Trends in hypothesis testing and related variables in nursing research: a retrospective exploratory study

Aim To compare the inclusion and the influences of selected variables on hypothesis testing during the 1980s and 1990s.

Background In spite of the emphasis on conducting inquiry consistent with the tenets of logical positivism, there have been no studies investigating the frequency and patterns of hypothesis testing in nursing research.

Data sources The sample was obtained from the journal Nursing Research which was the research journal with the highest circulation during the study period under study. All quantitative studies published during the two decades including briefs and historical studies were included in the analyses.

Review methods A retrospective design was used to select the sample. Five years from the 1980s and 1990s each were randomly selected from the journal, Nursing Research. Of the 582 studies, 517 met inclusion criteria.

Discussion Findings suggest that there has been a decline in the use of hypothesis testing in the last decades of the 20th century. Further research is needed to identify the factors that influence the conduct of research with hypothesis testing.

Conclusion Hypothesis testing in nursing research showed a steady decline from the 1980s to 1990s. Research purposes of explanation, and prediction/control increased the likelihood of hypothesis testing.

Implications for practice Hypothesis testing strengthens the quality of the quantitative studies, increases the generality of findings and provides dependable knowledge. This is particularly true for quantitative studies that aim to explain, explain and predict/control phenomena and/or test theories. The findings also have implications for doctoral programmes, research preparation of nurse-investigators, and theory testing.

Keywords Hypothesis testing, nursing research, nursing science

Introduction In its early development, nursing embraced logical positivism as a means to develop its scientific base (Jortner 1980, Norbeck 1987, Beard 2002, Rodgers 2005). Drew (1988) stated logical positivism had dominated the academic world and that nursing was among the professions that embraced this view. Inherent in logical positivism has been the establishment of research hypotheses to test assumptions and theories. Rodgers (2005) stated that while logical positivism was fraught with high expectations, it provided an impetus of rigour and integrity in the pursuit of enquiry.

Although there has been a significant shift in the patterns of knowledge generation during the past two decades, empirical studies remain predominant (Garper 1978, Meleis 2007, Polit and Tatano Beck 2008. In the empirical-analytical model, hypothesis testing...
testing is one of the fundamental forms of analysis. Kerlinger (1973) perceived hypothesis testing as the most powerful tool invented to achieve dependable knowledge. While a research question rephrases the purpose of the enquiry, a hypothesis predicts the relationship between the variables and the outcome a priori.

In spite of the emphasis on conducting enquiry consistent with the tenets of logical positivism, there have been no studies investigating the trends in hypothesis testing in nursing research. Our observations suggested that although the empirical-analytical model in nursing research became widely accepted by the early 1970s, the use of explicit hypothesis testing has been sparse. Therefore, the rationale of conducting this study was to test the validity of our observations by exploring the trends in hypothesis testing in tandem with examining the influences of selected variables on hypothesis testing during the 1980s and 1990s.

Our research questions were:

- What is the frequency of the inclusion of hypothesis testing in nursing studies during the 1980s and 1990s?
- Are there differences in the inclusion and the number of hypotheses tested during the two decades?
- What is the relationship between research purposes, the first author's degree and the inclusion of hypotheses during this period?
- Is there a difference in research purposes and authors’ degrees during the two decades?

Nursing research has experienced significant growth since the 1970s. The views on ways of knowing have shifted since the 1970s (Garper 1978, Boyer 1990, Aydelotte 1992, Riley et al 2002, Meleis 2007), but early nurse-scholars brought the tenets of the empirical-analytical model to their research and teaching (Grace 1978). Ortner (1980, 2000) argued that the credibility of nursing science depended on the rigour and quality of scientific work and the generalisability of findings. Others (Disintainer 1986, Norbeck 1987, Goodard 1990, Rodgers 2005) saw nursing enquiry as the investigation of relationships and the testing of theories and competing hypotheses. Dublin (1969) and Roy and Roberts (1981) also perceived hypothesis testing as an essential tool in linking theory to the empirical world and guarding against random factors producing the results. Hence, the nursing literature of the 1970s and 1980s suggested the use of the empirical model, in which hypothesis testing was a fundamental step, was well established. Although hypothesis testing is often associated with theory testing and conceptual frameworks, scholars extend it to other types of research. Polit and Tatano Beck (2008) stated that hypotheses are not exclusively derived from theory and supported hypothesis testing for most of the quantitative studies.

