Women, work, and well-being 1950–2000: a review and methodological critique

Petra L. Klumb*, Thomas Lampert

Berlin University of Technology, Ernst-Reuter-Platz 1, H 8, TU Berlin 10587, Germany

Abstract

In this research synthesis, we summarize 161 measures of the effects of women’s employment on well being reported between 1950 and 2000. Variations in the conceptualization and measurement of employment and health outcomes and the difficulty in distinguishing social selection from social causation limit the inferences that can be drawn from the evidence. Therefore, we distinguish two types of studies. Longitudinal studies measuring relevant covariates at the first measurement occasion and statistically controlling them in multivariate analyses providing effect-size information are classified as Type II studies. The remaining studies are classified as Type I studies. The main findings were that (1) results from methodologically sound Type II studies confirm the cross-sectional finding that paid employment has no adverse effects on women; (2) the outcome groups psychological distress, subjective health, cardiovascular risks and disease, and mortality do not converge completely.

Keywords: Germany; Review; Women’s health; Gender; Employment; Well-being

Introduction

In many industrialized countries, such as the US, the participation of women in the labor force increased markedly across the 20th century (Bureau of Labor Statistics, 2000). In other countries, such as Germany, the participation rate of women remained fairly constant across the 20th century, but women from younger cohorts became more likely to work full-time and to return to the labor market between or after childbirth (Maas, Borchelt, & Mayer, 1999). These historical changes provided a unique opportunity for researchers to investigate the effects of multiple roles on women’s well being. Consequently, there was a growing interest in the impact of (the additional role of) paid employment after World War II. Despite numerous efforts, however, the understanding of the underlying processes is still rather limited. This review attempts to summarize 50 years of research on the impact of employment on women’s physical and mental health. Because of their importance in evaluating the evidence, issues of conceptualization, measurement, and study design will be discussed first.

Conceptual and methodological issues

Two competing models guide research on the effects of multiple roles: the stress hypothesis (also known as role strain or scarcity; Goode, 1960) and the enhancement hypothesis (also known as role expansion or accumulation; Marks, 1977; Sieber, 1974). Because social roles are associated with daily demands and hassles as well as with major life events, the stress hypothesis holds that an additional role reduces an individual’s well being. At the same time, role involvement provides access to resources that may be instrumental in coping with demands. According to the enhancement hypothesis, therefore, individuals with an additional role should be in better health than individuals without this role—despite the additional demands and stressors it entails (Gove & Tudor, 1973; Kessler & McRae, 1982; Ross, Mirowsky, & Huber, 1983a). Other models encompass further properties of the living situation, individual expectations, and societal norms (job-stress model, person-environment fit model (P-E
fit), role quality model, gender model). However, the majority of researchers has based their investigations on role theory and have chosen the variables of interest accordingly.

**Choice of predictors**

Researchers usually operationalize the exposure or treatment variable as employment status. Women who are employed at the time of the study (sometimes including those looking for a job) are compared to those who are not in paid employment. As many researchers have pointed out, this dichotomy lacks reliability, i.e., it is temporally unstable, and may lack validity, i.e., it does not capture interindividual differences in labor market participation very well (Pavalko & Smith, 1999; Repetti, Matthews, & Waldron, 1989). The comparison category may be made up of such different groups as permanent homemakers, women on maternity leave, women who are—temporarily or permanently—ill or disabled, unemployed women looking for jobs, and students. Because of the fragmentation of female labor market participation, characterized by moves in and out of the labor force as well as by changes of occupation or career paths, assessing employment status at a single point in time seems to be a suboptimal way of capturing women’s work experiences (but see Weatherall, Joshi, & Macran, 1994, p. 293, who reported no substantial gain “by knowing whether a worker in 1981 was also in that category in 1971”). Therefore, some researchers draw on time-use information (daily, weekly, study-period, or life-time level) to quantify role involvement (Adelmann, Antonucci, Crohan, & Coleman, 1990; Bird & Fremont, 1991; Hibbard & Pope, 1991; Kotler & Wingard, 1989; Moen, Dempster-Clay, & Williams, 1992; Pavalko & Smith, 1999). Alternatively, the number of enacted roles may be used as a predictor. This predictor has been represented by different kinds and numbers of categories (e.g., Pietromonaco, Manis, & Frohardt-Lane, 1986, distinguished 5 roles; Reitzes & Mutran, 1994, distinguished 12 roles). Some perspectives focus on subjective experiences and expectations. For example, Barnett and Baruch’s concept of role quality is operationalized via the balance of rewards and concerns afforded by a role (Barnett & Baruch, 1985), while in the P-E fit approach, expectations and preferences are compared to actual role involvement (Waldron & Herold, 1986; Wethington & Kessler, 1989). Objective working conditions are rarely considered in this area (but see Hall, 1992).

**Choice of outcome measures**

Different research traditions tend to prefer specific outcome measures for conceptual, methodological, and historical reasons. In the initial studies on the effects of women’s employment, measures of psychological distress were used as dependent variables (see Warr & Parry, 1982b, for a classification). This was a result of the attention attracted by the sex differential in depression, which was ascribed to differences in the rewards provided by the traditional sex roles (Radloff, 1975; Rosenfield, 1980; Ross et al., 1983a). Large surveys usually include ratings of respondents’ general physical health or complaints (Bartley, Sacker, Firth, & Fitzpatrick, 1999; Waldron & Herold, 1986). Researchers in the tradition of the job-stress model usually prefer biological outcome variables such as neuroendocrine changes or cardiovascular risk factors; in large surveys, however, they have to rely on self-reports (Emslie, Hunt, & Macintyre, 1999; Lennon & Rosenfield, 1992). Researchers with a background in demography or social epidemiology have related employment to prospectively collected mortality information (Martikainen, 1995). Convergence or dissociations within or between these groups of measures have rarely been established in primary research (but see, e.g., Bekker, de Jong, Zijlstra, & van Landeghem, 2000). Nevertheless, a comparison of different outcomes and their time courses is of theoretical as well as practical importance.

