Conditions, Interventions, and Outcomes: A Quantitative Analysis of Nursing Research (1981-1990)

Ivo L. Abraham, PhD, CS, FAAN
Zona L. Chalifoux, MSN, RN
Georges C. M. Evers, PhD, RN

Nursing, as an effort institutionalized by societies to fulfill specific needs and demands for individuals and groups, is practical rather than theoretical in nature. However, this orientation, coupled with society's expectation to serve individuals and groups both effectively and efficiently, does not relieve nursing of the necessity to base its practices upon scientific knowledge. Specifically, there is a mandate for nursing, as for any other discipline within the clinical life sciences, to document the efficacy of some interventions and the inefficacy of others. Thus, it will become possible for the discipline of nursing to know which interventions help individuals and groups and which do not, and understand why.

In this paper, we report on a quantitative analysis of intervention/outcome studies published in six general nursing research journals between 1981 and 1990. Having evaluated the research reports along such dimensions as publication characteristics, subjects, studied, design issues, statistical power, interventions, outcomes, and conditions, we attempt to sketch trends and issues pertinent to the (small) proportion of published studies that tested the effects of nursing interventions on patients with or without specified conditions.

Review of the Literature

Critical reviews, such as those by Ellis (1977), Gunter and Miller (1977), and Highriter (1977), have facilitated the development of nursing research by describing its status in terms of content and method at particular time points from 1952 to 1977. These reviews, although innovative at the time, lack the scope and depth of analysis necessary to track the development of nursing as a scientific discipline with a practice referent (Ellis, 1984). Recent review studies, which examined more current publication trends in nursing research, have added greater precision to nursing's understanding of its evolution as a scientific discipline.

The early 1980s witnessed a transition in the focus of nursing research from nurses and nursing students (Abdellah, 1977; Gortner & Nahm, 1977) to nursing practice (Brown, Tanner, & Padrick, 1984; Jacobson & Meininger, 1985). Between 1985 and 1981, in four leading nursing journals, studies with a nursing action central to the study increased by 156% and those concerned with patient behavior increased by 300% (Moustafa, 1985). The ten years from 1977 to 1986 saw more than a sixfold increase in the number of studies on nursing practice reported in six general nursing journals (Moody, Wilson, Smyth, Schwartz, Tittle, & Van Cott, 1988). Over one third of these studies (36%)

were related to (but did not necessarily test interventions), with the remaining two-thirds being focused on assessment.

Of the 720 articles reviewed by Moody *et al.* (1988), adult health was the most frequent clinical focus (39.0%), followed by maternal infant (12.0%), critical care (10.0%), gerontology (9.3%), child health (8.0%), women's health (7.4%), and community health (.4%). The low frequency of research directed toward the elderly, children, and adolescents has also been pointed out by others (Brown *et al.*, 1984; Heater, Becker, & Olson, 1988; O'Connell & Duffy, 1978).

Sample characteristics such as gender and minority status have received little attention in previous reviews of nursing research. For example, the majority of studies reported from 1977 to 1986 were not gender-specific (65%), while 22% studied only women and 8% only men (for 5% of studies, pertinent information was absent) (Moody *et al*., 1988). A real, yet statistically nonsignificant increase in women's health studies since 1970 was noted by Dunbar (1982). Twice as many studies focused on maternal role as compared to women's general health issues.

Most nursing research studies have been observational/descriptive and this trend is independent of time (Brown *et al.*, 1984; Jacobson & Meininger, 1985; Moody *et al*., 1988). Jacobson and Meininger (1985) reviewed 317 selected, non-methodological non-historical studies published in three general research journals over a 27 year period. Only 27% were experimental or quasi-experimental, and the remaining (73%) were observational. Importantly, experimental studies peaked in the mid 1960s (41%) and declined through to the early 1980s (20%).
When used, quasi-experimental designs have been more prevalent than true experimental designs (24% versus 6%; Moody *et al.*, 1988). Jacobson and Meininger's (1985) found that the most frequent designs between 1956 and 1983 were the nonrandomized trial with multiple groups (33 of 56 studies, or 58.9%) and the repeated applications or withdrawals of treatments with subjects serving as their own controls (13 of 56, or 23.2%).

Univariate/bivariate statistics have been the most common class of statistical methods, regardless of whether or not studies were (quasi-)experimental (Moody *et al*., 1988). Several authors have identified consistent problems with data analysis in nursing research: using multivariate statistics with inadequate sample size (Moody *et al*., 1988); using univariate/bivariate methods when multivariate are indicated and might yield more informative findings (Abraham, Nadzam, & Fitzpatrick, 1989); failure to use nonparametric statistics when indicated (Moody *et al.*, 1988); lack of sample size and power calculations (Jacobson & Meininger, 1986); and omission of attrition rates with reason for withdrawal after random assignment (Jacobson & Meininger, 1986).

Methods

Scope

We reviewed ten volumes of six refereed general nursing journals committed to research and scholarship published during the ten-year period from 1981 to 1990: *Nursing Research (NR), Research in Nursing and Health (RINAH), Western Journal of Nursing Research (WJNR), Advances in Nursing Science (ANS), International Journal of Nursing Studies (IJNS),* and *Journal of Advanced Nursing (JAN).* Every article in each of these journals was reviewed as to whether it constituted a report, in full or in part, of a nursing intervention study. A nursing intervention study was defined as an investigation, in which a nursing or a nursing-relevant intervention is manipulated actively (experimentally or quasi-experimentally) or passively (ex post facto) with the goal of effecting outcomes in a sample of subject. Articles meeting this requirement were coded and evaluated in terms of publication characteristics, subjects studied, design and analysis issues, statistical power, interventions, and outcomes. Excluded from consideration were editorials, commentaries on articles and rejoinders (except if a full article in itself), letters to the editor, and short "column"-type publications. The *International Journal of Nursing Studies* and the *Journal of Advanced Nursing* were included to add a European dimension to the investigation. Even though in this investigation the unit of measurement was an article on an intervention study, for ease of reporting we equate each article with a study.