We conducted a pilot study to investigate the trends in hypothesis testing from the 1980s to the 1990s by examining the articles published in Nursing Research. We analysed all quantitative studies published during 1985 and 1995. Results showed that 63 studies were published in 1985. Of these, 17 (27.2 per cent) included at least one hypothesis – 43 hypotheses in total. By contrast, 53 studies were published in 1995 and 11 of these (21 per cent) included hypotheses – 21 hypotheses in all. These numbers represented a 35 per cent change in the total number of quantitative studies with at least one hypothesis, and a 104 per cent change in the total number of hypotheses generated between these midpoint years. Hence, the pilot study findings suggested that hypothesis testing in nursing research was not only low, but may be decreasing.

Based on literature review, our observations, and the pilot study findings, we formulated the following hypotheses:

- There will be differences between the 1980s and 1990s in the number of studies that included hypothesis testing.
- There will be differences between the 1980s and 1990s in the total number of hypotheses tested.
- The study purpose and the first author's degree will influence whether or not a study will have hypothesis testing.
- There will be differences in the purposes of the studies and the first author's degree between the 1980s and 1990s.

The sample was obtained from the journal Nursing Research, published six times a year during the two decades under study (O’Heath et al 1990). Nursing Research was selected because it was the research journal with the highest circulation during this period and had 85 per cent of its publications dedicated to research.

Included in the sample were all quantitative studies published during the two decades including briefs and historical studies. A retrospective design with cluster sampling was used to obtain the sample from five years of each decade, using the table of random numbers (Table 1). The 1980s
Table 1 Number of articles reviewed by randomly selected year

<table>
<thead>
<tr>
<th>Decade</th>
<th>Randomly selected year</th>
<th>Number of studies reviewed</th>
</tr>
</thead>
<tbody>
<tr>
<td>1980s</td>
<td>1980</td>
<td>74</td>
</tr>
<tr>
<td></td>
<td>1982</td>
<td>71</td>
</tr>
<tr>
<td></td>
<td>1983</td>
<td>60</td>
</tr>
<tr>
<td></td>
<td>1985</td>
<td>65</td>
</tr>
<tr>
<td></td>
<td>1987</td>
<td>58</td>
</tr>
<tr>
<td>1990s</td>
<td>1990</td>
<td>56</td>
</tr>
<tr>
<td></td>
<td>1993</td>
<td>46</td>
</tr>
<tr>
<td></td>
<td>1994</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td>1995</td>
<td>53</td>
</tr>
<tr>
<td></td>
<td>1996</td>
<td>44</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2 Classification of levels and description of study purpose

<table>
<thead>
<tr>
<th>Level of purpose</th>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Descriptive</td>
<td>Observe, count, delineate, elucidate and classify phenomena; identify the characteristics and the frequency of phenomena.</td>
</tr>
<tr>
<td>2</td>
<td>Exploratory</td>
<td>Shed new light on manifestations of a phenomenon; identify factors causing phenomena; investigate the antecedents of phenomena; identify the effect of and the relationships between factors causing phenomena.</td>
</tr>
<tr>
<td>3</td>
<td>Explanatory</td>
<td>Explain the underpinning of the phenomenon; explain the systematic relationship among the factors causing phenomena; test models, theories, conceptual frameworks or their components; explain how and why phenomena work.</td>
</tr>
<tr>
<td>4</td>
<td>Prediction/control</td>
<td>Predict and control phenomena based on previous research; make phenomena occur or alter their natures.</td>
</tr>
</tbody>
</table>

Polit and Tatano Beck (2008) yielded 328 and 1990s yielded 254 studies, giving a total of 582 studies. Content analysis showed that 517 (88 per cent) of these studies met the criterion of being a quantitative study. The variables investigated in this study were:

- Study purpose.
- Number of hypotheses in each of the studies.
- The first author’s degree.

