Social science surveys and political polls are valuable tools. They can help us (1) understand society; (2) craft and evaluate public policies to address social issues; (3) establish better linkages between the people and the government; and (4) enhance the transparency and legitimacy of elections.

But, be aware—survey and polls can also be tools of unfounded persuasion and misinformation.

The solution is not to dismiss all polls as biased and misleading, or to accept all as the unvarnished truth. Rather, it is to ask the right questions, understand the limitations of polls and value and use the information provided by polls appropriately.

Here are six questions everyone should ask about surveys in order to use them more effectively:

1. **How was the survey conducted? Was it face to face, or over the phone? In public places or at households?**

   The *survey mode* and interviewer traits can affect the results. Many surveys are conducted face-to-face, but they are also increasingly conducted by phone and the Internet. The mode has implications for the extent to which social desirability can affect the results. Sometimes people want to hide views they believe goes against social norms, an effect we call *social desirability*. Many people may be tempted to say that they voted when they did not, that they practice their religion when they do not, and so on. Behaviors like using drugs may be of interest to public health officials, but people will under-report undesirable or illegal behavior. Face-to-face surveys—particularly those conducted in public places—are highly susceptible to social desirability.

   More private modes such as paper and pencil questionnaires may reduce the under-reporting of these behaviors. On the other hand, because trust can be established between an interviewer and respondent, face-to-face interview can improve the accuracy of reporting over situations where, for instance, people believe that phones have been tapped or that written surveys can be attributed to them. The important point is to know what type of survey research was conducted and to think about how this may affect the results.

   So too, you want to think about who conducted the survey. Conformity bias can occur when the respondent changes his or her answer—perhaps even unconsciously—to conform to the views he or she attributes to an interviewer, based on his or her observable qualities. For example, participants may report more feminist attitudes to women than to men or higher support for a secular party to an interviewer who is not dressed in overtly religious clothing. Participants might know that the survey firm is affiliated with a certain party and change their answers. This is one possible explanation for 2006 Palestinian Center for Policy and Survey polls which suggested Fateh was the favored party in legislative elections, although HAMAS later won.
2. Who was included in the survey? How was sampling conducted? How representative is the sample?

We often read survey results that say, “70% of Tunisians believe…” or “25% of women think…” The underlying notion is that the people asked in the survey (the sample) are representative of Tunisians, women or other groups. But, how representative is the sample? And how can we tell whether the 1200 Tunisians in the sample, for instance, are likely similar to the full Tunisian population?

It is sometimes tempting to dismiss surveys altogether; after all, how can such a small sample represent the whole Tunisian population? Others are more likely to accept the results unconditionally: If the report says it represents Tunisians, then it does.

We suggest you neither dismiss nor accept results unconditionally. Instead, look at the way the sample was chosen to decide how likely the sample reflects the attitudes of the full population.

Scientific (also called probabilistic) sampling is a technique that allows survey researchers to measure attitudes and behaviors in a sample and generalize about those attitudes and behaviors in the population. In a probabilistic, or scientific, sample, every member of the population has a known and non-zero probability of selection. That is, if the sample is to represent the full Tunisian population, then both men and women, young and old, should be able to be in the sample. Conducting the survey in boys’ schools—where only young boys are present—may tell you something about young boys, but it would not be correct to say that the results tell you what the full population believes.

Two common ways to draw a probabilistic sample are:

(a) A simple random sample: Obtain a list of every member of the population, such as every student in the university. Draw a random sample of the list.

(b) Use Probability Proportional to Size (PPS) sampling at the household level (see: http://www.rhrc.org/resources/general_fieldtools/toolkit/55b%20PPS%20sampling%20technique.doc). In this case, the researcher randomly selects Primary Sampling Units (PSUs) and randomly select secondary units and so on within the PSUs. At the household level, the interviewer then uses Kish Tables to randomly select a participant within the household. This method is a type of multi-stage sampling.

A third type of sampling, frequently used, is called quota sampling. In this case, the researchers decide what percentage of the population is young or old, men or women, for instance, and send interviewers to survey individuals who match set characteristics. This may result in reasonably representative sample, but it is not a probabilistic sample.

So, the first question to ask is: how was sampling done? Look for probabilistic sampling, where the population that the sample is drawn from matches the one that the reported in the results.
3. **How much confidence do we have in the results reported? That is, how large is the margin of error?**

When we see reports that “70 percent of Tunisians (or Algerians, Libyans or Egyptians) believe they will vote in the next election”, we know this is the percentage of respondents in the survey—not of all Tunisians—who said they would vote, for instance, a certain way. To what extent can we believe this represents the willingness in the full population?

To think about this, it is first important to note that there are two kinds of errors that affect whether the survey or polling results accurately reflect the true beliefs of the population: sampling error and non-sampling errors.

**Sampling Error.** Most surveys and polls state something like, “The results of this survey are accurate at the 95% confidence level plus or minus 3 percentage points.” Sampling error is the error from surveying a sample of the population rather than the entire population. Sampling error is unavoidable because, by chance, any random sample drawn could differ slightly from the population. On average, the estimate (e.g. the proportion of the population that will vote for the Green Party) in the sample will be accurate for the population. But, a sample could, by chance alone, slightly over-estimate or under-estimate the popularity of a party in the population. The most important factor affecting the size of the margin of error is sample size. The larger the sample size, the smaller the margin of error (also called sampling error) will be.

**Non-Sampling Errors.** There are several types of non-sampling errors: (1) measurement error (e.g. social desirability bias); (2) coverage error; (3) non-participation error; and (4) non-response error. These errors cannot be measured and they can seriously bias the results of a poll.

Because such errors are possible, you want to ask researchers how they addressed them. Earlier, we discussed some common sources of measurement error: social desirability and conformity bias. Now let us discuss some other forms of non-sampling errors.

4. **The fourth question to ask is: If different groups vary in their probability of being selected, are weights used in reporting the results?**

Sometimes the sampling approach yields data that does not fully represent the population. In a representative sample, all members of the population have a known and non-zero probability of being selected. Yet, if the surveyors use a mode that excludes some members of the population, the results may be biased. For example, if a phone survey is used, members of the population who do not have a phone have a zero chance of being included in the sample. This is called **coverage error,** and it can bias the results if the error is related to the variable being measured. For example, if wealthier people are more likely to have a phone and to vote for the Green party, then support for that party will be over-estimated, due to coverage error.

Thus, when considering whether the survey is representative, you want to know if every member of the population has a chance of being included in the sample. If they are not, then the responses
can be “weighted” to give a statistically more accurate estimate. This means that the answers of groups that are under-represented in the sample are “counted more” than those who are over-represented in the sample. You want to know if the results were weighted to take into account coverage error.

5. **What was the response rate and did members of all groups participate equally? If not, have weights been used when reporting the results?**

Some respondents are systematically less likely to respond to surveys than others. People who live in urban areas refuse to participate in surveys more than those in rural areas. If people in rural areas support the Green Party more than urban voters, than support for this party will be over-represented in the sample. We call this *non-participation error*. To minimize this error, interviewers often return to the home or re-contact the sample member to increase participation of selected individuals in urban areas to ensure that they respond. Still, they are often under-represented in final survey data. Thus, you want to know the response rates of different populations, and to ask if this was taken into account in reporting results. As noted, weighting is used to take into account the under-representation of certain groups in the sample.

6. **How much missing data is there and how might this affect the results?**

Sometimes respondents refuse to answer particular questions, either because they find the question too difficult to understand or they do not want to reveal their answer. This leads to *item non-response error*. When respondents skip questions by refusing to answer a particular question or answering don’t know to a particular question, there can be important differences between those who answer the question and those who do not. This can lead to very misleading results because the statistics will only reflect the answers of people who revealed them, not the answers of those who did not reveal their answer.

A famous example of this occurred in Nicaragua before elections in the early 1990s. Many people thought the survey researchers were from the Sandinistas, the dominant political party, and thus were more likely to answer if they supported the Sandinistas, and more likely to refuse to answer rather than stating that he or she wants to vote for opponents otherwise. Because supporters of Sandinistas were more likely to answer the survey question, they made up a larger percentage of responses than was representing their support in the full population. As a result, the survey was highly misleading.

Surveys are important sources of information, but the information they provide should be examined carefully. Ask six important questions in evaluating polling data, and be informed consumers of a wealth of information.

For more information, see: [http://www.ropercenter.uconn.edu/education/polling_fundamentals_error.html](http://www.ropercenter.uconn.edu/education/polling_fundamentals_error.html) and [http://www.radio-canada.ca/nouvelles/Politique/2012/08/22/001-sondages-elections-methodes.shtml](http://www.radio-canada.ca/nouvelles/Politique/2012/08/22/001-sondages-elections-methodes.shtml)