**Potential moderators**

Rather than asking whether employment is adverse or beneficial per se (stress vs. enhancement), the question seems to be under which conditions it becomes adverse or beneficial. The same roles can have different effects depending on the prevailing conditions or the context (Waldron, Weiss, & Hughes, 1998). The role-substitution hypothesis suggests that roles yielding similar resources may have a substitutive relationship, such as that of employment and marriage. Taking on an additional substitutive role will be of little additional benefit to health. The role-complementation hypothesis suggests that specific role combinations, such as that of parenthood and marriage, have complementary or synergistic effects. In these combinations, the benefit from one role is enhanced by the other.

Beyond the influences of employment, marital, and parental status, exposure to stressors and the availability of resources vary as a function of age, occupation, and/or occupational status (Moen & Yu, 1999). Conflicts between roles change over the life course, as do the underlying demands and resources. Middle-aged women, blue-collar workers and those at the lower end of the status hierarchy are thought to be most exposed to stressors while having few resources available (Arber, Gilbert, & Dale, 1985). Consequently, it may be that only middle-class women experience a positive relationship between employment and well being. Alternatively, working-class women may benefit more from...
employment because their non-occupational environment is less favorable than that of middle-class women (Warr & Parry, 1982b).

**Confounded measures**

If all the variables of interest are assessed via self-report, the association between predictors and criteria may be based in part on shared method variance or other biases (Watson & Pennebaker, 1989). Partialling the prior from the subsequent measure of an outcome variable has been proposed as a way of controlling for the effects of stable third variables such as negative affectivity (Spector, Zapf, Chen, & Frese, 2000). Another way of controlling biases is to use multiple measures of predictors or outcomes.

**Causation vs. selection**

Since the seminal studies by Jahoda, Lazarsfeld, & Zeisel (1975), most researchers have seemed to assume a causal link between employment and well-being. However, the statistical associations reported in the literature (mostly based on posttest-only designs with unequalvalent groups) do not allow causal inferences to be drawn because alternative explanations cannot be ruled out (Cook & Campbell, 1979). Prior interindividual differences in health may lead to differential acquisition or relinquishment of roles. This phenomenon is known as reverse causation or the healthy-worker effect (Kandel & Davies, 1986; Waldron, Herold, & Dunn, 1982a; Waldron, Herold, Dunn, & Staum, 1982b). Employment status and health may also be the result of indirect selective processes, the antecedents of which (e.g., education) may or may not have been measured. This phenomenon is referred to as gravitation or drift. Most likely, there is a reciprocal relationship, i.e., selective and causative processes operate simultaneously and interactively (Waldron et al., 1982b).

The impact of employment has been investigated with many different study designs and analytical strategies. A small number of studies have employed macro-level data, i.e., data with some geopolitical entity as the unit of analysis (Pampel & Zimmer, 1989). The usual design, however, is an observational study based on individuals. Observational studies may be cross-sectional or longitudinal; most take a cross-sectional approach.

Most authors seem to implicitly assume that longitudinal studies would confirm their correlational findings. Some authors have been able to demonstrate that beneficial effects persist above and beyond selection (Verbrugge, 1983, 1986; Wethington & Kessler, 1989) whereas others have not (Jennings, Mazaik, & McKinlay, 1984). The latter have argued that the higher prevalence of chronic health conditions among homemakers can be accounted for almost completely by the proportion of women who report ill health as a reason for not being employed.

For causal inferences to be warranted, a longitudinal design alone is not sufficient. The minimum requirements for a rigorous test of the causal effect of employment are as follows: (a) observing employment status (or better: employment history), (b) assessing all relevant confounding factors (related to both employment and health), (c) assessing subsequent change in health, and (d) controlling the confounding factors statistically when regressing change in health on employment status (for elaborations of the causation issue, see, Cook & Campbell, 1979; Zapf, Dormann, & Frese, 1996). Because very few studies fulfill all of these requirements, it remains to be demonstrated whether methodologically sound longitudinal studies confirm the evidence generated by less rigorous approaches.

The above reasoning can be condensed into the following questions: (a) Do methodologically sound longitudinal studies confirm the cross-sectional findings? (b) Do different outcome measures yield converging results? In order to answer these questions, for each group of outcome measures we summarized the methodologically sound longitudinal studies that conformed to the minimum requirements (“Type II studies”) and compared their results to those of the remaining studies (“Type I studies”).

**Method**

**Literature searches**

The aim of the literature search was to identify all relevant empirical studies on the impact of employment on women’s health conducted between January 1950 and December 2000. In addition to published studies, we also tried to locate unpublished documents. We aimed to conduct as precise a search as possible, with “precision” referring to the ratio of relevant documents found to all those retrieved (White, 1994). Our search activities comprised three main strategies. The first was a computer-based search of major abstract databases (Medline, PsychInfo, and Sociofile) in which we combined title keywords from three domains: gender, sex, women, female, wife, housewife, homemaker, mother, or maternal with employment, work, occupation, or job with health, well-being, depression, disease, illness, morbidity, mortality, cardiovascular, or endocrine. The second strategy consisted in following up references in journal articles, book chapters, and books on employment and women’s health or related topics. Again, we searched for documents with at least one keyword from each domain in the title. The third strategy was to browse through journals and books in order to retrieve relevant documents not detected by the
previous two strategies. We looked through journals considered to be particularly relevant: the American Journal of Epidemiology, the American Journal of Public Health, the Journal of Health and Social Behavior, the Journal of Personality and Social Psychology, Social Science & Medicine, and Women & Health. In addition, we contacted researchers active in the relevant areas and inquired about unpublished work. To find further unpublished research papers and reports, we used additional resources provided by various institutions and organizations; e.g., the Abstract Newsletter published by the National Technical Information Service (NTIS), an agency of the US Department of Commerce, the Social Science Literature Information System (SOLIS) produced by the German Informationsszentrum Sozialwissenschaften (Social Science Information Center), and the Direct Access to Information (DATRAX II), an online retrieval system containing citations to dissertations and masters’ theses. We located a total of about 400 documents representing the pool of studies from which those to be included in our analyses were then selected.