Polit and Tatano Beck (2008) state that all studies except descriptive studies should test a hypothesis, when appropriate. Hence, the purposes of the 582 studies were coded as per Table 2.

Inter-rater-reliability – testing the initial agreement between the raters regarding the identification of study purpose – yielded a significant kappa statistic (κ=0.2(516), p<0.01) showing an 84 per cent agreement. The variable ‘hypothesis testing’ was coded as ‘yes’ if it was explicitly stated or could be inferred that at least one hypothesis was being tested, and ‘no’ if it was neither stated nor inferable that this was the case.

Finally, the number of hypotheses in each study was coded as a continuous variable. We categorised the degree type held by the first author as nominal data: doctor of philosophy (PhD), doctor of nursing science (DNS/DNSc), doctor of education (EdD) and ‘other’.

The data were entered into Excel files and analysed using SPSS version 14. For all the hypotheses, the level of significance was set at .05. We used the chi-square test to analyse the differences between the observed and expected frequencies, Student’s t-test to compare the mean differences in continuous variables, the general linear model (GLM) to analyse the continuous response variables to the independent variables, and logistic regression analysis (LRA) to examine the interaction effects of the intervening variables on hypothesis testing. All analyses were conducted to ascertain the trends in and between the two decades.

Results

Of the 517 studies in the sample, 50 per cent were exploratory, 25 per cent descriptive, 13 per cent explanatory and 12 per cent were prediction/control. In the sample, 150 (29 per cent) studies included at least one hypothesis and there were 433 hypotheses in all. There was a slight decline in the number of quantitative studies: 273 in the 1980s versus 244 in the 1990s.

Further, the number of studies with at least one hypothesis declined strikingly each year, with a brief rise in 1990 (Figure 1). Of the 273 studies published during the 1980s, 98 (36 per cent) included at least
one hypothesis versus only 52 (21 per cent) of those published during the 1990s - an 88 per cent decline between the decades.

**Testing hypothesis one** First, we tested the effect of the decade on the frequency of hypothesis testing using the $\beta$LM with the following variables in the model: decade, study purpose and the first author’s degree. The results showed that decade had a statistically significant effect on the frequency of hypothesis testing - specifically, that articles published in the 1980s included more hypotheses than those published in the 1990s (Table 3). Study purpose also had a significant effect on whether a study included a hypothesis. By contrast, the first author’s degree had no influence. Hence, analysis showed that the 1980s had significantly more studies that included at least one hypothesis than the 1990s while taking the study purpose and the first author degree into account.

Since the analysis showed that study purpose had a significant effect, we then analysed which study purposes were associated with hypothesis testing, using the variables ‘decade’, ‘first author’s degree’ and ‘study purpose’ in four separate models. We found that the study purpose of ‘description’ was inversely related (the estimate was -2.31) to hypothesis testing, while exploratory studies had no effect. Further, the odds ratio estimates showed that purpose of ‘description’ decreased hypothesis testing by a factor of 0.1 or 10 per cent (Table 4).

In contrast, ‘explanatory’ and ‘prediction/ control’ studies were positive predictors of hypothesis testing. The odds ratio estimates showed explanatory studies increased hypothesis testing by a factor of 2.5, while prediction/control studies did so by a factor of 3.2, while still accounting for the effects of the decade and the first author’s degree. More importantly, in each of the four models, the significant effect of the decade persisted. Therefore, hypothesis one was accepted.

**Testing hypothesis two** Next, we addressed the effect of decade on the number of hypotheses tested. Descriptive analysis showed that 433 hypotheses were tested during the two decades by 517 studies. Of these, 313 were tested during the 1980s and 120 in the 1990s. The range of the hypotheses tested was 0-11, with a mean of 0.83, and a mode of 2. Figure 2 shows the number of hypotheses tested by decade.

Using the $\beta$LM with the number of hypotheses as a dependent variable, we tested decade, research purpose and first author’s degree in the model, first with all the 517 studies in the sample. The results showed the decade and the study purpose had significant positive effect, but the first author’s degree had no effect (Table 5, page 42).

We then analysed the effects of the same variables on hypothesis testing with only those 150 studies that included one or more hypothesis. The results showed that the decade still had a

**Figure 1** Trends in studies with at least one hypothesis during the 1980s and 1990s

**Table 3** Effects of decade, study purpose and first author’s education on hypothesis testing (n=517): logistic regression analysis

<table>
<thead>
<tr>
<th>Variable</th>
<th>$\chi^2$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decade</td>
<td>17.44(1)</td>
<td>&lt;.05*</td>
</tr>
<tr>
<td>Study purpose</td>
<td>45.37(3)</td>
<td>&lt;.05*</td>
</tr>
<tr>
<td>First author’s degree</td>
<td>0.95(3)</td>
<td>&lt;.05</td>
</tr>
</tbody>
</table>

* Statistically significant level

**Table 4** The relationship between hypothesis testing and study purpose (n=517): logistic regression analysis for analysis of effects

<table>
<thead>
<tr>
<th>Variable/purpose</th>
<th>$\chi^2$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description</td>
<td>32.52(1)</td>
<td>&lt;.05* (estimate -2.31)</td>
</tr>
<tr>
<td>Explore</td>
<td>0.75(1)</td>
<td>&gt;.05</td>
</tr>
<tr>
<td>Explain</td>
<td>10.30(1)</td>
<td>&lt;.05*</td>
</tr>
<tr>
<td>Prediction/control</td>
<td>17.19(1)</td>
<td>&lt;.05*</td>
</tr>
</tbody>
</table>

* Statistically significant level
significant effect on the number of hypotheses tested during the two decades. More significantly, it was the decade alone that predicted the number of hypotheses tested, with neither the study purpose nor the author degree having any effect (Table 6). Therefore, hypothesis two was also accepted.

**Testing hypothesis three** Hypothesis three was tested with two sub-hypotheses. The first test (3a) examined whether the study purpose influenced the inclusion of hypothesis testing. This hypothesis was accepted as the results of the testing of hypotheses one and two with this variable in statistical models. Hypothesis 3b tested whether the first author’s degree influenced the inclusion of hypothesis testing. This hypothesis was rejected as the results of the testing of hypotheses one and two with this variable in all models repeatedly showed the first author’s degree had no effect.

**Testing hypothesis four** Hypothesis four was tested with two sub-hypotheses. Hypothesis 4a stated that the level of the research purpose will differ between the two decades. Results showed that the level of the research purpose differed significantly across the two decades: \( \chi^2 = 25.82(3), p<.01 \). Descriptive studies remained essentially unchanged, while exploratory studies decreased by 43 per cent, and explanatory and prediction/control studies each increased by 17 per cent. Hypothesis 4a was accepted.

Hypothesis 4b stated that there would be differences between the two decades in the type of the degrees held by the first authors. Of the 517 studies, 61 per cent of first authors held PhDs, 6 per cent held EdDs, 11 per cent held DNS/DNScs and 22 per cent held other degrees. The testing of this hypothesis showed that there were significant differences between the two decades: \( \chi^2=38.36(3), p<.01 \). While about half (51 per cent) of first authors in the 1980s held PhDs, this increased to 71 per cent during the 1990s. Moreover, first authors holding other degrees declined in number by 18 per cent from the 1980s to the 1990s. Therefore, hypothesis 4b was accepted.

