Causality, menopause, and depression: a critical review of the literature

Louise Nicol-Smith

Section Editor
by Hume

Abstract

Objective—To assess whether causal criteria can be used to find out whether there is support in published research for maintaining that menopause causes depression.

Design—Ninety four articles from 30 years of research examining the relation of natural menopause to depression were traced by using Medline and systematic follow up of reference lists. Specified exclusion and inclusion criteria were applied, and the resulting 43 epidemiological primary research articles were classified and tabulated according to sample and measures used and the researchers' own conclusion as to whether or not an association had been established. This material was qualitatively evaluated with Hill's nine criteria for causality.

Result—There is insufficient evidence at present to maintain that menopause causes depression. In addition to methodological and statistical problems, a temporal problem in the menopause concept hinders research in this area.

Conclusion—Causal criteria can usefully be used to structure a literature review. Further theoretical work is required to integrate standard clinical epidemiological concepts.

Introduction

Decisions about treatment are guided by a clinician's understanding of the causality of the complaint. When a middle aged woman presents with symptoms of depression the clinician needs to know whether the hormonal fluctuations that are known to occur during that stage of life are causing or contributing to the clinical picture. In this paper the contribution of research towards this causal understanding is dealt with directly. The treacherous philosophical issues surrounding the nature of causality are well known and perhaps most succinctly stated by David Hume in the 18th century.¹ Faced with decisions regarding public health and calls for legal and financial accountability, however, some epidemiologists are now arguing that a more realistic and pragmatic stance to causality is necessary.^{2,3} This trend is reflected in the method used in this study to examine whether there is support in the literature for maintaining that menopause causes depression.

Materials and methods

DATA SEARCHES

A database search for relevant literature was carried out with Medline. By using the categories available I could not distinguish between articles that considered affective symptomatology in general and those that were primarily concerned with the clinical disorder of depression. No clear distinction between these two occurs either in the literature or the databases. I considered all English language articles registered between 1966 and May 1996 with the focus of article or

textword on the exploded categories "emotions," "affective symptoms," "mental disorders," and "climacteric." A wide variety of articles from different academic disciplines emerged. Reference lists were used to locate further literature. I made every effort to maintain a neutral stance when locating literature, however, knowing that there is selection bias in the publication process⁴ and that authors have a tendency to cite articles that support their own conclusions,⁵ an element of selection bias cannot be ruled out. Of the 43 relevant primary research articles located,⁶⁻⁴⁸ 22 were located through the database. Though this figure is low (51%), it compares favourably with previously reported figures for searches of this kind of electronic database.⁴⁹⁻⁵⁰ Twenty four review articles⁵¹⁻⁷⁴ emerged during the search, and in 10 of these the authors' conclusion was that depression was or could be related to menopause. Only one of these, however, specified how primary articles had been selected,⁵² and none discussed the validity of, or integrated systematically, the findings. These were therefore dismissed as informative but inconclusive.⁴⁹⁻⁷⁵ Figure 1 lists the primary research articles that constitute the material for this paper.

METHOD

When I evaluated the material my null hypothesis was that menopause does *not* cause depression. This viewpoint comes from the belief that it is a greater error to assume that menopause is causing depression when it is not than to assume it is not causing depression when it is. This could result in tried and tested methods for treating depression being overlooked in the false belief that the menopause was underlying the complaint. Therefore, when I examined the articles I searched for evidence that the menopause could cause depression, which would falsify the null hypothesis. Research that showed no association between menopause and depression was not considered further. Hill's nine criteria⁷⁷ have provided a useful framework for deciding whether the most likely interpretation of an observed association is causality.⁷⁸⁻⁸⁰ I used these criteria to evaluate the 17 studies in which an association had been found.

EVALUATION WITH HILL'S CRITERIA

Experiment—In this type of inquiry an experiment which would provide the strongest evidence of a causal association is ruled out. Random allocation to groups, exposure of one group to the risk agent (menopause), and then following the groups for the occurrence of the outcome (depression) is impossible. All the studies located used a non-experimental design.

Strength—Causal inferences are more easily made about strong associations. Weak associations, on the other hand, are much more likely to be the result of unsuspected biases. While randomisation and control of exposure are not available as a method of controlling for confounding or other biases for the non-experimental researcher, other study designs can go some way

Department of
Psychology, University of
Oslo, Oslo, Norway
Louise Nicol-Smith,
psychologist

Correspondence to:
Fururabben 19, 1345
Østerås, Norway.

BMJ 1996;313:1229-32

towards preventing faulty causal inferences. According to Rothman, the strength of different types of design is primarily related to selection of subjects.⁸¹ In a cohort design, subjects are free from disease at the onset of the study and are allocated to groups according to exposure. These groups are then followed forward in

time until the development of disease. In case-control studies subjects are allocated according to disease, and their past experience is examined for possible risk factors. Although several of the studies located used non-diseased control groups or followed a group of subjects over time, none could be described as using

* Final menstruation occurs naturally at about 51 years of age (USA).¹²⁹

† Medline textword.

‡ Medline focus of article.

§ Monthly postcard follow up with perimenopausal and postmenopausal examination after three and 12 months of amenorrhoea.

GHQ: Goldberg's general health questionnaire

GPI: Goldberg's clinical interview schedule

PSE: Wing's present state examination

Fig 1—Details of primary research articles in critical review. These are listed alphabetically according to year of publication. Several articles from the same research team are bracketed together under the most recent publication. Note: Sample sizes are those at the onset of the study; the numbers available for data analysis may be much smaller. Sample sizes may also have been inflated because of heterogeneity—some women could be taking either hormone replacement therapy or have had a surgical menopause. However, this was considered by the authors not to affect their original conclusion, which in every case related to natural menopause. In the repeated measures studies measurements at different time points may have been regarded as being independent of each other: one subject being counted several times as she passed through several menopausal stages.

[] Definite time boundary

⋮ Indefinite time boundary

◊ Group limit outside table boundary

⋮ Repeated measure with indication of time interval

■ Group claimed by the author to be more depressed than other groups in the study

← Time elapsed after final menstruation before being allocated to group following arrow

these definitions of study design. I contend that all the research listed in figure 1 is essentially cross sectional in design or "the contemporaneous classification of people with respect to both exposure and disease" even when subjects were reassessed at several time points (p 70 of Rothman⁸¹). Rothman's definitions of study design are based on an assumption of a disease process which occurs over time and as a result of a number of necessary and sufficient causes. Current information, he says, is generally too recent to be meaningful aetiologically. Study design has not, therefore, added weight to the findings of any of the individual studies. This, of course, does not preclude the fact that some of the studies have better internal or external validity as a result of various design features.

Consistency—Consistency refers to the repeated observation of an association in different populations under different circumstances. This can be indicative of causality. The right hand side of figure 1 shows the authors' allocation of subjects to groups according to menopausal status or age, or both. Groups which are hatched are claimed by the author(s) to be more depressed than other groups in the study based on *their analysis* of the data collected. Visual examination of these hatched areas can be interpreted as indicating that there is possibly a slight tendency towards increased symptomatology in the period leading up to the final menstruation. It can also be seen, however, that there is little if any consistency between the samples, measurements, or findings of the studies that claimed to find an association.

Specificity—This criterion relates to a specific cause producing a specific effect. If this can be demonstrated it is strongly indicative of causality. None of the psychological effects attributed to the menopause occur exclusively in mid-life so specificity cannot be established.

Temporality—Temporality relates to the necessity of a cause being seen to precede its effect. In figure 1 the time scale around the area marking the final menstruation is unclear. This is deliberate and reflects the researchers' difficulty in precisely defining menopausal status at this time. This clearly affects the possibility of being able to draw temporal conclusions. This problem is further discussed below.

