



A case study of long-term Delphi accuracy

Rick Parente*, Janet Anderson-Parente

Towson University, Towson, Maryland, 21286, USA

ARTICLE INFO

Article history:

Received 15 January 2011
 Received in revised form 1 July 2011
 Accepted 3 July 2011
 Available online 6 August 2011

Keywords:

Delphi
 Accuracy
 Forecasting
 Prediction

ABSTRACT

We evaluated the long-term accuracy of predictions from a Delphic poll of the future of the mental health profession [1]. Six hundred participants provided predictions about the likely occurrence and probable time courses for 18 scenarios that could possibly have occurred over the 30-year interval between 1981 to present. Each of the panelists was polled twice with distributional feedback from the first poll provided on the second poll questionnaires. The data from the second polling was used to make predictions regarding the future of various issues that may have affected the mental health profession over the forecast interval. It has now been 30 years since the original poll was conducted; the purpose of this study was to evaluate the accuracy of the original group predictions. Results indicated that the Delphi panelists correctly predicted the occurrence of 14/18 scenarios. For those scenarios that did occur, the time course predictions were accurate within approximately 1–5 years.

© 2011 Elsevier Inc. All rights reserved.

According to Wikipedia, the on-line encyclopedia, “The **Delphi method** [pron: delfi] is a structured communication technique, originally developed as a systematic, interactive forecasting method which relies on a panel of experts” [2]. The Delphi technology involves extracting a consensus opinion from a panel of experts regarding the likelihood of future scenarios, probability judgments, or for developing likely scenarios. Consensus, in the form of decreased variability of panelists' responses, usually increases with iterative polling over two or more rounds; it is assumed that decreased variability of the panelists' predictions with iteration increases the accuracy of the group forecasts on the later rounds.

Usually a facilitator provides group consensus feedback regarding the various scenarios to the members of the group along with justifications for outlying opinions and then encourages the panel members to revise their predictions in light of this feedback. The procedure terminates after consensus is reached and the final group predictions are offered as an educated guess about the outcome of future events [3].

Although there are several versions of Delphi, [3–6] the basic technology involves four characteristics: Amalgamation of a group consensus opinion, iterative polling between rounds that include feedback, anonymity of participation, and statistical predictions of the probability of occurrence and likely time course for the scenarios. This combination of procedures usually produces predictions that are more accurate than chance or the average forecast from an individual group member [3]. Indeed, some published studies have shown that the structure of the Delphic polling procedure produces predictions that are more accurate than those obtained from unstructured groups such as group meetings, opinion polls, or focus groups [7,5].

Why does the Delphi Method produce accurate forecasts? Perhaps the most widely accepted explanation of the effect is the “Theory of Errors” proposed by Dalkey [8]. The basic assumption follows from the old adage; *two heads are better than one*. Extending this wisdom; *N heads are better than two*. In other words, forecasts (or decisions) that derive from the consensus of a structured group of individuals will be more accurate than those obtained from at least half of the group. Most groups provide a distribution of responses that may or may not contain the actual event when it occurs. However, when the true answer falls outside this range, either above or below it, then the median response of the group will be closer to the true answer than half of the group. When the true answer falls within the distribution of responses, the median response will usually be closer to the true

* Corresponding author.

E-mail address: rparente07@yahoo.com (R. Parente).

answer than most of the group. In general, regardless of where the true answer falls along the time line, the group consensus prediction will usually be more accurate than at least half of the individuals in the group. There is no guarantee, however, that the group median response will be more accurate than the best individual predictions.

The effect of providing feedback has a somewhat less obvious effect on accuracy. Feedback may move the response on later rounds in the direction of the true answer or towards the median response of the group that receives feedback during subsequent rounds. In most Delphi studies, providing feedback on subsequent rounds reduced variability of the panelists' responses. It was assumed that decreased variability was an index of consensus and that consensus was correlated with accuracy. Unfortunately, there have been few studies that have experimentally validated this assumption. Although some early research demonstrated that decreased variability of responding on iterated polls correlated with increased accuracy [9,10], later studies that separated out the effects of iterated polling with and without feedback did not find a consistent relationship between accuracy and variability [11,12].

