SURVEYS · TYPES AND APPLICATIONS

Surveys of Experts versus General Populations: Differentiated Knowledge Sources

Surveys of experts and surveys of the general public answer different questions and fail in different ways. This article compares the two as sources of knowledge, covering sampling, question design, aggregation, bias and how to combine them.

ModQu research article · 3,007 words · 26 references (APA 7th edition)

When an organisation wants to know something it cannot observe directly, it often asks people. But which people? A pharmaceutical company may survey physicians about prescribing practice or survey patients about their experience of a condition. A government may ask economists whether a policy will raise growth or ask citizens whether they support it. A retailer may ask category managers where the market is heading or ask shoppers what they intend to buy. These are different kinds of survey, and they produce different kinds of knowledge.

A **general population survey** measures the characteristics, experiences, behaviour and opinions of the members of a population. Its authority comes from representativeness: a properly drawn and weighted sample stands for everyone. An **expert survey** asks a small number of specially qualified people to report facts they know, to make judgements or to forecast. Its authority comes from the respondents' knowledge. The first treats each respondent as a unit of the population to be counted; the second treats each respondent as an informant, a measuring instrument for something outside themselves.

This article compares the two designs, explains what each is good for, examines how each can go wrong and sets out how they can be combined to produce stronger knowledge than either alone.

Two different questions

The simplest way to tell the designs apart is to ask what the answer is about. In a general population survey, the object of measurement is the respondent: their age, income, purchases, satisfaction or voting intention. A respondent cannot be "wrong" about their own preference in the way an expert can be wrong about the future, although they can misreport it. The estimate is an aggregate of individual facts, and the main question is whether the sample represents the population (Groves et al., 2009).

In an expert survey, the object of measurement is usually outside the respondent: the position of a political party, the likely size of a market, the probability of a technical failure or the prevalence of a practice within an industry. The respondent's answer is a judgement, and it can be more or less accurate. Representativeness matters less than competence. Ten well-informed specialists may tell us more about a technical question than ten thousand members of the public, while those ten thousand are the only valid source for what the public itself thinks.

The distinction is not always sharp. A survey of physicians about their own prescribing is a general population survey of a professional population. A survey of physicians asking them to estimate how often a drug is misused in their region is an expert survey. The same people can be population members for one question and informants for another, and well-designed questionnaires keep the two roles separate.

Figure 1 summarises the main differences.