Biological plausibility and gradient—The very existence of psychoneuroendocrinology testifies to the theoretical plausibility of hormones modifying affect. I found several studies on humans and animals which had used biological markers for menopause or depression.⁹⁴⁻¹²⁰ In this context, however, the incompleteness of the hypothetical models, the use of small clinical samples, and lack of replication prevent the drawing of causal conclusions from this material, at present.

Coherence and analogy—These last and, according to Hill, least important criteria are understood to relate to the logical relation between the chosen subject of interest and other specialties where a causal link is already established. If a causal link between the menopause and depression was demonstrated would this conflict with existing knowledge and be incoherent? Are other already established causal links directly analogous? The answer to the first question is clearly no, and conventional wisdom would merely be substantiated. The answer to the second question is more complex. Several topics of investigation can be considered to be analogous through a common hormonal factor. Mood changes after hormone replacement therapy and surgical removal of the ovaries have been reported, and, while involutional melancholia and the premenstrual syndrome are no longer included in the major disease classification systems, disorders associated with the puerperium are included in both the *Diagnostic and Statistical Manual of Mental Disorders*, fourth edition (DSM-IV)¹²¹ and ICD-10 (international classification

of diseases, 10th revision).¹²² It is argued that strong evidence from these related topics might predispose to accepting that hormonal disturbance during the menopause might also be causing depression, but more direct evidence is also required before causality can be inferred.

Results

On the basis of the foregoing arguments I therefore conclude that there is insufficient evidence to discard the null hypothesis. At present there is no substantial evidence that either a natural menopause with its accompanying changes in hormone concentrations or psychosocial factors exclusive to middle age put women at increased risk of depression. This conclusion has implications for both clinicians and researchers. For clinicians, this suggests that at present women suffering from depression in middle age should not be treated differently from those attending at other ages. The related issue of whether hormone replacement therapy has a beneficial effect on mood has not been dealt with in this paper and requires a further review with methodology appropriate for the assessment of controlled clinical drug trials. For researchers, however, the questions still remain.

Discussion

Does menopause cause depression and, indeed, can depression cause menopause? This latter question has been examined by at least one researcher.¹²³ And where does quality of life and wellbeing fit in? There is an increasing interest in this area which I did not consider for this review.¹²⁴⁻¹²⁶ With the increasing use of hormone replacement therapy where can we find samples of naturally menopausal women and how representative are women in the Western world who are possibly attending menopause clinics?

In preparing this paper, I examined the individual studies in depth. The problems confronting researchers are many: from how to define and measure their concepts, through methodology, to statistics. A detailed discussion of all these points is inappropriate here, and the interested reader is directed particularly towards articles on methodology by the researchers themselves.¹²⁷⁻¹³³ Two issues emerged, however, when I was carrying out this systematic review which do not seem to have been dealt with elsewhere.

CONCEPTUAL CONSIDERATIONS

Firstly, the measuring of the menopause seems to be largely dependent on tradition. The generally accepted research definition of the menopause as the final menstruation is unhelpful when it comes to constructing categories which can be used in data analysis. Summarising the researchers' categories of menopausal status was problematic when I was constructing figure 1 and is clearly shown in the disjunction in the time scale. While the probability of being menopausal increases with the duration of amenorrhea and age, this can be determined only retrospectively. This would seem to be at odds with a basic axiom of scientific methodology. If there is no objective way of detecting when a phenomenon is either present or not present (thereby creating at least two categories) then it cannot be quantified. By using this argument, menopause, despite its long research history and lay acceptance, can seem to be just such a phenomenon, and the construction of a variable based on arbitrary lengths of time from last menstrual period can be regarded as a method of avoiding this basic conceptual issue. A parallel, which may help to illustrate this point, would be to try to relate misbehaviour in 14 year old boys to a "teenage" variable based on time since the voice broke. When we are attempting to infer causality it may be more fruitful to

examine directly the hypothetical underlying biological processes or look for sociopsychological factors specific to that age group. Speculation about the social implications and reasons for having groups of people defined as teenage or menopausal is interesting but beyond the scope of this empirical review.