Dimensions and Variables

Research reports were coded along several methodological and content dimensions and variables.

Publication Characteristics

Journal in which article was published.
Year of publication.
Number of Authors.

Multidisciplinary: if the paper had two or more authors, it was assessed for whether or not the team of authors was multidisciplinary.

Subjects

Age Group: the predominant age group studied (0-17 child; 18-59 adult; 60+ older adults). If not indicated specifically, mean age of the sample was considered. For animal studies, this variable was coded as the age group for which the animals served as a model.

Condition-Specific: based on sampling criteria or other information, whether a given study was specific to one or more clinical conditions. Admittedly subject to debate but for practical classification reasons, perinatal and other childbearing-related conditions were coded as condition-specific.

Organ System: for condition-specific studies, the organ system(s) involved were coded: nervous system, central and peripheral (including, psychiatric conditions); cardiovascular system; endocrine system; urinary system; hematological system; immune system; digestive system; respiratory system; reproductive system (male/female; for women this included normal pregnancies); sensory system; musculoskeletal system; skin; or other.

Minorities: studies were coded as to whether they did (purposively or by sampling availability) or did not include minority subjects; or whether, from the report, it was impossible to determine (and therefore unknown) if minorities were included. This and the next variable were included in part because of the recent criticisms that biomedical research has not sufficiently included minorities and women.

Women: studies were coded as to whether they did (purposively or by sampling availability) or did not include women subjects; or whether, from the report, it was impossible to determine (and therefore unknown) if women were included.

Design Issues

Type of Design: design were coded as either single to $4 factor studies with and without repeated measures, blocked (simple blocking and latin, greco-latin, and hyper-greco-latin squares), hierarchical, split-plot, ex post facto, or other (i.e., unclassifiable "hybrid" designs). Coding was on the basis of our analysis of the design, not necessarily what was stated by the authors.

Number of Groups featured in the design, including treated or untreated control/comparison groups.

Control Group: did study include a control or comparison group, treated or untreated?

Random Assignment: were *all* subjects randomly assigned to study conditions; or for one-group studies in which subjects received sequential treatments, were treatments randomly ordered across subjects?

Time Points: number of observation points featured in the design (not necessarily equal to number of time points reported in data analysis).

Attrition: did the study suffer loss of subjects over the course of the investigation? Often, by necessity scoring this variable was done on the basis of indirect information (e.g., subsample sizes at lagged time point not summing to stated sample size, degrees of freedom discrepant from stated sample size, etc.).

Sample Size Left: for studies with attrition, what percentage of the original sample size was left (i.e., inverse of attrition rate)?

Statistics: descriptive (noninferential), univariate or multivariate (inferential). Multivariate was defined conventionally as involving the statistical analysis of two or more dependent variables.

Effect Size: was a quantitative, standardized estimator of the size of difference among groups (i.e., effect of interventions) included (Abraham, Schultz, Polis, Vines, & Smith, 1987)?

Statistical Power

Studies testing the effectiveness of nursing interventions use statistical methods to test a null hypothesis ("intervention has no effect") when in fact the interest of the investigators is in the alternative hypothesis ("intervention has an effect"). Statistical power, the probability of not committing a Type II or b error, assures the integrity of the (nonstatistical) inferences drawn about the alternative hypothesis. While .80 has become a generally accepted power level, it still allows a 20% error margin in inferences about the alternative hypothesis and the presumed effects of an intervention. Power of .90 is more appropriate.

Power Analysis: did the report refer to a power analysis? This was scored as "yes" if the article reported the power analysis, made mention or implied that a power analysis had been conducted.

Power .80 Attained and ***Power .90 Attained***: using power tables from Kirk (1982), to what degree (percentage) did the smallest cell size in the study correspond to the recommended cell size for detecting a difference of 1.00 standardized units with powers of .80 and .90, respectively, at a=.05? A value $100 indicates that the smallest cell size met or exceeded the respective .80 or .90 power requirements. -++Two dichotomous were created variables specifying whether levels of .80 or .90 were attained. Because many articles did not include sufficient information to replicate power computations, generic power tables were used. Note that the Kirk (1982) tables are targeted towards (quasi-)experimental studies and do not require variance estimates, are quite liberal, thus making it relatively easy for investigators to attain recommended sample sizes.

Interventions

Number of Interventions reported on in the article. If the design included a treated control/comparison group, this was counted as an intervention. If this group was untreated and thus did not receive any intervention of any kind (even not “routine care”), it was not counted as an intervention.

Type of Intervention: each intervention was initially coded as either physiological, psychological (intrapersonal), social (interpersonal, including psychoeducational), spiritual, organizational, financial, service utilization, professional, or other. Because of its distribution (see Results section for further details), this variable was used to reclassify studies as either physiological, psychosocial, or administrative intervention studies, or any combination thereof; and subsequently to reclassify studies as either physiological or psychosocial.