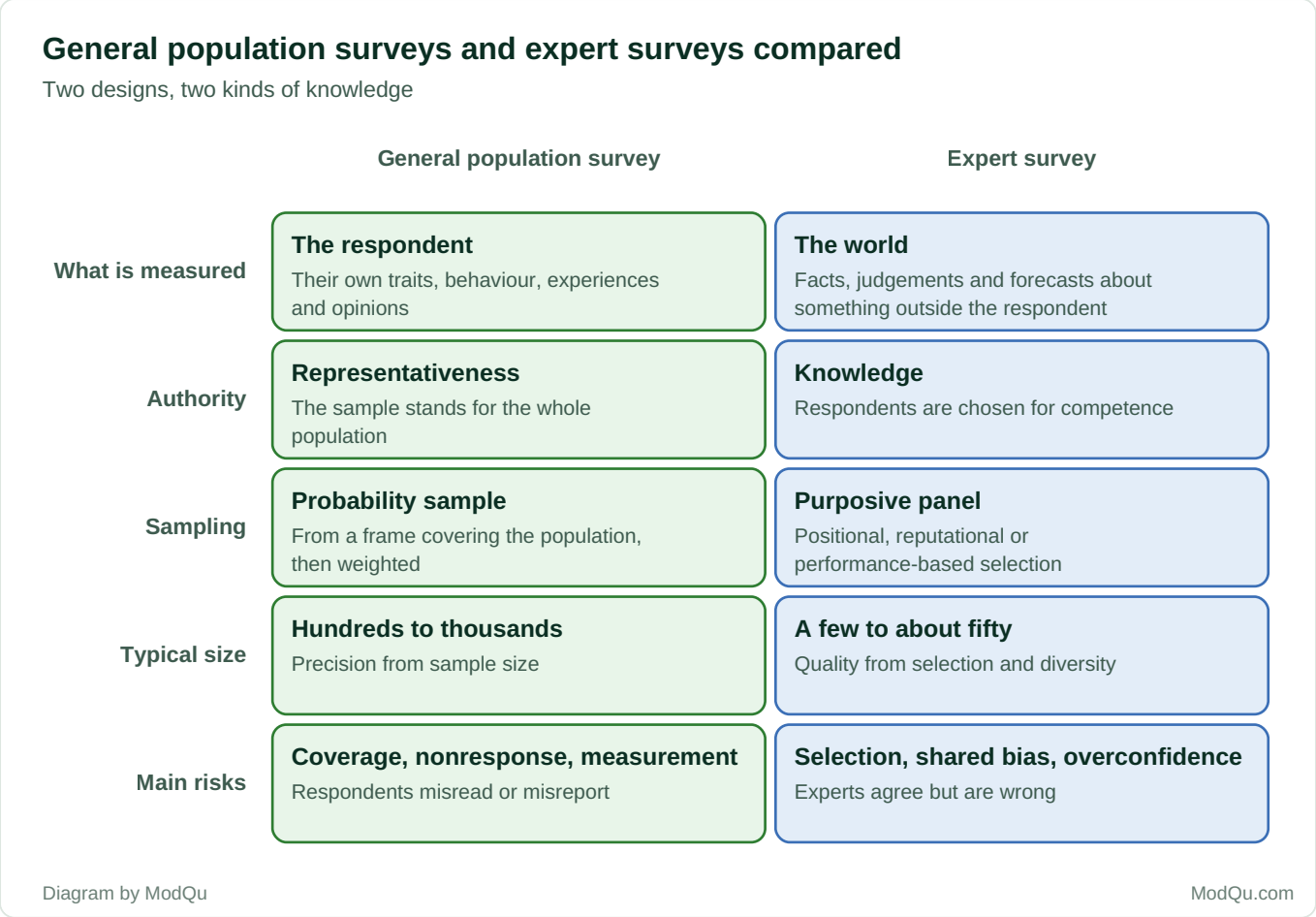


Figure 1. How general population surveys and expert surveys differ. Diagram by ModQu.

Sampling: populations versus panels

General population surveys

The theory of sampling a general population is well developed. A **sampling frame** lists or covers the population, a **probability sample** gives every member a known chance of selection, and design-based estimation allows the results to be generalised with a stated margin of error (Kish, 1965). The main threats are coverage error, nonresponse and measurement error. Because the respondents are interchangeable members of the population, no single person matters much. What matters is that the sample as a whole mirrors the population.

Response rates in general population surveys have fallen for decades, but a low response rate does not automatically mean a biased estimate. Groves and Peytcheva (2008) showed in a meta-analysis that the response rate was a weak predictor of **nonresponse bias**, which depends on whether the reasons for not responding are related to the variable being measured.

Expert surveys

Experts are rarely listed in a complete frame, and the definition of who counts as an expert is itself a design decision. Common approaches are:

- **Positional selection:** everyone holding a defined role, such as all chief financial officers in firms above a size threshold or all members of a professional body.
- **Reputational selection:** nomination by peers, often through snowball methods.
- **Credential or performance selection:** people with specified qualifications, publications or a track record of accurate judgement.

Each approach builds in assumptions. Positional selection may include people who hold a title but lack relevant knowledge. Reputational selection favours the visible and well connected. Performance selection is the most defensible where a track record exists but is rarely available.

Expert surveys are also small. A study of a specialist field may have 15 to 50 respondents, and sometimes fewer. Statistical generalisation to a population of experts is less important than ensuring that the panel covers the relevant range of knowledge and viewpoints. Good practice is to define the domain of expertise first, then select experts so that different schools of thought, regions or types of organisation are represented (Okoli & Pawlowski, 2004).

Experts and professionals are also hard to reach. They are busy, receive many survey requests and often have gatekeepers. Systematic reviews of surveys of physicians found that response rates could be improved by monetary incentives, especially those paid in advance, and by mixing contact modes (VanGeest et al., 2007). A later meta-analysis of surveys of health care professionals confirmed that incentives and the mode of contact both affected response (Cho et al., 2013). Similar principles apply to business executives and other professional groups: a clear purpose, a credible sponsor, a short questionnaire and a reason to care about the topic (Dillman et al., 2014).

Question design for informants

Questions for the general public must be understandable by people with varied education and no specialist vocabulary. Questions for experts can use technical language, but they face a different problem: the expert is asked to report knowledge or judgement about something they observe imperfectly.

Research on **key informants** in organisations shows the risks. Phillips (1981) found substantial measurement error when different informants in the same organisation described the same organisational characteristics, and argued that a single informant's report should not be treated as a perfect measure. Kumar et al. (1993) set out how to choose informants who

are knowledgeable about the specific issue, how to use several informants per organisation and how to combine their reports. Their advice applies to any survey that relies on informants: ask about matters the respondent can observe, ask several respondents where possible and check how much they agree.