Criteria for inclusion of Type I studies

In order to be included in the synthesis, a study had to (a) be a primary study with empirical findings based on quantitative methods of data collection and analysis, (b) include employment indicators as predictors and mental or physical health variables as criteria, (c) observe women aged between 16 and 68 years (studies on both sexes were only taken into account if the data were reported separately for women and men; results on older women were considered if information on middle-aged women was also available), (d) report an effect size and/or provide information about the statistical significance of the results (in the absence of an effect size, information about statistical significance in combination with the direction of the effect enabled us to apply vote-counting procedures). In the case of non-independence of studies, i.e., several reports based on the same sample population, we included only non-redundant independent and outcome measures. Study quality was not evaluated in this step. Overall, 140 articles reporting 153 studies met our criteria for inclusion.

Recorded variables

Study characteristics

We recorded the database, the measurement occasions, and the country in which the data were collected. Furthermore, we coded the type of employment indicator and the type of outcome variables used (psychological distress, general physical morbidity, neuroendocrine reactivity, cardiovascular risk and disease, and mortality), participants’ age, sex, marital and parental status, as well as their education and occupational status, if available. In addition, we recorded the sampling procedure and response rate, study design, and statistical analysis performed. For longitudinal studies, we recorded whether confounding variables were controlled for and whether alternative hypotheses such as reverse causation were tested. Moreover, the year and type of publication (journal article, book chapter, monograph, research report), the disciplinary background and the sex predominant in the author group were coded. Finally, we recorded whether the main hypotheses were deduced from theoretical models, and which models were concerned (role theory, person–environment fit, job model, gender model).

Most of the journal articles were published in public health journals (85% or 57%), followed by psychological and sociological journals (19% or 13% and 26% or 17%, respectively). Few studies were published before 1980, but the number of relevant publications grew rapidly during the 1980s, peaking in 1989 and 1992. After these peaks, the number of articles and book chapters began to decline again.

Criteria for identification of Type II studies

In a first step, we selected all studies with more than one measurement occasion (including prospective mortality analyses; N = 32). For a study to be included in the Type II sample, outcome and relevant covariates also had to be measured at the first measurement occasion and statistically controlled in multivariate analyses providing effect-size information. It emerged that 27 of the longitudinal studies controlled for the outcome assessed at the first measurement occasion, 21 of these also controlled in some way for other confounding variables. After excluding studies that (a) were redundant (N = 3), (b) reported exclusively effects of change in employment status on health (N = 3), (c) were restricted to a specific illness (N = 1), and (d) did not report effect sizes (N = 1) 13 studies remained which met the Type II criteria. With such small numbers of studies, central tendencies lack robustness because of large confidence intervals. Therefore, we decided to display each result separately instead of reporting central tendencies. Both authors assessed the 17 effect sizes of the Type II studies (intra-class correlation coefficient = 0.99) and discussed discrepancies until consensus was reached.

Summary

All 13 Type II studies were published in 1982 or later and 11 of them presented findings from the United States. The non-response rate was low in all studies. Four of the studies concentrated exclusively on married women. Information with regard to education,
occupational status, and the number and age of children was scarce. Table 1 displays the synopsis of study characteristics.

Results

Table 2 contains the vote-counting results for Type I studies and Type II studies. In general, the outcome of Type II studies is residualized change adjusted for varying confounders. Therefore, the size of the coefficients reported below reflects the increase in complaints, i.e., the smaller the coefficient, the smaller the increase.

The impact of employment on psychological distress

Ninety-one Type I studies and two Type II studies investigated the relationship between employment and psychological distress. Of the Type I studies, 48 reported beneficial effects, 6 adverse effects, and 18 no effects of employment. For 19 studies, the effects depended on the scores of other variables. Both Type II studies found employment to have beneficial effects. The data reported by Aneshensel (1986) originated from the Los Angeles Depression Study (LADS, \(N = 490\)), a 1-year panel survey. She found the multivariate odds of becoming depressed (as measured by the CES-D scale, Radloff, 1977) to vary with the quality of experience. Women with low-strain jobs had the smallest odds ratio (equated to 1.00), those with high-strain jobs had somewhat higher odds ratio (1.44), and those not employed had substantially higher odds ratio (3.01). We calculated the average effect of employment to be a 60% decrease in the risk of becoming depressed over the study period. Marital and employment roles did not interact in their impact on depression. Pavalko and Smith (1999) drew their data from the 1982 to 1989 surveys of the United States National Longitudinal Survey of Mature Women (NLSMW, \(N = 2763\)). Compared to women with other employment patterns, women who remained out of the labor force had the highest increase in psychological distress as measured by the CES-D (\(b = 2.25, p < 0.01\), one fourth of a standard deviation). This benefit of employment was found independent of initial status and did not reflect a mere selection effect. The levels of distress were not found to be a function of the length of labor force participation, however (\(b = 0.11, \text{SE} = 0.12, \text{ns}\)).

Alternative models (reverse effects) were tested in one of the Type II studies. Pavalko and Smith (1999) found differences in distress between women who left the labor force and those who remained employed or joined the labor force to be attributable mostly to health-related labor force exits.

Thus, both Type I and Type II studies found employment to be associated with reduced psychological distress in women. There was no indication that employment status interacted with other variables in its effects on distress.

The impact of employment on general physical morbidity

Fifty-three Type I studies and seven Type II studies investigated the relationship between employment and general physical morbidity. Of the Type I studies, 35 reported beneficial effects, 8 no effects, and 2 adverse effects of employment. For eight studies, the effects depended on the scores of other variables. Three of the Type II studies found beneficial effects, three found no effects, and one study found an adverse effect of employment. Ebi-Kryston, Higgins, and Keller (1990) used the 1959–60 and 1978–79 cycles of a large prospective epidemiological investigation in Tecumseh, Michigan (\(N = 764\)). They found self-reported health limitations and restrictions to be equally high for employed women and for homemakers at both measurement occasions. In a logistic regression model controlling for age, hypertension, and body mass index, the odds ratio relating employment to the prevalence rate of health limitations at follow-up did not differ reliably between homemakers (OR = 1) and employed women (OR = 0.76).