Outcomes

Number of Outcomes reported in the article. If a major outcome was broken down into suboutcomes, the major outcome was counted in addition to the suboutcomes when a separate statistical analysis was reported for it.

Type of Outcome: each outcome was coded using the same scheme used for type of intervention (see above). Because of its distribution (see Results section for further details), this variable was used to reclassify studies as either physiological, psychosocial, or administrative outcome studies, or any combination thereof; and subsequently to reclassify studies as either physiological, psychosocial, or combined outcome studies.

Findings

Publication Characteristics

As Table 1 shows, we reviewed a total of 2746 articles, of which only a small proportion (10.7% or 293) were intervention studies. Journals featured between 3.5% and 19.5% reports on intervention studies, for an average of 9.87% (*SD*=6.46). Thus, in the ten-year period reviewed, only one out of nine articles published in the six journals reported on a (quantitative) investigation of nursing interventions.

Intervention reports had between 1 and 9 authors (*M*=2.35, *SD*=1.52), with 39.2% papers authored by one and 22.5% by two. Four out of five (81.2%) papers had three authors or less. Only 45.1% of author teams ($2 authors) were multidisciplinary.

Subjects

Table 2 summarizes the characteristics of subjects studied. Most studies included in the sample (67.6%) focused on adults, yet note that we were unable to identify the predominant age group in 19 (6.9%) studies. Most studies (58.7%) were condition-specific, with the cardiovascular, reproductive, nervous, digestive, and respiratory systems accounting for 82.1% of these investigations. Only 29 articles (9.9%) excluded women, and 36 (12.4%) articles did not provide sufficient information to judge whether women were included. Remarkably, 203 (69.3%) articles were unclear as to whether minorities were involved, and only one out of four (73 or 24.9%) mentioned that minorities were included, either purposively or by sample availability.

Design Issues

One-factor designs were described in over two-thirds (68.9%) of all articles. Of these, one-factor designs with repeated measures were the most common design encountered (158 reports or 53.9%; see Table 3). Factorial designs (i.e, two or more factors) were used in only 14% of all studies reported. There were only 7 studies (2.4%) that employed blocking as a means of controlling one or more nuisance variables. All used simple blocked designs, and

no studies with true latin-squares or higher order blocked designs were identified (even though some authors incorrectly claimed higher order designs). Note that 20 (6.8%) studies featured a design that could not be classified according to established criteria. Most of these were unusual if not "hybrid" designs, often of questionable validity.
Table 4 reviews several aspects of the designs found in the sample of study reports. Designs featured between 1 and 20 groups, with a mean of 2.67 (*SD*=2.34). Sixty-four studies (21.8%) employed (weak) one-group designs, and 92.5% of studies had four or less groups. Control groups were found in 60.8% of studies. Random assignment was used in only 44.7% of studies. Most studies were longitudinal with at least two time points (82.1%), and only 52 studies (17.9%) employed (weak) posttest-only designs. Subjects were observed an average of 6.31 times (*SD*=11.07) over the course of a protocol; 192 longitudinal studies (65.5%) had between 2 and 7 time points.

Sixty-nine articles (23.5%) did not contain any direct or indirect information on attrition. Of the remaining 224 studies, 54.5% suffered attrition. Sample sizes left after attrition ranged from 0% to 99% of the original sample size (*M*=74.2%, *SD*=23.8, *Mdn*=83%).

Data analysis methods were overwhelmingly univariate (78.8%). Interestingly, an additional 17 studies (5.8%) made inferences about the effects of nursing interventions using only descriptive, noninferential statistics. No articles reported quantitative estimates of the effect of the interventions studied.

Statistical Power

Only 9.6% of all articles reported, made reference to, or implied a statistical power analysis, leaving 90.4% of studies without sample size justification. Forty-five articles (15.4%) did not report detailed cell sizes, making it impossible to perform power calculations per Kirk (1982).

As Table 5 shows, studies attained between 3% and 1447% of the minimum cell size needed to detect differences of 1.00 standardized units *C* with power of .80 (a=.05). This distribution was skewed (*Skewness*=4.84, *SE*=.15), very flat (*Kurtosis*=33.05, *SE*=.31), and highly variable (S^2=22636.58, *SD*=150.46). Even though the mean percentage of attainment exceeded 100 (*M*=130.04), the median was only 94. Using 100 as the cut-off, 51.6% were classified as not meeting .80 power requirements. There was no association between reporting (yes/no) and meeting .80 power requirements (yes/no) (c^2[1]=.430, *n.s.*).

As noted earlier, a power of .80 may not be sufficient for intervention studies, as conceptually these studies concern the alternative hypothesis more than the null hypothesis. When power calculations at the .90 level were performed, studies attained between 3% and 1070% of required cell sizes (*C*=1.00, a=.05). Again, this variable was distributed nonnormally (*Skewness*=4.81, *SE*=.15; *Kurtosis*=32.65, *SE*=.31) and with great variability (S^2=12409.68, *SD*=111.40), making the median of 70% more informative than the mean (*M*=97.25), even though neither reached 100%. Using 100 as the cut-off, over two-thirds (69.0%) of the studies did not meet .90 power requirements. Performing power calculations and meeting .90 power requirements were statistically independent (c^2[1]=.001, *n.s.*).

Interventions

Interventions described in articles ranged from 1 to 23 with a mean of 2.55 (*SD*=2.38); 67.9% of articles focused on only one or two interventions. On the basis of the coding scheme (see the section on Methods), studies were classified by intervention mix as either physiological, psychosocial (comprised of psychological, social, and spiritual), or administrative; or combined physiological and psychosocial, physiological and administrative, and psycho-social and administrative. No studies tested interventions that combined all three dimensions. Table 6 presents this classification. As can be seen from this Table, 82.9% of studies had exclusively physiological or psychosocial interventions. As the interest of this paper is predominantly clinical, analyses focused on similarities and differences between physiological and psychosocial intervention studies.

Of all physiological and psychosocial studies, only 10.3% referred to a power analysis, only 47.5% met .80 and only 29.9% met .90 power levels, 53.1% did not use random assignment methods, and 42.8% suffered attrition (23.9% had unknown attrition). These proportions are similar to those obtained for the sample at large. Physiological and psychosocial studies did not differ significantly on these methodological characteristics. Contingency analysis (2x2) revealed no associations between type of intervention study and the following design characteristics: reporting that a power analysis was performed (c^2[1]=2.12, *n.s.*), meeting requirements for power at the .80 (c^2[1]=.003, *n.s.*) and the .90 levels (c^2[1]=1.45, *n.s.*), use of random assignment methods (c^2[1]=3.02, *n.s.*), and presence of attrition (c^2[1]=3.29, *n.s.*).

Further, physiological and psychosocial studies did not differ on the mean number of interventions typically included (M_{phy}=2.81, *SD*=2.56; M_{psy}=2.25, *SD*=2.39, *t*[241]=1.75, *n.s.*), time points in the design (M_{phy}=8.21, *SD*=12.83; M_{psy}=5.68, *SD*=10.87, *t*[239]=1.65, *n.s.*), percent of original sample left after attrition (M_{phy}=70.89, *SD*=27.77; M_{psy}=75.78, *SD*=20.45, *t*[103]=!1.04, *n.s.*), percent of required .80 power level attained (M_{phy}=107.33, *SD*=94.78; M_{psy}=130.06, *SD*=120.41, *t*[202]=!1.47, *n.s.*), percent of required .90 power level attained (M_{phy}=80.30, *SD*=70.28; M_{psy}=97.16, *SD*=89.31, *t*[202]=!1.47, *n.s.*), and number of authors (M_{phy}=2.47, *SD*=1.45; M_{psy}=2.23, *SD*=1.54, *t*[241]=1.26, *n.s.*). However, psychosocial studies tended to focus on more outcomes (M_{phy}=4.64, *SD*=4.04; M_{psy}=7.00, *SD*=6.50, *t*[241]=!.40, *p*<.002) and adopt designs with more groups (M_{phy}=2.27, *SD*=1.91; M_{psy}=2.94, *SD*=2.40, *t*[241]=!2.41, *p*<.02).

Outcomes

Articles cited between 1 and 44 outcomes, with a median of 4 and a mean of 6.16 (*SD*=5.83); 68.6% of all articles related six or less outcome variables. As Table 7 shows, on the basis of the coding scheme (see Methods section), studies were classified by outcome mix as either solely physiological, psychosocial (comprised of psychological, social, and spiritual), or administrative; or combined physiological and psychosocial, physiological and administrative, psychosocial and administrative, or physiological and psychosocial and administrative. As 82.9% of studies had exclusively physiological, psychosocial, or a combination of physiological and psychosocial outcomes, analyses are focused on similarities and differences between these three types of outcome studies.

Of these three types of outcome studies, only 8.3% referred to a power analysis, 52.7% met .80 but only 29.1% met .90 power levels, 56.4% did not use random assignment methods, and 39.1% suffered attrition (25.5% had unknown attrition). These proportions are similar to those obtained for the sample at large and when studies are classified by intervention mix. Physiological, psychosocial, and combined outcome studies did not differ significantly on these methodological characteristics. Contingency analysis (3x2) found no associations between type of outcome study and the following design issues: reporting that a power analysis was performed (c^2[1]=2.50, *n.s.*), meeting requirements for power at the .80 (c^2[1]=2.09, *n.s.*) and the .90 levels (c^2[1]=5,17, *n.s.*), use of random assignment methods (c^2[1]=3.67, *n.s.*), and presence of attrition (c^2[1]=2.41, *n.s.*).

In addition, studies with physiological, psychological, or combined outcome variables did not differ on the mean number of interventions typically included (M_{phy}=2.92, *SD*=2.65; M_{psy}=2.04, *SD*=1.85; M_{com}=2.44, *SD*=2.61; *F*[2,240]=2.74, *n.s.*), groups included in the design (M_{phy}=2.34, *SD*=2.01; M_{psy}=2.57, *SD*=1.21; M_{com}=3.26, *SD*=3.43; *F*[2,240]=2.74, *n.s.*), percent of original sample left after attrition (M_{phy}=73.10, *SD*=28.67; M_{psy}=73.03, *SD*=24.50; M_{com}=75.59, *SD*=22.50; *F*[2,94]=1.00, *n.s.*), percent of required .80 power level attained (M_{phy}=109.02, *SD*=131.51; M_{psy}=131.51, *SD*=135.18; M_{com}=133.38, *SD*=106.09; *F*[2,200]=.78 *n.s.*) and percent of required .90 power level attained (M_{phy}=81.40, *SD*=105.50; M_{psy}=98.27, *SD*=100.37; M_{com}=99.76, *SD*=105.50; *F*[2,200]=.81, *n.s.*). However, ANOVAs with Scheffé tests for multiple comparisons revealed that physiological outcome studies observed subjects over more time points than psychosocial studies (M_{phy}=8.06, *SD*=12.37; M_{psy}=3.28, *SD*=6.51; M_{com}=6.34, *SD*=10.40; *F*[2,237]=4.28, *p*<.02).