**Discussion**

Researchers frequently study publication productivity trends in their discipline (Farren 1991, Traynor *et al* 2001, Potempa *et al* 2008, Polit and Tatano Beck 2009). Similarly, we investigated 517 studies for the inclusion, the number and the influences of selected variables on hypothesis testing during the 1980s and 1990s. To our knowledge, this is the first study to investigate trends in hypothesis testing in tandem with variables related to hypothesis testing. However, the sample was obtained from a single journal, hence the interpretation of findings requires caution.

Despite its limitations, the analysis showed that there has been a significant decline in the number of studies that included at least one hypothesis as well as a decline from the 1980s to the 1990s in the number of hypotheses. The decline occurred in spite of a 17 per cent increase in the 1990s of those study purposes associated with hypothesis testing such as explanation and prediction/control. This finding corresponds well with the theoretical expectations of hypothesis testing being present in studies conducted for explanation and prediction/control.

The reasons for the decline in hypothesis testing in nursing research in the 1990s are unclear. Our
Data suggest that part of the decline may be due to a number of factors such as a decline in quantitative studies, increasing recognition of other ways of knowledge generation, and an increase in qualitative studies during the 1990s. The observed decline may also be related to graduate educational preparation and author characteristics (research experience or preference, and topic focus).

Also important is the finding that the type of degree held by the first author made no difference regarding whether a study would include hypothesis testing. This finding is particularly striking because the number of first authors with PhDs in the journal increased by about 20 per cent in the 1990s. Hence, nursing research not only showed a trend towards a decline in hypothesis testing, it did so in spite of an increase in the first authors who held PhDs – it is traditionally expected that the PhD specifically will prepare people for the rigours of traditional research (Lash 1987, Lash 1992, Marion et al 2003, Kim et al 2006, Anderson 2008).

A major strength of this study was its large sample size (n=517) and the investigation of the influences of related variables on hypothesis testing.

Limitations of the study
The major limitation of the study was the use of a single research journal for sample selection. Our sample is not representative of all the research studies published during the 1980s and 1990s. In this regard, the study shares the same sample bias of studies that use a single hospital, clinic or school. The selection of the journal, however, was based on its being the journal with the highest circulation during the 1980s and 1990s, and dedicating 85 per cent of its articles to research.

During the content analysis of our sample, we noted the high quality of these studies, some of which have had a strong effect on nursing care. The prevalence of hypothesis testing during the 1980s and 1990s, however, was our variable of interest and we limited our analyses to the changes in that variable.

Finally, we did not investigate the changes in the number of qualitative studies. This can offer a possible explanation for the decline in the number of quantitative studies and hence in hypothesis testing. However, we used the LRA to take into account the changes in the frequencies of related variables. Similarly, we recognise that not all factors that could influence hypothesis testing were investigated, as we did not explore the topics investigated (biological, social, educational) or the institutions with which the authors were associated.

The investigation was also limited to the studies published up to 2000. With the advance of the ‘evidence-based practice’ movement, studies published during the decade beginning in 2000 could exhibit entirely different trends in hypothesis testing. Hence, follow-up studies on this topic are needed.

Implications of findings
The use of hypothesis testing strengthens the quality of the quantitative studies, increases the generality of findings and provides the discipline of nursing with dependable knowledge. This is particularly true for those quantitative studies that aim to explore, explain and predict/control phenomena and/or test theories. The findings also have implications for doctoral programmes...
Conclusions

Despite the limitations, findings suggest that not only has the frequency of hypothesis testing in nursing research been low, but there has been a decline in its use in the last decades of the 20th century. Further research is needed to identify the factors that influence the conducting of research with hypothesis testing in multiple national and international journals. Similarly, doctoral dissertations and masters’ theses may be examined to ascertain if graduate programmes encourage hypothesis testing in student research.

The topic of our enquiry is new and follow-up studies are needed using multiple national and international journals. If replicated, our findings may have relevance for doctoral programmes and the discipline of nursing as a scientific field.

Online archive

For related information visit our online archive of more than 6,000 articles and search using the keywords.