In the NLSMW Pavalko and Smith (1999) found that women who remained out of the labor force displayed the greatest increase in health limitations (\(b = 0.64, p < 0.001\), one sixth of a standard deviation). Like the finding on psychological distress, this result was interpreted as both a benefit of employment and a selection effect. In contrast to psychological distress, however, the increase in health limitations varied as a function of the length of labor force participation (\(b = -0.02, p < 0.05\)). All the effects were uniform, however, across job types, quantity of housework, and number of children present in the home.

Ross and Mirowsky (1995) used a probability sample of US households from the National Survey of Personal Health Practices and Consequences. Of the respondents aged 20–64 who were interviewed by telephone in 1979 and again in 1980, 1479 were female. The decline in both perceived health and physical functioning over the 1-year period was slower for women in full-time employment than for those without a full-time job (\(b = 0.04, p < 0.01, b = 0.04, p < 0.001\), respectively). There was no interaction with marriage or race.

Waldron and collaborators (1982b) obtained longitudinal data (1967, 1969, 1972, 1977) from a national sample of women aged 30–44 participating in the 1967 initial interview for the United States National Longitudinal Surveys of Labor Market Experience (NLSLME, \(N = 3606\)). They did not find significant relationships between labor force status in 1967 and subsequent change in health (the rate ratios for women
<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Data</th>
<th>Country</th>
<th>Predictors</th>
<th>Outcomes</th>
<th>No. of women</th>
<th>Initial characteristics</th>
<th>Statistical analysis</th>
<th>Main results</th>
<th>Moderator effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aneshensel (1986)</td>
<td>Los Angeles Depression Study, 1979–80 (T1–T4)</td>
<td>US</td>
<td>Role strain</td>
<td>Change in depressivity</td>
<td>490</td>
<td>Women of all ages</td>
<td>Log-linear models</td>
<td>Women with low-strain jobs had smallest odds of becoming depressed, those with high-strain jobs had higher odds, and those not employed highest</td>
<td>Marital status: no interactions</td>
</tr>
<tr>
<td>Ebi-Kryston et al. (1990)</td>
<td>Tecumseh Study, 1959–60 and 1978–79 (T1, T2)</td>
<td>US</td>
<td>Employment status at T1 and T2</td>
<td>Incidence of cardiovascular symptoms, physical health, controlled for hypertension and BMI, and mortality after 20 years</td>
<td>866</td>
<td>Women aged 20–44 years</td>
<td>Logistic regression</td>
<td>No differences in cardiovascular symptoms and physical health by employment status, but higher mortality risk for those women employed at baseline or previously compared to those never employed</td>
<td>No interactions reported</td>
</tr>
<tr>
<td>Hibbard and Pope (1991)</td>
<td>Center for Health Research of the north-western Region of Kaiser Permanente, 1970–71 and 1985–86 (T1, T2)</td>
<td>US</td>
<td>Employment status and multiple roles at T1</td>
<td>New cardiovascular morbidity and mortality after 15 years (controlled for self-reported health status at baseline)</td>
<td>1160</td>
<td>Women and men aged 18–65 years</td>
<td>Cox proportional hazards models</td>
<td>Employment related to a reduced risk of death, but not to the risk of new cardiovascular morbidity</td>
<td>Parental status: employed women with at least one child had a lower risk of new cardiovascular morbidity than those without children</td>
</tr>
<tr>
<td>Kotler and Wingard (1989)</td>
<td>Alameda County Study, 1965 (T1)</td>
<td>US</td>
<td>Employment history and status at T1</td>
<td>Mortality after 18 years (controlled for physical health at baseline)</td>
<td>1969</td>
<td>Women and men aged 35–64 years</td>
<td>Relative risks and Mantel-Haenszel $\chi^2$</td>
<td>No differences in mortality risk between women full-time employed for at least 50% of their lives and housewives</td>
<td>No. of children: housewives with children had a higher mortality risk, especially those with four or more children in the home</td>
</tr>
<tr>
<td>Moen et al. (1992)</td>
<td>Data from an upstate New York community, 1956 and 1986 (T1, T2)</td>
<td>US</td>
<td>Patterns of employment and multiple roles</td>
<td>Functional and subjective health after 30 years (controlled for poor health status at baseline)</td>
<td>313</td>
<td>Married mothers aged 55–81 years at T2</td>
<td>OLS regression</td>
<td>Being continuously employed negatively related to subjective health, but not to functional ability; ever having been employed negatively related to functional ability but not to subjective health; neither outcome related to duration of employment</td>
<td>Sample of married mothers, no interactions reported</td>
</tr>
<tr>
<td>Pavalko and Smith (1999)</td>
<td>National Longitudinal Survey of Mature Women, 1982, 87 and 89 (T1–T3)</td>
<td>US</td>
<td>Patterns of employment</td>
<td>Change in physical health limitations and psychological distress</td>
<td>2763</td>
<td>Women aged 45–59</td>
<td>OLS regression</td>
<td>Compared to women with other employment pattern, women who remained out of the labor force had highest increase in health limitations and distress (for health limitations, but not for distress, the increase varied as a function of the length of labor force participation)</td>
<td>Job types, No. of children: no interactions</td>
</tr>
<tr>
<td><strong>Reviere and Eberstein (1992)</strong></td>
<td>National Health and Nutrition Survey, 1971–74 and 1982–84 (T1–T4)</td>
<td>US</td>
<td>Employment status at T1 and T2</td>
<td>New coronary heart disease at T2</td>
<td>3097</td>
<td>Women aged 25–44 years</td>
<td>Logistic regression</td>
<td>Compared to women with the same job at both times, homemakers at both times had a higher risk of being diagnosed with heart disease between baseline and follow-up</td>
<td>Marital status: for working women, becoming unmarried was related to a lower rate of heart disease</td>
</tr>
<tr>
<td><strong>Ross and Mirowsky (1995)</strong></td>
<td>National Survey of US Personal Health Practices and Consequences, 1979 and 1980 (T1, T2)</td>
<td>US</td>
<td>Employment status at T1</td>
<td>Change in perceived health and physical functioning</td>
<td>2436</td>
<td>Women and men aged 20–64 years, 85% White, 65% married</td>
<td>OLS regression</td>
<td>Full-time employment related to slower declines in perceived health and physical functioning</td>
<td>Marital status, race: no interactions</td>
</tr>
<tr>
<td><strong>Vagerö and Lahelma (1998)</strong></td>
<td>Multiple Census-Linked Deaths Registry, 1970, 1980</td>
<td>SE</td>
<td>Employment status at T2</td>
<td>Mortality between 1980 and 1986 (sensitivity analyses estimating selection effects)</td>
<td>Married women who were homemakers and between 20 and 49 in 1970</td>
<td>Age-standardized death rates</td>
<td>Women who moved into full-time or part-time employment had lower mortality risks than homemakers</td>
<td>Occupational class: no interactions</td>
<td></td>
</tr>
<tr>
<td><strong>Waldron and Jacobs (1989)</strong></td>
<td>National Longitudinal Surveys of Labor Market Experience, 1977 and 1982 (T1, T2)</td>
<td>US</td>
<td>Employment status and multiple roles at T1</td>
<td>Change in health problems</td>
<td>30% white and 90% black women (5% with preschool-age children)</td>
<td>Women aged 30–44 years (40–54 at the beginning of the study interval)</td>
<td>OLS regression</td>
<td>No differences in health trends by labor-force status</td>
<td>Marital and parental status, race: no interaction with parental status, for white women, no interaction with marital status, for black women, effect only for those with children</td>
</tr>
<tr>
<td><strong>Waldron et al. (1982b)</strong></td>
<td>National Longitudinal Surveys of Labor Market Experience, 1967, 1969, 1972 and 1977 (T1, T2, T3, T4)</td>
<td>US</td>
<td>Employment status at T1</td>
<td>Change in self-rated health</td>
<td>3606 white women and 1390 black women</td>
<td>Women aged 30–44 years, analyses only for married women</td>
<td>Log linear models</td>
<td>No significant relationships between labor force status and subsequent change in health</td>
<td>Usual occupation, race: no interactions</td>
</tr>
<tr>
<td><strong>Waldron et al. (1998)</strong></td>
<td>National Longitudinal Surveys of Labor Market Experience, 1978, 1983 and 1988 (T1, T2, T3)</td>
<td>US</td>
<td>Employment status at T1</td>
<td>Change in physical health problems</td>
<td>~3000</td>
<td>Women aged 14–24 years</td>
<td>OLS regression</td>
<td>Employment related to a decelerated increase in health problems for unmarried women</td>
<td>Marital status, parental status: positive effects lower for married women, effect no function of parental status, but decreased with increasing no. of preschoolers</td>
</tr>
<tr>
<td><strong>Weatherall et al. (1994)</strong></td>
<td>OPCS Longitudinal GB Study, 1971 (T1)</td>
<td>GB</td>
<td>Employment status at T1</td>
<td>Mortality within 1976 and 1985 (first 5 years of follow-up omitted)</td>
<td>93,576</td>
<td>Married women aged 15–59 years in 1971</td>
<td>Logistic regression</td>
<td>For all employed women, risk of death lower than for homemakers</td>
<td>Parental status, occupational class, working time: women with neither job nor child at greater risk than all others; non-manual workers had largest advantage, part-time workers slightly greater advantage than full-time workers</td>
</tr>
</tbody>
</table>
in the labor force as compared to those out of the labor force in 1967 were 6.1–5.3 and 11.1–10.1 for the two initial levels of self-rated health). Neither did they find labor force status to interact with usual occupation or race in its effects on health.