Expert surveys often ask for quantities or probabilities. Asking for a single number hides uncertainty, so formal **expert elicitation** asks for ranges or full probability distributions, for example the 5th, 50th and 95th percentiles of a quantity (Morgan & Henrion, 1990). O'Hagan et al. (2006) describe how to structure such questions, how to train experts briefly before they answer and how to check the coherence of their judgements. These techniques produce richer and more honest information than a simple "best guess" question.

Bogner et al. (2009) distinguish three kinds of knowledge that experts can provide: technical knowledge of facts and procedures, process knowledge of how things work in practice, and interpretive knowledge of how situations should be understood. Questionnaires should be clear about which kind they are seeking, because experts are often more reliable on the first than on the third.

Aggregation: crowds and panels

How individual answers are combined differs sharply between the two designs.

In a general population survey, aggregation is a matter of estimation: a weighted mean or proportion estimates the population value. Every respondent's answer counts according to their weight, and disagreement between respondents is information about the population, not error.

In an expert survey, disagreement is a sign of uncertainty about the thing being judged. The question becomes how best to combine judgements into a single estimate. The simplest method, averaging, is often surprisingly effective. Galton (1907) reported that the median of nearly 800 guesses of an ox's weight at a country fair was within about one per cent of the true weight. Surowiecki (2004) popularised this as the "wisdom of crowds", arguing that the average of many independent, diverse judgements can be more accurate than most individuals.

Averaging works because individual errors partly cancel. Clemen (1989) reviewed a large literature on combining forecasts and found that simple averages of forecasts were usually more accurate than the individual forecasts, and often as good as more complex weighting schemes. Larrick and Soll (2006) showed that people tend to underestimate the value of averaging and prefer to pick the opinion of whoever seems most knowledgeable, which often reduces accuracy.

Averaging has a limit. It cancels independent errors, but it cannot remove a bias that all the judges share. Figure 2 illustrates this with a simple model. As more judges are added, the error of the average falls, but it levels off at the size of the shared bias. A crowd of non-experts with a large shared bias may never become as accurate as a smaller group of experts whose shared bias is small. Equally, a group of experts trained in the same way may share blind spots that diversity would correct.