STATISTICAL CONSIDERATIONS

Secondly, little attention has been drawn to the difficulties of carrying out other than descriptive analysis of data with this degree of complexity. When design control cannot be carried out for confounding variables, multivariate statistical control becomes imperative. Hot flushes, stressful life events, socioeconomic factors, previous depression, marital status, expectations regarding the menopause, age, and race are just some of the other variables which were also claimed to be associated with depression in these studies. In addition, while it has been argued in this review that collecting data at several points in time does not necessarily improve study design, longitudinal data clearly has advantages regarding faulty subject recall, a pervasive problem for researchers in this specialty which was first pointed out in 1933.⁴⁸

The task of carrying out multivariate, repeated measure data analysis with an idiosyncratic definition of menopausal status can, however, easily become overwhelming. I encountered a wide variety of creative statistical manoeuvres in reading through this literature. Collaboration between researchers to replicate results in comparable datasets would lend strength to their findings and help to identify strengths and weaknesses in the analytic techniques used.

CONCLUSION

I hope that the usefulness of conducting a systematic literature review from an epidemiological viewpoint has been shown. It was difficult, however, to incorporate elementary epidemiological concepts regarding study design into Hill's criteria. In addition, the value of using Hill's criteria to establish causality is far from being generally accepted.^{3 81 134} While David Hume is renowned for his scepticism regarding whether causality could ever be established objectively, he was also, perhaps surprisingly, the first to demonstrate the day to day usefulness of adopting causal criteria. He proposed eight rules which he could use in his own logical reasoning. In addition to contiguity in space and time

(association), priority (temporality), and constant union (consistency), he also refers to specificity, gradient, analogy, and the complex interrelation between causes which, he states, need not necessarily be unidirectional or even linear in regard to their effect (listed in part III, section IV of Hume¹). There seems to have been remarkably little refinement of these rules since 1739, and it is respectfully suggested that this is a topic worthy of theoretical attention.

A preliminary version of this paper was submitted in 1993 as part of the requirements for the degree of Cand Psychol. I am indebted to Professor Arne Holte for initiating this review; Professor Peter Laake, section for medical statistics, for his teaching and encouragement in preparing this final paper; Kjell Mathiesen for technical and editorial assistance; and the librarians at the faculty of medicine for their good humour in helping to locate the literature.

Funding: This study was partly supported by funds from the Norwegian Climacteric Project.

Conflict of interest: None.

- 1 Hume D. *A treatise of human nature. Book 1: Of the understanding*. Harmondsworth: Penguin, 1984. (First published 1739.)
- 2 Renton A. Epidemiology and causation: a realist view. *J Epidemiol Community Health* 1994;48:79-85.
- 3 Susser M. What is a cause and how do we know one? A grammar for pragmatic epidemiology. *Am J Epidemiol* 1991;133:635-48.
- 4 Altman DG. The scandal of poor medical research (editorial). *BMJ* 1994;308:283-4.
- 5 Gøtzsche PC. Reference bias in reports of drug trials. *BMJ* 1987;295:654-6.
- Full details of references 6-48, the primary research articles in figure 1 are available from the author or on the Internet at <http://www.bmj.com/bmj/>
- 49 Oxman AD, Guyatt GH. Guidelines for reading literature reviews. *Can Med Assoc J* 1988;138:697-703.
- 50 Chalmers I, Dickerson K, Chalmers TC. Getting to grips with Archie Cochrane's agenda. *BMJ* 1992;305:786-8.
- Full details of references 51-74, the annotated references for review articles, are available from the author or on the Internet at <http://www.bmj.com/bmj/>
- 75 Muir Gray JA. Two classes of creativity-improving systematic reviews. *J Epidemiol Community Health* 1994;48:4-5.
- 76 Mulrow CD. The medical review article: state of the science. *Ann Intern Med* 1987;106:485-8.
- 77 Hill AB. The environment and disease: association or causality? *Proc R Soc Med* 1965;58:295-300.
- 78 Sackett DL, Haynes RB, Guyatt GH, Tugwell P. *Clinical epidemiology—a basic science for clinical medicine*. Boston: Little, Brown, 1985.
- 79 Holland PW. Statistics and causal inference. *J Am Statist Assoc* 1986;81:945-60.
- 80 Peterson HB, Kleinbaum DG. Interpreting the literature in obstetrics and gynecology. 1. Key concepts in epidemiology and biostatistics. *Obstet Gynecol* 1991;78:710-7.
- 81 Rothman KJ. *Modern epidemiology*. Boston: Little, Brown, 1986.
- Full details of references 82-93, the depression instruments listed in figure 1, are available from the author or on the Internet at <http://www.bmj.com/bmj/>
- Full details of references 94-120, the neurobiological articles, are available from the author or on the Internet at <http://www.bmj.com/bmj/>
- 121 American Psychiatric Association. *Diagnostic and statistical manual of mental disorders: DSM-IV*. Washington, DC: American Psychiatric Association, 1994.
- 122 *The ICD-10 classification of mental and behavioural disorders: clinical descriptions and diagnostic guidelines*. Geneva: World Health Organisation 1992.
- 123 Harlow LH, Cramer DW, Annis KM. Association of medically treated depression and age at menopause. *Am J Epidemiol* 1995;141:1170-6.
- 124 Pocock SJ. A perspective on the role of quality-of-life assessment in clinical trials. *Cont Clin Trials* 1991;12:257-65S.
- 125 Daly E, Gray A, Barlow D, McPherson K, Roche M, Vessey M. Measuring the impact of menopausal symptoms on quality-of-life. *BMJ* 1993;307:836-40.
- 126 Oldenhave A, Jaszmann LJ, Haspels AA, Everaerd WT. Impact of climacteric on well-being. *Am J Obstet Gynecol* 1993;168:772-80.
- 127 Matthews KA, Kuller LH, Wing RR, Meilahn EN. Biobehavioral aspects of menopause: lessons from the healthy women study. *Exp Gerontol* 1994;29:337-42.
- 128 Kaufert PA. A health and social profile of the menopausal woman. *Exp Gerontol* 1994;29:343-50.
- 129 McKinlay SM. Issues in design, measurement, and analysis for menopause research. *Exp Gerontol* 1994;29:479-93.
- 130 Kuller LH, Meilahn EN, Cauley JA, Gutai JP, Matthews KA. Epidemiologic studies of menopause: changes in risk factors and disease. *Exp Gerontol* 1994;29:495-509.
- 131 Brambilla DJ, McKinlay SM, Johannes CB. Defining the perimenopause for application in epidemiological investigations. *Am J Epidemiol* 1994;140:1091-5.
- 132 Kaufert PA, Gilbert P, Hassard T. Researching the symptoms of menopause: an exercise in methodology. *Maturitas* 1988;10:117-31.
- 133 Kaufert P, Syrotuik J. Symptom reporting at the menopause. *Soc Sci Med* 1981;15E:173-84.
- 134 Weed DL. On the logic of causal inference. *Am J Epidemiol* 1986;123:965-79.

(Accepted 10 September 1996)

Key messages

- Over the past 30 years, at least 94 studies published in the English language have investigated the relation between menopause and depression
- Causal and methodological criteria can be used to structure the findings and draw conclusions in a literature review
- Menopause has so far not been shown to cause depression
- In addition to methodological and statistical constraints, a temporal problem related to the menopause concept hinders research in this area
- Theoretical work is required to integrate causal criteria and standard clinical epidemiological concepts