Waldron and Jacobs (1989) also used longitudinal data (1977–1982) from the NLSLME ($N = 2392$). The mean increase in the health problems score between 1977 and 1982 was $+0.57$. From the graphical display of the relationship between health trends and roles held, we took the average effect of labor force participation to be $b = 0.05$.

According to the conventional 5% level of significance, health trends were unrelated to 1977 labor force status for married white women as well as for unmarried white women (but less clearly so). Having children did not make a difference but only 5% of the mothers had preschool-age children.

Waldron and collaborators (1998) also used longitudinal data (1978, 1983, 1988) from the NLSLME, with a sample of women aged 14–24 in 1968 ($N = 3331$ and 2989 for the first and the second interval). Employment was found to be associated with a decelerated increase in health problems for unmarried women ($b = -0.39, p < 0.01$, and $b = -0.75, p < 0.0001$, for 1978–1983 and 1983–1988, respectively). The beneficial effect of employment was substantially lower for married women, ($b = 0.06$ and $b = -0.18$). This differential was consistent with the role-substitution hypothesis, according to which employment and marriage have a substitutive relationship with regard to the provision of health-related resources. Although the effects of employment on health did not differ as a function of parental status, they became less positive as the number of preschoolers in the home increased. We calculated the average effect of employment to be a deceleration in the increase in health problems of $-0.10$. This was the only study to report adverse effects of employment. Can this deviant result be explained by any characteristic of the study? In the discussion, we will consider the two salient differences to the other studies: The sample consisted of married women, exclusively, and the

### Table 2
**Impact of employment on psychological distress, physical morbidity, cardiovascular risk factors, and mortality for studies conforming (Type II) and not conforming (Type I) to the criteria for methodologically sound studies**