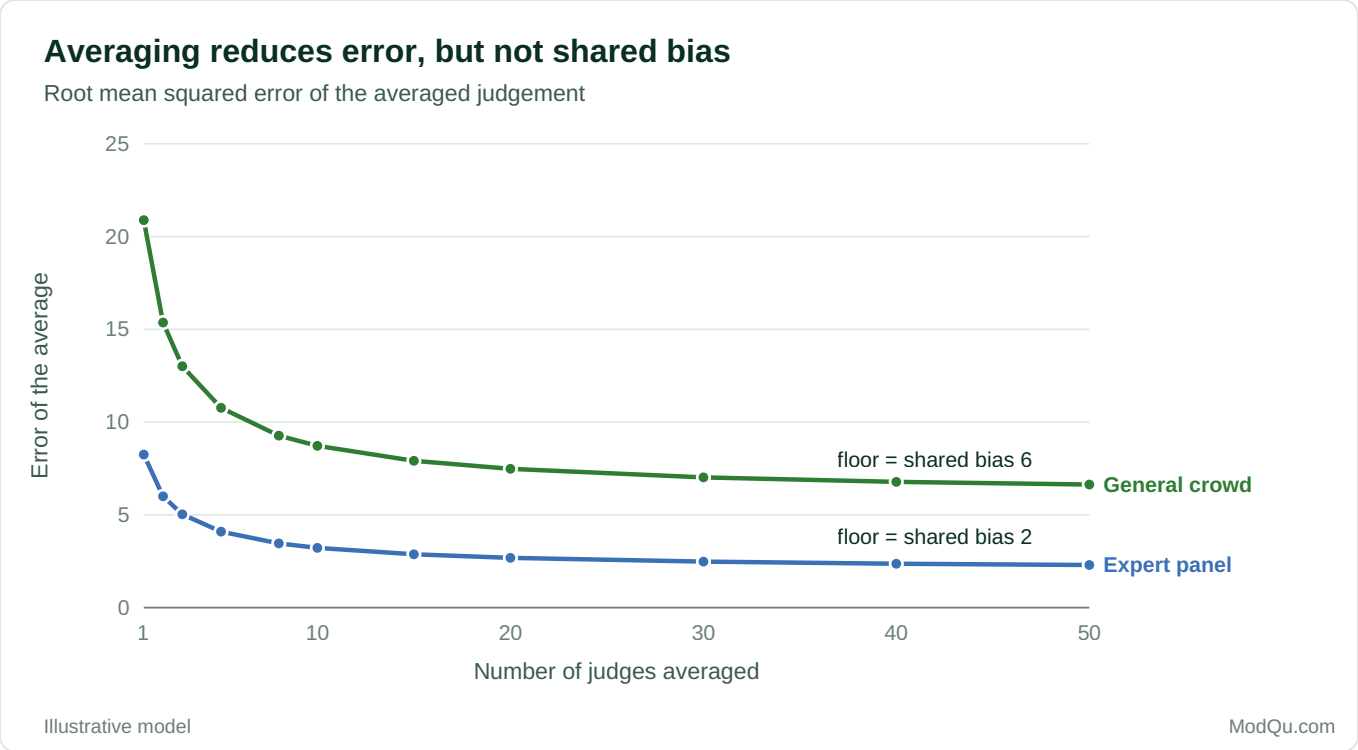


Figure 2. Error of an averaged judgement as judges are added. Root mean squared error equals the square root of shared bias squared plus individual error variance divided by the number of judges. Crowd: individual error 20, shared bias 6; experts: individual error 8, shared bias 2. Illustrative model; chart by ModQu.

Better methods weight judges by their demonstrated skill. Budescu and Chen (2015) proposed identifying the members of a crowd whose judgements consistently improve the aggregate and giving them more weight, which improved accuracy over a simple average. In forecasting tournaments, Mellers et al. (2014) found that training in probabilistic reasoning, working in teams and tracking individual accuracy all improved forecasts.

Structured group methods

The **Delphi method** is a structured way of surveying experts in several rounds. It was developed at the RAND Corporation, and Dalkey and Helmer (1963) described an early application. Experts answer anonymously, receive a statistical summary of the group's answers and their reasons, and then revise their own answers. The process usually continues for two to four rounds. Anonymity reduces the influence of dominant personalities, and feedback allows experts to learn from each other without face-to-face pressure (Linstone & Turoff, 1975).

Rowe and Wright (1999) reviewed evaluations of Delphi and found that it generally produced more accurate judgements than unstructured groups and than the first-round average, although the evidence was mixed and depended on how the method was carried out. The convergence that Delphi produces is not always a sign of improved accuracy; it may reflect conformity to the group summary. Reporting both first-round and final distributions makes this visible.

How good are experts?

Expertise is valuable but not uniformly so. Tetlock (2005) followed several hundred experts making thousands of predictions about political and economic events over many years. On average, their forecasts were only slightly better than simple statistical rules and, for long-range questions, barely better than chance. Experts who explained the world with a single big idea did worse than those who drew on many sources and were willing to revise their views.

Kahneman and Klein (2009) set out when expert intuition can be trusted. Intuitive expertise develops when the environment is regular enough to be predictable and the expert has had many opportunities to learn with rapid, clear feedback. Fire-fighters, chess players and experienced nurses meet these conditions in many situations. Stock pickers and long-range political forecasters usually do not. An expert survey therefore deserves more trust when it asks about matters within the experts' direct, repeated experience and less trust when it asks them to forecast irregular events.

Experts can also be overconfident. Calibration studies ask people for intervals they are, say, 90 per cent sure contain the true value; experts' intervals often contain the true value far less often than claimed. Formal elicitation and training reduce this tendency (O'Hagan et al., 2006).

Expert surveys in political science illustrate both strengths and checks. Researchers use them to locate political parties on ideological scales. Steenbergen and Marks (2007) evaluated such judgements by comparing them with other sources, including party manifestos, voter perceptions and the views of party elites, and found that expert placements were generally consistent with them. The validation strategy, comparing expert judgements with independent measures, is a model for other fields.

Where experts and the public disagree

Experts and members of the public often reach different answers to what appears to be the same question. Slovic (1987) showed that lay judgements of risk are shaped by qualities such as dread, unfamiliarity and lack of control, while experts' judgements track expected fatalities more closely. Neither view is simply "right": the public's judgements reflect values and concerns that technical estimates leave out.

Sapienza and Zingales (2013) compared responses of leading academic economists with those of a general population sample on the same economic questions. The two groups differed widely on many items, and the public's answers changed little when told what experts thought. The study shows that expert and public surveys measure different things: one measures professional judgement, the other public belief, and both matter for decisions that depend on public support.

For businesses the same contrast appears between what industry specialists expect and what customers want. Analysts may judge a product feature to be technically superior, while customers may not value it. A plan built only on expert opinion risks missing demand; a plan built only on customer surveys risks missing what is technically possible or coming soon.

Combining the two sources

The two designs are complementary. Figure 1 shows that each is strongest where the other is weakest. Practical ways to combine them include:

1. **Experts to frame, public to measure.** Use expert interviews or a small expert survey to identify the issues, options and vocabulary, then measure their prevalence and importance in a general population survey.
2. **Public to describe, experts to interpret.** Use a population survey to establish facts about behaviour or opinion, then ask experts to explain the patterns and judge their implications.
3. **Parallel questions.** Ask both groups the same question, as Sapienza and Zingales (2013) did, and treat differences as a finding in their own right.
4. **Expert estimates for what cannot be measured.** Where a population survey cannot measure something, such as future market size, expert elicitation fills the gap, with uncertainty reported explicitly.

In each case the report should state clearly which conclusions rest on which source. Readers should know whether a figure is an estimate of what customers say or a judgement by specialists about what will happen.

A worked example

Consider a hypothetical manufacturer of home heating equipment deciding whether to invest in a new heat-pump product line. It needs to know three things: how large the market will be in five years, what features buyers value and how installers will respond.

The first question concerns the future and depends on technology costs, regulation and energy prices. No member of the public can report it. The firm commissions an expert elicitation with 18 specialists drawn from equipment manufacturing, energy policy, utilities and installation. Each gives a median and a 90 per cent range for market size. Two Delphi rounds narrow the spread somewhat, and the final report shows both the median of the medians and the range of views, with a note that most experts share the assumption that current subsidies continue.

The second question concerns buyers' preferences, and only buyers can answer it. A general population survey of homeowners, drawn from an address-based frame and weighted to known population totals, measures awareness, purchase intention and the importance of cost, comfort and running costs.

The third question falls between the two. Installers are both a professional population, whose own business practices can be counted, and informants about customers they meet every day. The firm surveys a sample of installers, asking separately about their own practices and about what they observe among customers.

Each source answers a different question, and the conclusions are stronger because the firm did not ask any group to answer a question it was poorly placed to answer.

Practical guidance

For a general population survey:

- Define the target population and choose a frame that covers it.
- Use probability sampling where the decision requires precise population estimates.
- Write questions that ordinary respondents can understand and answer.
- Weight the data and report sampling error and response rates.

For an expert survey:

- Define the expertise required for the question and select experts against that definition.
- Seek diversity in background and viewpoint as well as depth of knowledge.
- Ask for ranges or probabilities rather than single numbers.
- Ask about matters within the experts' direct experience where possible.
- Use structured methods, such as Delphi, for judgements that benefit from iteration, and report the spread of views as well as the central estimate.
- Where possible, check expert judgements against independent data or later outcomes.

For both, document who was asked, why, how their answers were combined and what the results can and cannot support.

Common mistakes

- **Treating an expert panel as a population sample.** Twenty experts' views are not a statistical estimate of "what the industry thinks" unless the panel was drawn to represent the industry.
- **Treating public opinion as expert knowledge.** Public estimates of technical quantities, such as the size of a risk, measure perceptions, not facts.
- **Overweighting single confident voices.** Prominent or senior experts are not necessarily the most accurate.
- **Presenting consensus as certainty.** Agreement among experts who share the same training or information may reflect a shared bias.
- **Ignoring the role of each respondent.** Mixing questions about the respondent's own behaviour with questions asking them to judge others, without clear separation, makes the data hard to interpret.

Key takeaways

General population surveys and expert surveys are different instruments. The first counts and describes; its strength is representativeness, and its main risks are coverage, nonresponse and measurement error. The second consults informed judgement; its strength is knowledge, and its main risks are selection, shared bias and overconfidence. The best evidence often combines them, using each for the questions it answers best and making clear which conclusions come from which source.