Intervention and Outcome Mixes

When the intervention and outcome mixes (Tables 6 and 7) were crosstabulated, significant associations were found. The unreduced contingency analysis was significant (c^2[30]=157.56, *p*<.000001) with a weak association (*V*=.33, *p*<.000001) between intervention and outcome mixes. More relevant are the findings from the reduced clinical crosstabulation, the data for which are presented in Table 8. There was a significant association between clinical intervention and outcome mixes (c^2[2]=81.65, *p*<.000001), which was quite strong (*V*=.62, *p*<.000001). Physiological intervention studies focused predominantly on physiological outcomes (78.2%). In contrast, while 47.1% of psychosocial intervention studies focused exclusively on psychosocial outcomes, another 35.3% investigated combined outcomes. Note that 17.6% of these studies investigated the effects of psychosocial interventions solely on physiological outcomes.

Condition-Specific Investigations

Condition-specific and condition-nonspecific studies were remarkably similar in design and clinical focus (see also Table 9). These two types of studies did not differ on the mean number of interventions (M_{cs}=2.49, *SD*=2.61; M_{cns}=2.64, *SD*=2.02; *t*[291]=!.52, *n.s.*), outcomes (M_{cs}=6.05, *SD*=6.12; M_{cns}=6.31, *SD*=5.40; *t*[291]=!.39, *n.s.*), time points (M_{cs}=5.76, *SD*=9.95; M_{cns}=7.09, *SD*=12.46; *t*[288]=!1.01, *n.s.*) and groups (M_{cs}=2.70, *SD*=2.52; M_{cns}=2.64, *SD*=2.07; *t*[291]=.22, *n.s.*) included in the design, percent of requirements for power at the .80 level attained (M_{cs}=140.45, *SD*=176.83; M_{cns}=115.13, *SD*=100.42; *t*[246]=.22, *n.s.*), percent of requirements for power at the .90 level attained (M_{cs}=104.73, *SD*=130.82; M_{cns}=86.55, *SD*=74.71; *t*[246]=1.27, *n.s.*), and percent of sample size left after attrition (M_{cs}=73.01, *SD*=23.72; M_{cns}=76.22, *SD*=24.16; *t*[123]=!.72, *n.s.*). Further, contingency analysis (2x2) revealed that there was no association between condition-specific and condition-nonspecific studies, on the one hand, and the following characteristics, on the other hand: type of intervention (c^2[1]=1.52, *n.s.*) or outcome study (c^2[1]=1.28, *n.s.*), use of random assignment methods (c^2[1]=4.72, *n.s.*), integration of control group into the design (c^2[1]=2.50, *n.s.*), meeting .80 (c^2[1]=.006, *n.s.*) and .90 power requirements (c^2[1]=.030, *n.s.*), and presence of attrition (c^2[1]=2.47, *n.s.*).

Major Conditions

Table 9 summarizes the characteristics of studies focused on clinical conditions related to the cardiovascular, reproductive, nervous, digestive, and respiratory systems. Investigations across these organ systems were consistently alike in terms of mean number of interventions tested (*F*[4,136]=1.01, *n.s.*), number of outcomes studied (*F*[4,136]=.69, *n.s.*), groups featured in the design (*F*[4,136]=2.01, *n.s.*), percentage of .80 (*F*[4,116]=.19) and .90 power requirements met (*F*[4,116]=.19, *n.s.*), and percentage of sample size left after attrition ($F[4,65]=.32$, *n.s*). In studies involving the respiratory system, subjects were observed significantly more than in studies pertaining to the nervous and digestive systems ($F[3.54]$, $p<.01$). However, given that there were only 11 respiratory system investigations, of which three had more than 20 time points, this finding merits caution in interpretation.

Contingency analyses (5x2) revealed no associations between type of system involved and the following characteristics: intervention mix ($c^2[4]=6.11$, *n.s.*), use of random assignment methods ($c^2[4]=7.86$, *n.s.*), sample size estimation using power analysis ($c^2[4]=7.49$, *n.s.*), statistical methods used ($c^2[4]=12.65$, *n.s.*), attrition ($c^2[4]1=x.xx$, *n.s.*), and meeting .90 ($c^2[4]=5.99$, *n.s.*) power requirements. However, there were significant, yet nonsystematic, associations between major system involved and outcome mix ($c^2[4]=15.94$, $p<.05$), use of control groups ($c^2[4]=16.79$, $p<.003$) and whether studies met .80 power requirements ($c^2[4]=11.38$, $p<.03$).

Discussion

The question as to whether the discipline of nursing is committed to the empirical investigation of interventions to improve and maintain the health and well-being of patients and to prevent health problems from occurring or reoccurring, can only be answered equivocally from this present study. The fact that in the 1980s, arguably the decade in which nursing research may (or should) have experienced the greatest growth, only one out of nine articles published in six general research journals was concerned with testing the effects of interventions on patient outcomes, is of significant concern (we recognize that there may be a lag time between conduct of a study and publication of its findings). Moreover, the fact that the scientific and technical merit of investigations was often compromised due to consistent and recurrent methodological deficiencies, questions the authority of study findings in a field where intentional replication and extension are sporadic rather planned.

While we do not have data on preceding decades and therefore cannot examine growth, the paucity of intervention/ outcome studies in the past decade is of concern in itself. Certainly, much remains to be described, related, compared, interpreted, and understood about *person*, *environment*, and *health*. Descriptive studies, varying in focus, scope, and epistemology, will continue to contribute knowledge to the discipline. However, the fourth element, *nursing*, must be fully integrated as well to become the the pivotal element of knowledge development. This can only be achieved by fostering studies that investigate how the active or passive manipulation of (aspects of) *nursing*, applied to (some) healthy or ill *persons* and (parts of) their social and physical *environments*, impacts upon (aspects of) *health*.

Before discussing methodological and content issues, a note about authorship is in order. Most papers reporting intervention studies had single authors, and seldom did authorship exceed three. It would be inappropriate to draw conclusions about the nature of nursing research from authorship data. Yet, intuitively, these figures seem lower than what is typically seen in general and specialty research journals in other clinical life sciences.

Sample and Publication Bias

A limitation of the present study is that the sample was drawn from six general nursing research journals, when much nursing research is published in specialty nursing journals and in nonnursing, multidisciplinary periodicals. Notwithstanding this sampling bias, we can assume that investigators committed to advancing the discipline will publish in the discipline's leading journals with some regularity. Based on name and research recognition only, our impression from reviewing ten years of the chosen journals is that this is indeed the case. In addition, Downs

(1990) pointed out that many articles rejected by research journals are subsequently published in the so-called clinical journals. This is certainly not inappropriate, yet we concur with Downs (1990) that the format of publication needs to adhere to common principles of scientific reporting; specifically, methodological and conceptual information critical to verification, critique, and replication needs to be included.

Of greater concern is publication bias in the published intervention literature: the tendency, often pronounced, to favor "studies with positive or encouraging" or "significant and promising results" (Simes, 1987, p. 11). Intervention studies originate from a bias: the conceptually, empirically, or otherwise justified intellectual desire to document the presumed efficacy of some nursing interventions over other ones. Statistically significant results on most or all patient outcomes are believed to advance knowledge, thus equating new knowledge with statistically significant knowledge. In contrast, nonsignificant results tend to be "explained away" as being due to methodological or conceptual flaws. What is overlooked in this (erroneous) reasoning is that null results from well-conceptualized and well-designed studies with adequate statistical power advance knowledge just as much as do non-null results. Null results may not tell the discipline about new and effective nursing treatments, but they do reveal to the discipline new but ineffective interventions.

Not only does publication bias restrict what is known about effective and ineffective nursing interventions, it also prohibits comprehensive synthesis of research findings on nursing interventions. Narrative and integrative research reviews as well as quantitative research syntheses ("meta-analyses") cannot fairly integrate empirical knowledge if the sources of knowledge are biased towards novelty rather than innovation.

Methodology

Methodological strengths and weaknesses of studies published in the ten-year period surveyed were found to be consistent across all studies, and unrelated to either the intervention or outcome mix of studies, whether studies were condition-specific, and the type of major conditions they addressed. Such findings underscore the fact that the methodological qualities and shortcomings noted are not limited to certain types of interventions, outcomes, or patient populations, but instead are generalizable characteristics of clinical investigation in nursing.

Design

Most studies shared important methodological qualities. Studies tended to be longitudinal with at least two observations. In particular the designs with more than two time points provided an opportunity to go beyond merely assessing whether or not change occurred, and allowed for at least some examination of the process of change over time (Abraham & Neundorfer, 1990). Many studies were comparative, employed treated or untreated control groups, and thus were capable of assessing the differential effects of some nursing interventions over others. Designs accommodated mostly one or two factors. Coupled with the repeated measures implemented in many studies, these designs were appropriate for the experimental or quasi-experimental goals of the investigations.

On the down side, studies consistently presented with notable design flaws. Many authors made reference, explicitly or implicitly, to nuisance variables, yet few studies employed designs to block the influence of these variables. Random assignment is thought to disperse the potential effects of at least some of these variables, yet less than half of all studies were randomized investigations. Perhaps many investigators knew of potential threats to the internal validity of their studies, yet few attempted to modify their designs accordingly.

Statistical Power

Of great concern are the statistical power problems that plagued many studies. Conceptually, intervention studies are concerned more with the alternative hypothesis that the intervention(s) had an effect than with the null hypothesis of no such effect. Even though power of .80 has become a de facto standard, power of .90 and higher might be more appropriate for clinical investigations. That only a small proportion of studies did not mention statistical power analysis, performed either a priori to estimate needed sample size or post hoc to verify both statistically significant and nonsignificant results, is important in itself. Yet, most troublesome is that consistently about half of the studies did not attain the minimum cell size required for power of .80. With similar regularity about two-thirds failed to meet the requisites for power of .90. This means that the majority of studies contain a sizeable Type II or b error rate margin. Consequently, at the conceptual level, this leaves the discipline with inconclusiveness if not outright unclarity about the presumed effects of many nursing interventions.

Attrition

Related to statistical power is the issue of attrition. Attrition is common in clinical investigations and the finding that about half of the studies reviewed suffered attrition is not surprising. What is surprising, though, is that about one in four reports did not contain any information about subject loss or retention. Further, in many of the studies with subject loss, attrition was detected only because we closely examined subsample sizes, tables, degrees of freedom, etc.

Attrition affects the residual statistical power and the capability of a (longitudinal) study to detect differences between groups. The fact that the statistical power of the majority of studies was compromised underscores the need for oversampling. An attrition buffer of about 25% is suggested by our finding that, on average, 74.2% of original sample size was left at the completion of the study protocol.

A novel statistical approach to managing attrition is suggested in a recent paper by Efron and Feldman (1991) on compliance in drug effectiveness research, which can be extended to include the dichotomous situation of subject participation/loss. Based on compliance data from a clinical trial of a cholesterol-lowering drug (compliance being "the proportion of the intended dose actually taken" [p. 9] by subjects), these authors developed a regression model of compliance and response. This model was subsequently used to recover the true dose response curve, using in fact compliance as an explanatory variable.

Statistical Analysis

While most studies employed univariate models for statistical analysis when multivariate models might have been informative as well, the most significant statistical concern is the absence of effect size estimators. Without this information, all we really know is whether a given intervention did or did not have an effect. The size of this effect, which is potentially of great conceptual and clinical significance, remains unknown. At the worst, an intervention with small yet significant effect might end up being adopted as a clinical guideline or policy, when in fact the clinical efficacy is limited.

Relatedly, clinical investigations might benefit from statistical analysis methods that go beyond examining groups on the basis of means or other measures of central tendency. Variance effects, independently or in conjunction with mean effects, may reveal important information about how an intervention impacts on patients in ways that mean effects will not elucidate. Ultimately, through replication using variants of interventions, nursing studies should strive towards defining the nursing-equivalent of dose-response curves associated with certain interventions. Thus, it will be possible to achieve the goal of identifying, from the range of effective interventions within a class of interventions, the one with the most optimal patient outcome response in a given subpopulation of patients.

Populations Studied

During the period reviewed, intervention research in nursing was focused predominantly on adults. To their credit, nurse investigators, unlike many of their colleagues in the biomedical sciences, include women in their samples. This may be in part because of sample availability; yet many studies intentionally recruited women subjects. More importantly, nursing research critically concerns itself with women's health issues. It can only be hoped that the recent emphasis on and funding allocations pertaining to women's health within the NIH will strengthen and expand the discipline's leadership in care of healthy and ill women.

In contrast, nursing may not be much different from the other health sciences in its concern for the health of racial and ethnic minorities. The marked absence of any reference to minority representation in over two-thirds of intervention studies reviewed is not to say that in all instances minorities were excluded, or that all studies were focused on the proverbial white middle-class subject (even though, undoubtedly, both must have been the case in several studies). At the very least, the nonreporting of the racial composition of study samples shows that scientific sensitivity to differential epidemiologies is limited in nursing research. At the very worst, it may reflect an implicit discriminatory bias.

The distribution of condition-specific and condition-nonspecific studies stresses the discipline's commitment to both illness and health. However, the majority of studies were related to specific clinical conditions, which suggests that nursing and nursing care to a significant extent are illness-driven in their aims to restore, promote, or maintain health (in other words, nursing in part "treats" illness). This (at times unpopular) illness orientation provides, within the discipline, one of many platforms from which to pursue health; and, across disciplines, a common perspective for inquiry and care from divergent health and illness vantage points. Note also that the majority of clinical conditions specified in studies could be classified into five major categories of illnesses. This shows the resolution among nurse researchers to investigate interventions that are epidemiological warranted.

The relative paucity of investigations involving the elderly is of concern. Too often, late adulthood is seen as an extension of adulthood, and aging as the pathological deviation thereof. Not surprisingly, this has led to the often erroneous belief that what works with adults, works with the elderly. The fact that we face unprecedented demographic and epidemiological changes in the decades ahead, systematic attention to the development and testing of nursing interventions to restore or promote elder health is necessary, especially for selected high-risk subpopulations. Likewise, the limited number of clinical investigations on children is in discord with the needs of many children, in particular low birth weight infants, technology dependent children, children with chronic and/or life threatening illnesses, and poor or otherwise disadvantaged at risk children.

Interventions and Outcomes

Only a small proportion of articles in the journals reviewed were investigative reports of the effects of nursing interventions on clinical and administrative outcomes. Yet, it is encouraging to note that the majority of studies were clinical in nature and examined the effects of predominantly physiological and psychosocial nursing interventions on physiological, psychosocial, and combined patient outcomes. This not only attests to a strong interest amongst researchers in exploring new approaches to patient care; but also indicates a commitment to validating interventions that have become part of nursing through history, tradition, or trial-and-error.

The finding that most interventions studied were either physiological or psychosocial in nature exemplifies the two major strands of knowledge development in clinical nursing. Many studies within each type had only physiological or psychosocial outcome variables, respectively, which may indicate careful a priori consideration by researchers of conceptual and clinical relationships among independent and dependent variables. That many intervention studies of both types integrated physiological with psychosocial outcomes reflects an integration of nursing's holistic perspective in selected investigations.

On the downside, it must be noted that many studies had multiple outcome variables often quite indiscriminately so. This suggests a still common belief that its is better to measure too much than too little. Furthermore, having too many outcome variables can also make power analysis more difficult. Power analysis uses variance estimates for each of the outcome variables. Having more variables increases the likelihood of obtaining heterogeneous and extreme variance estimates. In turn, extreme estimates will yield more stringent sample size requirements.

As a team with varied nursing backgrounds, we also looked intuitively and nonsystematically at conceptual progress in our respective interest areas. A common observation was that quite a few studies were initiated without in-depth and critical consideration of prior findings and methodologies, often leading to conceptual repetitiveness and duplication of prior methodological errors.

Theory Development

While knowledge in nursing is acquired by research (and, to some extent, by practice), understanding is gained by theory (Walker, 1986). Although studies testing the effects of nursing interventions on patient outcomes may reveal effective and ineffective treatments, per se they do not necessarily help us understand how and why some interventions are beneficial and some not. Intervention studies constitute a platform from which theory can be generated, on the basis of which theory can be tested, and from which theories can be invoked to interpret and explain study findings. Some theories of relevance to nursing have been borrowed from other disciplines, others have been formulated from practical experiences, and still others have been created within conceptual models or paradigms. Perhaps most critical to understanding the findings of intervention/outcome research is the use of prevailing paradigms, as these facilitate the planning, implementation, and interpretation of science. In the absence of paradigm-linked research, knowledge will develop slowly and haphazardly with no clear indication of conclusiveness. Relatedly, the relevance of theory to the research in question will be compromised, and the contributions of research to the development and refinement of theory will be confined (Evers, 1989).

Implications for Reporting of Clinical Investigations

The findings of this study have implications for the reporting of intervention studies in nursing. Foremost, the difficulty experienced in coding many articles, because critical information was either absent or cryptic, calls for formulating new and revising existing publication standards. These standards should facilitate methodological verification, replication of studies by other investigations, and conceptual and clinical interpretation by scientists and consumers. Essential content includes: (a) power analysis, preferably a priori and post hoc, but at the least the latter; (b) estimation of the effect size associated with the interventions tested; (c) sampling criteria; (d) sample composition, not only in terms of epidemiological and gender strata, but also in terms of racial strata, and ethnic differentiation within different minority groups; (e) randomization, including justification for deviations from random assignment; and (f) justification of design, especially if allusion is made to possible nuisance variables.

The credibility of nursing research to the scientific community will be determined by the extent to which it fulfills its promise and potential as a clinical science committed to investigating the effects of clinical interventions on clinical outcomes in people with or without clinical conditions. The authority of nursing research within society at large will be defined by the extent to which the public recognizes how nursing care critically contributes to preventing illness, maintaining or improving health, decreasing suffering, affirming quality of life, and assuring dignity of death.

References

Abdellah, F. (1970). Overview of nursing research 1955-1968, Part 1. *Nursing Research*, *19*, 6-17.

Abraham, I. L., Nadzam, D. M., & Fitzpatrick, J. J. (Eds.) (1989). *Statistics and quantitative methods in nursing: Issues and strategies for research and education.* Philadelphia, PA: Saunders.

Abraham, I. L., & Neundorfer, M. M. (1990). The design and analysis of time series studies. In L. Moody (Ed.), *Advancing nursing science through research*, vol. 2 (pp. 143-172). Beverly Hills, CA: Sage.

Abraham, I. L., Schultz, S., Polis, N., Vines, S., & Smith, M. C. (1987). Research-on-research: The meta-analysis of nursing and health research. In M. Cahoon (Ed.), *Research methodology* (pp. 126-147) (Recent Advances in Nursing, vol. 17). Edinburgh: Churchill Livingstone.

Brown, J., Tanner, C., & Padrick, K. (1984). Nursing's search for scientific knowledge. *Nursing Research*, *33*, 26-32.

Downs, F. S. (1990). Accuracy counts, too. *Nursing Research*, *39*, 131.

Dunbar, S. (1981). Women's health and nursing research. *Advances in Nursing Science*, 1-16.

Ellis, R. (1977). Fallibilities, fragments and frames: Contemplation on 25 years of nursing research in medical-surgical nursing. *Nursing Research*, *26*, 177-182.

Efron, B., & Feldman, D. (1991). Compliance as an explanatory variable in clinical trials. (With discussion). *Journal of the American Statistical Association, 86*, 9-26.

Ellis, R. (1984). Personal communication. Cleveland, Ohio.

Evers, G. C. M. (1989). *Appraisal of self-care agency ASA-Scale*. Assen/Maastricht, The Netherlands: Van Gorcum.

Gortner, S., & Nahm, H. (1977). An overview of nursing research in the United States. *Nursing Research*, *16*, 10-33.

Gunter, L. & Miller, J. (1977). Toward a nursing gerontology. *Nursing Research*, *26*, 208-221.

Heater, B., Necker, A., & Olson, R. (1988). Nursing interventions and patient outcomes: A meta-analysis of studies. *Nursing Research*, *37*, 303-307.

Highriter, M. (1977). The status of community health nursing research. *Nursing Research*, *26*, 183-192.

Jacobson, B., & Meininger, J. (1985). The designs and methods of published nursing research: 1956-1983. *Nursing Research*, *34*, 306-312.

Jacobson, B., & Meininger, J. (1986). Randomized experiments in nursing: The quality of reporting. *Nursing Research*, *35*, 379-382.

Kirk, R. (1982). *Experimental design* (2nd ed.). Monterey, CA: Wadsworth.

Moody, L., Wilson, M., Smyth, K., Schwartz, R., Tittle, M., & Van Cott, M. (1988). Analysis of a decade in nursing practice research. *Nursing Research*, *37*, 374-379.

Moustafa, N. (1985). Nursing research from 1977 to 1981. *Western Journal of Nursing Research*, *7*, 349-356.

O'Connell, K., & Duffey, M. (1978). Research in nursing practice: Its present scope. In N. Chaska (ed.), *The nursing profession: Views through the mist* (pp. 161-174). New York: McGraw-Hill.

Simes, R. J. (1987). Confronting publication bias: A cohort design for meta-analysis. *Statistics in Medicine*, *6*, 11-29.

Walker, L. O. (1986). Theory, practice, and research in perspective. In L. H. Nicoll (Ed.), *Perspectives on nursing theory*. Boston: Little, Brown.