<table>
<thead>
<tr>
<th>Effects reported in Type I studies</th>
<th>Psychological distress</th>
<th>General physical morbidity</th>
<th>Cardiovascular risk/disease</th>
<th>Mortality</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(N = 91)$</td>
<td>$(N = 53)$</td>
<td>$(N = 15)$</td>
<td>$(N = 3)$</td>
<td></td>
</tr>
<tr>
<td>Beneficial effect</td>
<td>48</td>
<td>35</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Adverse effect</td>
<td>6</td>
<td>2</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Depended on moderators</td>
<td>19</td>
<td>8</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>No effect</td>
<td>18</td>
<td>8</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>Effects reported in Type II studies</td>
<td>$(N = 2)$</td>
<td>$(N = 7)$</td>
<td>$(N = 3)$</td>
<td>$(N = 5)$</td>
</tr>
<tr>
<td>Aneshensel (1986)</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ebi-Kryston (1990)</td>
<td></td>
<td>0</td>
<td>0</td>
<td>+</td>
</tr>
<tr>
<td>Hibbard and Pope (1991)</td>
<td></td>
<td></td>
<td>0/M</td>
<td>+</td>
</tr>
<tr>
<td>Kotler and Wingard (1989)</td>
<td></td>
<td></td>
<td>0/M</td>
<td></td>
</tr>
<tr>
<td>Moen et al. (1992)</td>
<td></td>
<td></td>
<td>0/M</td>
<td></td>
</tr>
<tr>
<td>Pavalko and Smith (1999)</td>
<td></td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>Reviere and Eberstein (1992)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ross and Mirowsky (1995)</td>
<td></td>
<td>+</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vageró and Lahelma (1998)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Waldron and Jacobs (1989)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Waldron et al. (1982b), Study 1</td>
<td></td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Waldron et al. (1998)</td>
<td></td>
<td>+/M</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weatherall et al. (1994)</td>
<td></td>
<td></td>
<td></td>
<td>+/M</td>
</tr>
</tbody>
</table>

*Note: +, positive effect; −, negative effect; 0, null effect; M, moderator effects.*
catch-up design bridged an extraordinarily long period of time.

Only Waldron and collaborators (1982b) analyzed the relationships between health and subsequent change in labor force participation. Labor force entries were more likely for women who did not report any health-related limitations to their activities and for those who rated their health as good or excellent at baseline. Labor force exits were more likely for women who reported health-related activity limitations at baseline.

Thus, with regard to general physical health, the pattern of findings from the Type II studies diverged from the Type I studies. Whereas most Type I studies reported beneficial effects, Type II studies did not find employment to be associated with health. In the seven Type II studies, only one reliable moderation was reported in which employment status interacted with marital status in its effects on health.

The impact of employment on cardiovascular risk and disease

Fifteen Type I studies and three Type II studies investigated the relationship between employment and cardiovascular risk and disease. The risk factors for and indicators of cardiovascular disease examined include systolic and diastolic blood pressure, high- and low-density lipoprotein levels, cholesterol level, triglycerides, evidence of coronary heart disease, myocardial infarction, and coronary death.

Of the Type I studies, four reported beneficial effects, one adverse effects and seven no effects of employment. For three studies, the effects depended on the scores of other variables such as marital status, parental status, or occupational status. One of the Type II studies reported a beneficial effect, two studies reported no effect of employment. Ebi-Kryston and colleagues (1990) found women who were employed at both measurement occasions to be 2.68 times less likely to report a history of heart attack at follow-up than those who were homemakers at both times (after controlling for relevant covariates such as diagnosed hypertension and body-mass index). In contrast to this self-report measure, the proportion of women with ECG evidence of coronary heart disease did not differ reliably between employment groups, although those who were homemakers at both times were diagnosed as having coronary heart disease least often, and their systolic and diastolic blood pressure levels were slightly higher at both baseline and follow-up. Age-adjusted incidence rates between baseline and follow-up for shortness of breath, chest pain, and diagnosed hypertension did not differ between employment groups. These findings seemed to average out and were therefore represented by an odds ratio of one.

Hibbard and Pope (1991) analyzed a survey conducted in 1970–71 by the Center for Health Research of the United States Northwest Region of Kaiser Permanente (N = 1160). The outcome variable was ischemic heart disease diagnosed subsequent to the interview date. The adjusted 15-year relative hazard for women employed at baseline was 29% smaller than that for women not in employment, but the difference was not statistically reliable (CI 9–2.3). For mothers, this difference was reliable. Reviere and Eberstein (1992) used data from the United States National Health and Nutrition Survey (NHANES I, 1971–74) and the NHANES Epidemiologic Follow-up Survey (1982–84) restricted to women who reported at Time 1 that they had never had a diagnosis of heart disease (N = 3097). The probability of being diagnosed with heart disease between baseline and follow-up was increased by 0.624 (p < 0.01) for those who were homemakers at both times and slightly, but not reliably, for women who took up employment over the period of investigation, both compared to women with the same job at both times. For working women, becoming unmarried was related to a lower rate of heart disease.

For cardiovascular risks and diseases, the findings of the Type II studies thus replicated the relationships found across Type I studies: Only one study found a beneficial effect of employment, while the remaining two found no effects. One reliable moderation was reported in the three studies: for employed women only, parental status was associated with a positive effect.

The impact of employment on mortality

Three Type I studies and five Type II studies examined the relationship between employment and mortality. Two Type I studies reported beneficial effects of employment; for the third study, the effect depended on the scores of other variables. Four of the Type II studies reported a beneficial effect, one study reported no effect of employment. In the Tecumseh sample described above, Ebi-Kryston and colleagues (1990) found the age-adjusted mortality rates for women employed at baseline or previously and those never employed to differ reliably (4.6 vs. 5.2, p < 0.05).

The 15-year relative risks of death in the sample recruited in the Northwest Region of Kaiser Permanente (Hibbard & Pope, 1991, introduced above) were reduced by 41% for employed as compared to non-employed women after adjusting for age, education, and self-reported health. Vagerö and Lahelma’s (1998) study was based on the Swedish Multiple Census-Linked Deaths Registry (MCDR). The population of married women who were homemakers (excluding the “other non-employed”) and between 20 and 49 years of age in 1970 were included in the first step (N = 352,900). The findings were not adjusted for health at baseline, but the
authors conducted sensitivity analyses (see below). Women who worked at least 35 h a week (full time) or at least 20 h a week (long part time) were compared to homemakers; students and women working shorter hours were excluded. Regardless of the occupational class that women who joined the labor force moved into, they were found to be at an advantage over the homemaker group in terms of health (RR = 0.54).

Weatherall and colleagues (1994) used data from the OPCS Longitudinal Study subsample (Office of Population Censuses and Surveys) of married women aged 15–59 in 1971 (N = 93, 576) to examine all-cause mortality between 1976 and 85. Although the findings were not adjusted for health at baseline, we decided to include this study as a Type II Study because the authors omitted the first 5 years of the follow-up period to reduce the impact of health selection. The relative risks of death were lower for all employed women than for homemakers (except one, all 95% confidence intervals below 0.92). Manual workers had the smallest advantage, non-manual workers the largest. Part-time workers had a slightly greater advantage than full-time workers. We calculated the average advantage of employed women to be 20% (RR = 0.80).

There was no effect of employment in the United States Alameda County Study, however (Kotler & Wingard, 1989). Women aged 35–64 in 1965 (N = 1457) who were in full-time employment outside the home for at least 50% of their lives from age 22 on did not differ from homemakers in their age-adjusted 18-year risk of mortality (13.5% and 14.2%, respectively, RR = 0.95, CI = 0.06, 1.2).

Most researchers conceded that the relative importance of selection and causal explanations of the observed mortality differentials remains unclear. Vagerö and Lahelma (1998) performed a sensitivity analysis to estimate the impact of reverse causation. Simulating the mortality advantages with modest (twofold risk), strong (threefold risk), and extreme (tenfold risk) health selection while taking psychological and circulatory problems into account showed the expected advantages of employed women to be 13%, 21%, and 37%. They concluded that the observed advantage of employed women (54%) could not be fully explained by health selection.

The clearest picture thus emerged with regard to mortality, with all but one study demonstrating employed women to be at an advantage over non-employed women. The degree to which this was attributable to social causation above and beyond social selection could not be determined. Two studies reported moderator effects. In one study, women with neither job nor child were at greater risk than all others; in the other, housewives with children had a higher mortality risk, especially those with four or more children in the home. These effects are not consistent, however, and the groups on which they are based seem to be rather exceptional. Moreover, little is known about the intervening processes, which—given that mortality is such a distal outcome—are of particular interest.

Discussion

The main result of this research synthesis is that methodologically sound longitudinal studies confirm the findings of cross-sectional research showing that employment has either beneficial or neutral effects on women’s health. Overall, there is agreement between Type I and Type II studies, as well as between the different outcome measures, that employment has no adverse effects in the general population of employed women. Some effects are moderated by other variables such as marital and family status but these interactions are not consistent across studies. There are also some significant differences between studies, however. Studies that examined subjective distress and mortality showed employment to have a beneficial effect, whereas studies with subjective health or cardiovascular risks and disease as outcomes obtained neutral results. The only study to find an adverse effect of employment took subjective health as an outcome.

The main finding is directionally opposite to the stress hypothesis: The absence of adverse effects is evidence in support of its rival, the enhancement hypothesis. Although we tried to rule out artifactuality by including only methodologically sound studies and although the effect held under a broad range of conditions (e.g., different countries, different subject populations, different operationalizations of the theoretical concepts such as the four sets of outcome measures), thus indicating external and construct validity of the measures, causal inference is still not warranted. A research synthesis cannot surmount the design constraints of the primary studies: If these do not fulfill the criteria for causality, there is no way this limitation can be overcome. For instance, (long-term) side effects of their efforts to cope with multiple roles may have resulted in role relinquishment for women who did not succeed in adapting to the demands. Another limitation is the dependence of the accumulated database on the employed search strategies. We are confident, however, that potential biases are small because we used a broad range of search strategies.

Methodological implications

Even when adjusting for the past history of time-varying confounding factors, researchers may still have obtained biased estimates in Type II studies. Robins (1986) proposed several methods that yield valid estimates of the causal effect of a time-varying treatment in the presence of time-varying confounding factors.
One prerequisite that is difficult to achieve is the assessment of all prognostic factors: observational studies lack evidence about confounding by unmeasured factors. Sensitivity analyses seem to be a pragmatic solution for future studies (Robins, 1999). These serve to quantify the extent to which the inference concerning the treatment effect varies as a function of the magnitude of confounding by unmeasured factors. In addition, correlational research should be complemented by experimental studies in which the treatment effect can be separated from confounders (in the case of employment effects, this can be achieved through natural experiments or by varying the effects of intervening variables).

Given that—with the exception of one outlier—the effects appear to be homogeneous, our conclusions seem very plausible. This outlier in question is the adverse effect of employment on health found by Moen and collaborators (1992). Post hoc, we hypothesize two possible reasons for this discrepant finding. One is the exceptionally long follow-up period of 30 years; the durations of the other studies ranged from 1 to 18 years, with a mean and median of 10 years. It would be worth investigating the impact of different follow-up periods, because inadequate spacing of the measurement occasions could have led to some effects being missed (Fresen & Zapf, 1988). In the study by Hibbard and Pope (1991), for instance, there were no (reliable) effects of employment at both 5 and 10 years post-interview; the advantage seemed to take 15 years to manifest itself. There are physical processes such as arteriosclerosis that accumulate over decades before their effects are noticeable. Moreover, the time courses of different outcomes may differ (Cohen, Kessler, & Gordon, 1997), but the theoretical models are not sufficiently elaborated to allow for hypotheses regarding the functional forms linking their elements to be deduced. Because data on the effect of the temporal extension of the exposure variable are as yet scarce and inconsistent (Moen et al., 1992; Pavalko & Smith, 1999), future studies should aim at assessing dose–response relationships.

Alternatively, the discrepancy may be explained by other characteristics of the sample, such as ethnicity and marital status. Several studies have found that white married women do not benefit as much from employment as other groups of women do (Waldron & Jacobs, 1989), and that if employed in white-collar jobs, they may even be worse off than their non-employed counterparts (Waldron & Jacobs, 1988). Theoretically, this conforms with the role-substitution hypothesis (Waldron et al., 1998), but not all the reviewed results were in line with this hypothesis. Of the four Type II studies that examined exclusively married women, two found a beneficial effect (Vagero & Lahelma, 1998; Weatherall et al., 1994), one a neutral (Waldron et al., 1982b), and one an adverse effect of employment (Moen et al., 1992). These studies differed in two salient characteristics, namely type of outcome measure and national context. Neutral and adverse effects were reported for subjective/functional health as outcome and US contexts whereas beneficial effects were reported for mortality as outcome and European contexts (Sweden and Great Britain). Scandinavian countries differ from most other countries in that women’s total labor force participation is high, part-time jobs are of comparable status to full-time jobs, and the availability of quality child care corresponds to the existing demand (Arber & Lahelma, 1993). Such cultural differences may modify the relationship between employment and well-being. Unfortunately, because of the small number of Type II studies, it was not possible to determine the relative contributions of different study characteristics to the differences in reported effects in order to identify conditions under which employment has adverse or beneficial consequences.