References

Bogner, A., Littig, B., & Menz, W. (Eds.). (2009). *Interviewing experts*. Palgrave Macmillan. https://doi.org/10.1057/9780230244276

Budescu, D. V., & Chen, E. (2015). Identifying expertise to extract the wisdom of crowds. *Management Science*, *61*(2), 267–280. https://doi.org/10.1287/mnsc.2014.1909

Cho, Y. I., Johnson, T. P., & VanGeest, J. B. (2013). Enhancing surveys of health care professionals: A meta-analysis of techniques to improve response. *Evaluation & the Health Professions*, *36*(3), 382–407. https://doi.org/10.1177/0163278713496425

Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting*, *5*(4), 559–583. https://doi.org/10.1016/0169-2070(89)90012-5

Dalkey, N., & Helmer, O. (1963). An experimental application of the Delphi method to the use of experts. *Management Science*, *9*(3), 458–467. https://doi.org/10.1287/mnsc.9.3.458

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). *Internet, phone, mail, and mixed-mode surveys: The tailored design method* (4th ed.). Wiley.

Galton, F. (1907). Vox populi. *Nature*, *75*(1949), 450–451. https://doi.org/10.1038/075450a0

Groves, R. M., & Peytcheva, E. (2008). The impact of nonresponse rates on nonresponse bias: A meta-analysis. *Public Opinion Quarterly*, *72*(2), 167–189. https://doi.org/10.1093/poq/nfn011

Groves, R. M., Fowler, F. J., Jr., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey methodology* (2nd ed.). Wiley.

Kahneman, D., & Klein, G. (2009). Conditions for intuitive expertise: A failure to disagree. *American Psychologist*, *64*(6), 515–526. https://doi.org/10.1037/a0016755

Kish, L. (1965). *Survey sampling*. Wiley.

Kumar, N., Stern, L. W., & Anderson, J. C. (1993). Conducting interorganizational research using key informants. *Academy of Management Journal*, *36*(6), 1633–1651. https://doi.org/10.2307/256824

Larrick, R. P., & Soll, J. B. (2006). Intuitions about combining opinions: Misappreciation of the averaging principle. *Management Science*, *52*(1), 111–127. https://doi.org/10.1287/mnsc.1050.0459

Linstone, H. A., & Turoff, M. (Eds.). (1975). *The Delphi method: Techniques and applications*. Addison-Wesley.

Mellers, B., Ungar, L., Baron, J., Ramos, J., Gurcay, B., Fincher, K., Scott, S. E., Moore, D., Atanasov, P., Swift, S. A., Murray, T., Stone, E., & Tetlock, P. E. (2014). Psychological strategies for winning a geopolitical forecasting tournament. *Psychological Science, 25*(5), 1106–1115. https://doi.org/10.1177/0956797614524255

Morgan, M. G., & Henrion, M. (1990). *Uncertainty: A guide to dealing with uncertainty in quantitative risk and policy analysis*. Cambridge University Press. https://doi.org/10.1017/CBO9780511840609

O'Hagan, A., Buck, C. E., Daneshkhah, A., Eiser, J. R., Garthwaite, P. H., Jenkinson, D. J., Oakley, J. E., & Rakow, T. (2006). *Uncertain judgements: Eliciting experts' probabilities*. Wiley. https://doi.org/10.1002/0470033312

Okoli, C., & Pawlowski, S. D. (2004). The Delphi method as a research tool: An example, design considerations and applications. *Information & Management, 42*(1), 15–29. https://doi.org/10.1016/j.im.2003.11.002

Phillips, L. W. (1981). Assessing measurement error in key informant reports: A methodological note on organizational analysis in marketing. *Journal of Marketing Research, 18*(4), 395–415. https://doi.org/10.1177/002224378101800401

Rowe, G., & Wright, G. (1999). The Delphi technique as a forecasting tool: Issues and analysis. *International Journal of Forecasting, 15*(4), 353–375. https://doi.org/10.1016/S0169-2070(99)00018-7

Sapienza, P., & Zingales, L. (2013). Economic experts versus average Americans. *American Economic Review, 103*(3), 636–642. https://doi.org/10.1257/aer.103.3.636

Slovic, P. (1987). Perception of risk. *Science, 236*(4799), 280–285. https://doi.org/10.1126/science.3563507

Steenbergen, M. R., & Marks, G. (2007). Evaluating expert judgments. *European Journal of Political Research, 46*(3), 347–366. https://doi.org/10.1111/j.1475-6765.2007.00694.x

Surowiecki, J. (2004). *The wisdom of crowds*. Doubleday.

Tetlock, P. E. (2005). *Expert political judgment: How good is it? How can we know?* Princeton University Press.

VanGeest, J. B., Johnson, T. P., & Welch, V. L. (2007). Methodologies for improving response rates in surveys of physicians: A systematic review. *Evaluation & the Health Professions, 30*(4), 303–321. https://doi.org/10.1177/0163278707307899