomy, 2020; Latour, 1987). The findings of this study can be read as an example of what Collins (1981) termed “Interpretive Flexibility”—a term used to describe how different groups of scientists can interpret the meaning of particular findings in different, sometimes divergent, manners. Previously, I reported results that suggested that scientific disagreement may be sustained by different evidence selection behaviors (Leng, 2018; Leng, Leng, & Maclean, 2019), and shown that network analysis combined with citation context analysis via the claim-specific approach (Greenberg, 2009) is a powerful tool for detecting this. However, this study shows that such network analyses can also identify cases of divergent interpretations of the same empirical study.
In this study, the tendency to cite any given finding was related to both research foci terms in the titles and to how papers clustered together, suggesting that findings were selected for their relevance to a particular research question. Small and Greenlee (1980) first showed how co-citation analysis combined with citation context analysis could be used to trace the spread of particular results through the literature. Their approach has since been adopted to produce fine-grained maps of particular research areas with papers linked by co-citation links and classified by their “consensus passage”—the finding, concept, or method that a paper is most commonly cited for (Schneider, 2006; Small, 1986, 2011). Here, I have demonstrated a complementary method for tracing the spread of findings by constructing a directed citation network composed of the citing literature of a single paper, and identifying clusters within that network (Traag et al., 2019), for the purpose of studying the variation of use and interpretation.

With the development of digital archives that include citation context data, such as Semantic Scholar and Scite (Nicholson, Mordaunt et al., 2021), the kind of analysis performed in this paper should become much easier. Recently, Bornmann et al. (2020) introduced “Citation Concept Analysis” (CCA) for analyzing the cognitive influence of particular theoretical works. However, the authors note that “The method [CCA] cannot be used to measure the impact of conventional publications, those that do not introduce an important concept and that do not receive many citations.” (p. 1052). Although CCA cannot be used to measure uncited influence, it appears possible to extend CCA to trace the trajectory of empirical findings. If clustering of citation networks proves to be a reliable indicator of the tendency for papers within a cluster to cite the same findings, as it has proved for establishing papers focusing on specific topics (Klavans & Boyack, 2017; Leng & Leng, 2021; Shibata, Kajikawa et al., 2008), then network analysis in conjunction with CCA would offer a systematic method for tracing where findings and their interpretations reside in the literature. If so, network analysis could make an important addition for studies examining the differences that can arise in how particular studies are used and interpreted via citation context analysis (Anderson & Lemken, 2020).

### 6.1. High Citations and “Impact”

Very few of the 343 papers analyzed here addressed the totality of evidence in Paul et al.; most used single findings in a way that supported various specific claims regarding the causes of CHD. By current expectations of study rigor, Paul et al. was a woefully underpowered study. While the sample size of nearly 2,000 men was large for its time, it ultimately comprised an analysis of associations of many variables with an endpoint that affected only 88 men. Furthermore, Paul et al. did not report the findings from all of the variables that they had examined, but only a “portion” (1963, p. 20) of them, observations that were believed important because of the current theories of CHD. But it was already clear that CHD was a multifactorial disease, and the study, in retrospect, seemed likely to produce abundant false positive results if the analysis did not fully correct for multiple comparisons, or many false negative results where the effect size was small. As many of the findings reported by Paul et al. were generated from bivariate analysis, many of the observed associations might have been expected to disappear after controlling for other variables. Indeed, this was the fate of the reported association between coffee consumption and CHD; the original study authors published a paper five years later explaining that, when smoking was controlled for, the association between coffee intake and CHD became non-significant (Paul et al., 1968).

But the purpose of the present study was neither to praise Paul et al. nor to condemn it. This study was not about Paul et al., but about those who cited it. The bibliometric record marks Paul et al. as a study of exceptional impact. It was a major, original test of an important
hypothesis, and it had produced a wholly unexpected outcome. But it was not for this that Paul et al. was cited. Instead, the study was mainly cited for its various findings by different research communities focusing on different research questions. Yet, no single empirical finding was referred to in more than 25% of the citing literature; the high number of citations to Paul et al. is not explained by the importance of any one of its findings.

Thus, this is a paper that achieved highly cited status, but it seems to have made only a modest impact on the literatures that its findings came to be cited in. By concealing both the specific findings cited and their interpretations, citation counts, which are aggregates of different uses and interpretations (see reviews by Bornmann and Daniel (2008) and Tahamtan and Bornmann (2019)), can give a false impression of a paper having some specific influence on the literature. This fact might be obvious to those in science studies, but the assumption of uniformity of cited content has affected the literature on citation bias.

In most studies on citation bias (see Duyx et al., 2017), bias is detected by comparing how often papers that report statistically significant findings in favor of some hypothesis have been cited compared with papers reporting null findings or findings in the opposite direction to that predicted. However, it is rare that studies analyze exactly how papers are referenced via citation context analysis. The results of this study suggest that there is a problem with using citation counts to evaluate citation bias to studies with multiple findings. Such papers can be cited for one, many, or none of their findings (e.g., cited for methods), and those findings can be interpreted in different ways. Fewer than 25% of the citations to Paul et al. were directed at any one finding; thus its citation count alone tells us little about its impact on the literature, and comparing its citations to other studies of dietary fat consumption would be an inappropriate test of citation bias unless citation context analysis was performed for each study that it is compared to. Paul et al. is far from exceptional in the number of variables analyzed; for example, Medalie et al. (1973, p. 329), a later cohort study of dietary fat and CHD reported “over 100 single variables with the development of a first myocardial infarction.”

The use of citation counts alone may be less problematic when testing for citation bias towards a set of clinical trials if they are all trials of the same intervention (e.g., Jannot, Agoritsas et al., 2013; Miemer, Platts-Mills, & Jones, 2016; Ravnskov, 1992), which presumably limits what these studies can be cited for. However, caution is still needed. A paper may be referenced in support of a claim even when nothing in that paper could reasonably be taken as supportive of the claim. Errors can also be introduced into the interpretation of a finding, and that wrong interpretation can spread—leading to the strange situation where a paper is referenced for nothing actually stated in that paper, but for what is believed to be in it (Stang et al., 2018). As Small (1978, p. 338) stated:

[A] document could originally have been written to convey ideas other than the one it has come to symbolize. In other words, the process of becoming public property has transformed the document into something the author may not have intended. While such extreme instances are undoubtedly rare, and in most cases the intended message is the received message, the possibility of the social transformation of meaning must be recognised.

Of studies examining citation bias to observational studies via citation counts alone, only one has rigorously examined whether focal studies are reporting multiple determinants or outcomes and whether this has an impact on citation rates: Urlings, Duyx et al. (2019) found that studies of the relationship between the intake of trans-fats and CHD were cited significantly more often if they included additional determinants or more outcomes. de Vries, Roest et al. (2016) studied citation bias to studies examining whether an association exists between the
serotonin transporter gene and stress and depression, concluded that: “positive studies receive a disproportionate amount of attention”; however, they go on to highlight that, “where individual studies often include a variety of analyses and p values, it is difficult for any reader to tell the forest from the trees.” While it may well be the case that positive studies receive disproportionate attention, if a paper reports multiple determinants or outcomes it becomes more likely that some of these will be positive; whether such papers are cited for the positive outcomes or for other outcomes perhaps needs to be examined, not assumed.”

Thus, some variant of citation context analysis, such as the claim-specific approach (Greenberg, 2009), seems to be required if the intention of analysis is to understand the impact (or lack thereof) of a paper’s findings on the literature.

ACKNOWLEDGMENTS
I would like to thank Dr Niki Vermeulen, Professor Steve Sturdy, Dr Valeria Skaifida, Professor Steve Yearley, Professor Yves Gingras, and Professor Gareth Leng for providing me with feedback at various stages of this research. I also thank the three anonymous referees for their careful and constructive reviews of my manuscript.

COMPETING INTERESTS
The author has no competing interests.

FUNDING INFORMATION
This work was supported by the Economic and Social Research Council (ESRC) doctoral studentship, the ESRC’s Advanced Quantitative Methods (AQM) enhanced stipend, and the ESRC Postdoctoral Fellowship (Grant number: ES/V01790/1). The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

DATA AVAILABILITY
All data analyzed in this manuscript has been made openly accessible via Zenodo (Leng, 2021).

REFERENCES
Diversity in citations to a single study


De Vries, Y., Roest, A., Franzen, M., Munafò, M., & Bastiaanssen, J. (2016). Citation bias and selective focus on positive findings in the literature on the serotonin transporter gene (5-HTTLPR), life stress and depression. Psychological Medicine, 46(14), 2971–2979. https://doi.org/10.1017/S0033291716000080


Quantitative Science Studies 27
Diversity in citations to a single study


Misemer, B. S., Platts-Mills, T. F., & Jones, C. W. (2016). Citation bias favoring positive clinical trials of thrombolytics for acute
Diversity in citations to a single study

ischemic stroke: A cross-sectional analysis. 


Price, D. J. (1965). Networks of scientific papers. 


Schneider, J., Ye, D., Hill, A. M., & Whitehorn, A. S. (2020). Continued post-retraction citation of a fraudulent clinical trial report, 11 years after it was retracted for falsifying data. 


Serra-Garcia, M., & Gneezy, U. (2021). Nonreplicable publications are cited more than replicable ones. 


Small, H. & Greenlee, E. (1980). Citation context analysis of a co-citation cluster: Recombinant-DNA. 


https://doi.org/10.1093/oxfordjournals.aje.a112338, PubMed: 998607

Quantitative Science Studies