Theoretical implications

In general, the studies reviewed seemed to lack theoretical underpinning. According to our coding scheme, about a quarter of the authors performed theory-guided hypothesis testing. In the vast majority of cases, however, the choice of predictors and outcomes seemed to be based on common sense or to be determined by historical and pragmatic factors such as the availability of measurement instruments. In conclusion, we can only repeat what our predecessors, Warr and Parry (1982a, b, p. 514), wrote 20 years ago: “there is no justification for further empirical comparisons between the (…) well-being of women in general who have jobs and those who do not. Research must examine more precise hypotheses”.

Although not all the hypotheses that could be derived from role-theoretical models seem to have been tested, our call is not for more of the same. Theory development can only be advanced if researchers specify mediating processes. The role-theoretical models have to be elaborated to allow for specific hypotheses to be deduced and tested. There is a paucity of relevant data because existing research has hardly ever assessed intervening concepts. Instead, researchers have speculated post hoc about possible mediators and moderators to reconcile inconsistent findings.

Which mediating processes could account for the association between employment and emotional and physical well being? At least two different mechanisms are likely to be involved: competence or mastery (Karasek & Theorell, 1990; White, 1959) and social affirmation (Thoits, 1983). Environmental challenges form the starting point for the first mechanism. These afford opportunities for active learning, and accumulated learning experiences in turn stimulate the
development of skills and of a sense of mastery or competence. The expanded range of solutions to environmental challenges reduces the person’s perception of and reactivity to stress and raises her future activity level (Karasek & Theorell, 1990, p. 92). In an upward spiraling effect, strain is thus inhibited. Because gains in competence occur most often when challenges in the situation are matched by the individual’s level of control, environments high in challenges and high in control are thought to promote physiological growth and regeneration. Testing whether this is the mechanism that explains health differences between employed and non-employed women requires an instrument that is applicable to market and family work. The AVAH (Resch, 1992) is such an instrument. In an action theoretical approach, all action units with the same goal are pooled and described according to a number of dimensions: (a) levels of planning and decision making, (b) forms of cooperation, (c) levels of care intensity, and (d) flexibility of time use. To get this information, however, one has to go through a time-intensive interview and coding procedure. An alternative approach is the sampling of experience in work (paid and unpaid) and leisure activities (e.g., Klumb & Perrez, in press).

The second mechanism involves social interactions. Affirmative social interactions afford opportunities for experiencing relatedness, competence, and autonomy. The fulfillment of these core needs is seen as the basis for the association between social interactions and well-being (Reis, Sheldon, Gable, Roscoe, & Ryan, 2000). And as Thoits (1983) has argued, identities are sustained in reciprocal role relationships and social integration may enhance an individual’s sense of meaningful existence and purposeful behavior. This experience is crucial to mental health and well being.

Moreover, the social environment can have effects on environmental demands and well-being and moderate their relationship (Cohen & Wills, 1985; Seemann, Berkman, Blazer, & Rowe, 1994; Seeman & McEwen, 1996). For instance, social support can (a) reduce demands through practical support with problem solving or activities, (b) alter basal neuroendocrine profiles, and (c) modify the relationship between specific demands and reactivity by influencing the processes of primary and/or secondary appraisal and thereby attenuating stress responses (i.e., lower response levels and faster recovery to baseline).

Social networks do not have exclusively positive effects, however. They can create or exacerbate existing stress, through excessive demands or criticism, for example. It seems that women are more exposed to these negative effects than men (Rook, 1984). To decide whether a specific social interaction functions as a resource or as a demand researchers have to assess its specific context (Cohen, Evans, Stokols, & Krantz, 1986; Schönpflug & Battmann, 1988). We thus recommend ecological momentary assessments of the dynamics of social stressors and social resources, both in the family and at the workplace (Buunk & Peters, 1994) to test whether differences in social affirmation explain health differences between employed and non-employed women. This research relies on small samples because representativeness is not required and large samples are very costly. Yielding insights into processes it complements research with large representative samples.

Conceptualizing these intervening mechanisms on the basis of a framework that relates intra-individual and inter-individual differences would help to fill the gap we have identified in the literature. Intra-individual differences in exposure (positive as well as negative aspects) and in well being as well as their associations could be conceived of as functions of stable and time-varying characteristics such as the current constellation of personal and social attributes such as marital status/satisfaction. This perspective provides new possibilities for testing hypotheses regarding role constellations such as the role-substitution and role-complementation hypotheses.

With regard to outcome measures, it is important to complement self-reports with indicators that are independent of an individual’s symptom perception and reporting (Verbrugge, 1989). This can be accomplished by using multiple measures such as affective, cognitive, cardiovascular, and neuroendocrine ones. Moreover, it seems desirable to broaden the construct of well being so that it represents not only the absence of negative but also the presence of positive dimensions.

Acknowledgements

We wish to thank Claudia Semmler for her assistance with the literature search and collection. This research was funded by a grant from the Volkswagen Foundation to the first author.

Studies preceded by one asterisk were cross-sectional (Type I), studies preceded by two asterisks have longitudinal designs (Type I), those with three asterisks were coded as Type II studies. Studies without asterisk are those quoted in the introduction and discussion sections.

Author contributions: PK conceived the approach presented here, suggested criteria for including studies, for distinguishing Type I and Type II studies and the variables to be coded, coded Type II studies and drafted and revised the manuscript. TL performed the search and selection of studies, coded Type I and Type II studies, and documented the procedure.

References


Pietromonaco, P. R., Manis, J., & Frohardt-Lane, K. (1986). Psychological consequences of multiple social roles. Psychology of Women Quarterly, 10, 373–382 *


White, H. D. (1994). Scientific communication and literature retrieval. In H. Cooper, & L. V. Hedges (Eds.), The
handbook of research synthesis (pp. 41–55). New York: Russell Sage Foundation.


Further reading