Rowe and Wright [13] provided an excellent summary of Delphi accuracy in published studies up through 1999. Of the 21 published studies these authors reviewed, more than half of the studies reported accuracy levels after iterative polling that were higher than those obtained from an initial round of polling. Only two of the published studies reported the reverse effect. However, most of the early studies investigated accuracy of almanac predictions or short-term general interest events that could be evaluated within a constrained time frame, for example, everyday events like gas prices or election outcomes. Other studies have looked at accuracy of economic forecasts, weather predictions, or technological developments. Although most studies report improve accuracy of the Delphi Method relative to Face-To-Face discussions or more modern techniques like Prediction Markets [7], it is difficult to compare accuracy measures with such divergent measures.

Although there are very good summaries of the Delphi procedure, [12] there are very few studies of long-term predictive accuracy of the Delphi Method. Indeed, a Google search of published studies of Delphi accuracy that included the keywords: Delphi, Prediction, Forecast, Long-Term, and Accuracy did not return a single study that spanned more than 10 years. The lack of research in this area is probably due to the lengthy time frame necessary to validate the outcomes. However, it is important to study long-term accuracy of the Delphi Method because it may be the only means of forecasting technological breakthroughs or unique inventions that have no sequential history (e.g., manned landing on the moon) or geopolitical events that are difficult to predict (e.g., terrorist attacks). It is therefore important to determine if the technique actually produces accurate forecasts for these types of events. Although there are several studies that have documented short-term Delphi accuracy, [2] the data presented here are quite possibly the only study of Delphi accuracy that extends over a 30-year interval.

1. Description of the study

Anderson, Parente, and Gordon [1] used the Delphi Method to predict the future of the mental health profession over a 20-year time frame and beyond. Six hundred participants, 100 from each of six mental health professions participated. The professions included, Psychiatrists, Psychologists, Social Workers, Mental Health Counselors, Mental Health Administrators, and Psychiatric Nurses. The polling was done by mail and consisted of two rounds with median and distributional feedback provided to all participants on the second poll. The forecast questionnaire consisted of 18 scenarios that surveyed six different areas that were of concern to the mental health profession at the time: quality of graduate training; certification and licensure, accreditation practices; third-party reimbursement to practitioners at various training levels, employment prospects, and future trends. The authors purchased a random selection of members' mailing addresses from each professional association and these members served as the panelists. The 18 scenarios derived from research that was originally published by Asher [14].

2. Procedure

Participants were asked to decide first if each of the 18 scenarios would likely occur during the 20 year time frame from 1981 thru the year 2000 and beyond. A twenty year time frame was chosen because Asher. [14] noted that changes in the mental health profession were in a state of flux and it was therefore unlikely that predictions of those changes would be valid beyond the year 2000. If the participant thought that the event would occur then they were asked to provide a likely year along a timeline provided under each scenario. Therefore, each participant determined if the event was likely to occur. If they thought it would not occur during the 20 year time span, they were asked to respond by checking a box labeled "NEVER".

The 100 participants in each group received two polls. The second poll contained a summary of the responding for the 100 participants in that group. The feedback consisted of percentages under the predicted time courses for the various scenarios along with the percentage of respondents who had checked NEVER for each scenario. Every participant who responded to the second poll was thus aware of the consensus of his or her own peer group but not of that of the other groups. The participants were informed that the percentages were generated by individuals with professional training that was similar to their own. Although the participants were asked to consider the feedback, they were also told that the percentages should not be permitted to bias their responses. The 100 second-poll questionnaires from each group were also summarized as percentages which were used to derive predictions of the probability of occurrence (IF) and the time course (WHEN) for the various scenarios. A scenario was predicted to occur if a majority of the 600 participants endorsed it. For those scenarios that were predicted to occur, the median WHEN response was used to estimate the likely time of occurrence.

متن کامل مقاله

دریافت فوری ←

ISIArticles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات