



# **KONDA** **BIOPSY REPORT**

*(This report includes KONDA November'23  
Barometer Report's "Biopsy Report" section)*







# CONTENTS

<b>1. KONDA BIOPSY REPORT .....</b>	<b>4</b>
1.1. Why Biopsy? .....	6
1.2. First Stage: Mapping .....	7
1.3. Stage Two: Diagnosis .....	9
1.4. Third Phase: Treatment.....	18
1.5. Why Were the Polls Wrong? Evaluation of the KONDA Biopsy Report .....	19
1.5.1. Research method and sampling frame.....	22
1.5.2. Sampling width.....	22
1.5.3. Survey form design .....	23
1.5.4. Application in the field.....	23
1.5.5. Representation.....	24
1.5.6. Motivations for stating voting preference .....	24
1.5.7. Weighting .....	25
1.5.8. Conclusion.....	26





# 1. KONDA BIOPSY REPORT

Although our main purpose and production are based on social data, the KONDA brand is still largely associated with election research and predictions. Therefore, our evaluation as a company is based on how the election research we announce reflects the results of the ballot box.

We published the results of the field research we conducted exactly one week before the general elections held on May 14, 2023, 3 days before the election, as before every election. May 14 revealed a deviation from KONDA's measurement beyond the margin of error.

Many different claims have emerged with the assumption that measurements made through surveys have the power to influence public opinion or political actors. We hope that anyone who believes in the power of science and reputation, like us, does not listen to these claims. We expect institutions that claim to do similar work to us to abandon fallacies based on rumours, journalists to distrust these rumors and this issue to be discussed on a methodological basis.

We described the fact that the research data did not match the ballot box results as a situation that needed to be examined, and we turned it into a biopsy study, a summary of the results of which you can read in this report. In this study, we questioned everything from method selection to sample design, from our field methods to our weighting methods.

Eren Pultar, who worked at Konda until 2021 and has knowledge of all our processes, even from the outside, coordinated this process. At this point, I would like to thank her for putting forward this multi-faceted, complex and valuable work for us. This study, which was carried out with the support of different experts, including especially Prof. Dr. Nebi Sümer, has a much broader output than the summary you will read below, and we will share these with other experts and continue to improve ourselves.

As a result of the study, we would also like to find a very specific reason that caused us to make a different measurement than the election result. However, the more complex the organism we call society is, the more complex the methods and processes used to measure it are. Every step, every element affects each other. Therefore, rather than a defined error, we encounter approaches to be improved. The biopsy study is not an outcome, but rather a new beginning for KONDA's sustainable social measurement efforts.

As a result of the biopsy process, there is only one clear finding for us: In the face of the complexity of measuring society, it is necessary to continue to strive and improve our work without compromising science and truth...

We present our Biopsy Report for your review, hoping that it will benefit everyone who wants to understand society through knowledge, science and measurement.

Best Regards,

Aydin Erdem  
KONDA Research  
General manager



## 1.1. Why Biopsy?

We first conducted the study we called biopsy when our measurements before the November 1, 2015 elections were very different from the election results. We had conducted a review of why election polls could be wrong, but we had not reached a definitive conclusion that would explain the 8-point deviation in the vote of the largest party. Although we have made constant efforts to improve our processes since then, there has been no fundamental change.

Due to the atmosphere in this year's May 14 elections, many people were looking forward to Konda's election measurements. But our measurement in the latest study turned out to be wrong. Although we saw Kemal Kılıçdaroğlu's vote ahead in the presidential election, Tayyip Erdoğan finished the first round ahead with a 4-point difference. Because of this error, the measurements we made for the parliamentary elections, some of which were outside the margin of error, were overshadowed.



Even though the measurements made by the survey companies were generally not accurate, we aimed to correct this mistake by overhauling our own processes and using a comprehensive approach that could reveal all the stages that may cause the measurements to fail and pave the way for improvement rather than finding an easy, patchwork solution to the problem we first identified. We saw this as an opportunity to start a biopsy process.

Being wrong in election polls is neither unique to Turkey nor to the May 14 elections. The increasing difficulty of conducting surveys can now cause companies to make collective mistakes in many elections. For example, the fact that they generally mismeasured the results of the Brexit referendum and the 2016 presidential election in the USA, which Donald Trump won, increased the scrutiny on the mistake. This type of examination is called autopsy (post-mortem) in the literature. Even though the election is over, we call it a "biopsy" because we want to believe that the patient is still alive and the election polls are salvageable.

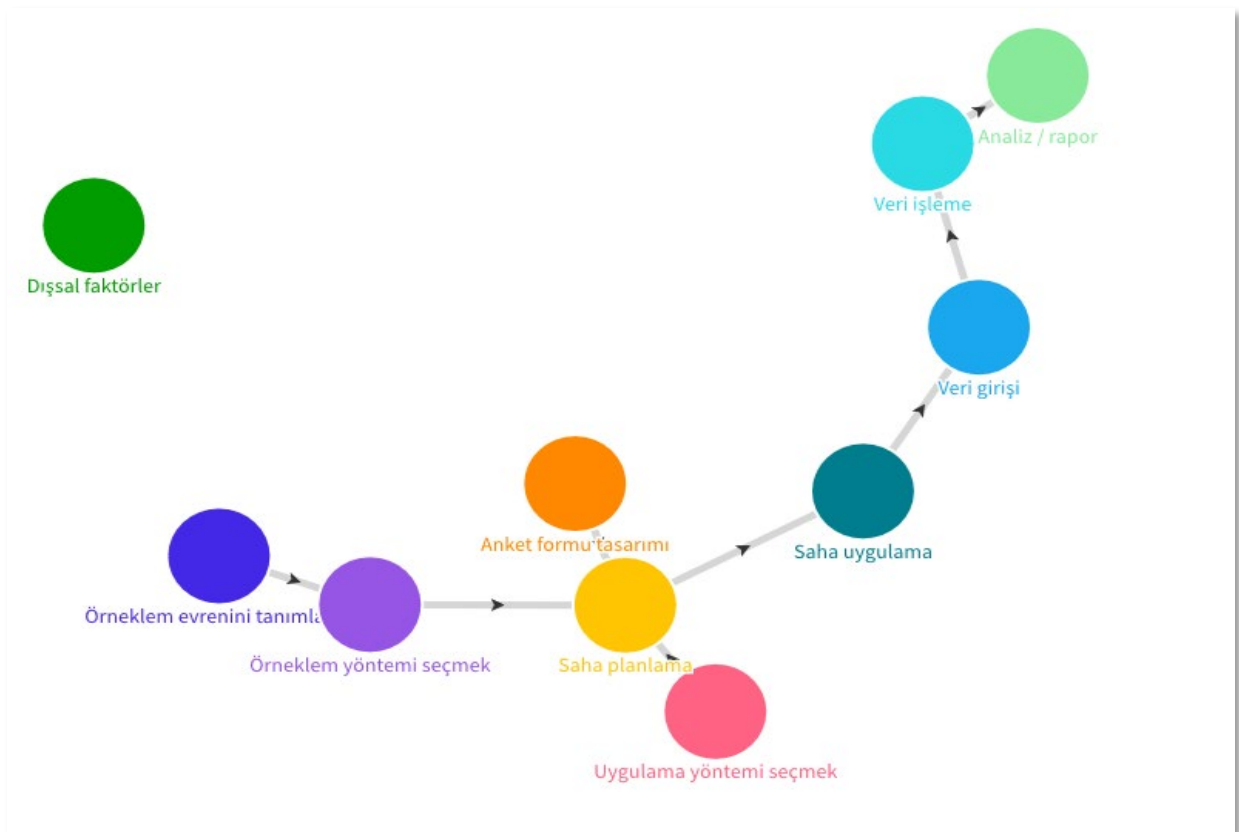
**We mapped the research process and made diagnoses; we are starting to apply treatment.**

Our biopsy process proceeds in three stages: Mapping and planning all the steps and decisions of the research process; To diagnose which of these steps and decisions may have errors or deficiencies; to apply treatments that can be applied for steps and decisions that can be corrected or improved.

## 1.2. First Stage: Mapping

We listed each decision and the work we did at each step in the process, from deciding to measure voting behavior in the first place to publishing the measurements three days before the election. Even though this process works smoothly for every Barometer every month, by naming them, we have created a map that will help us understand why we follow the steps we follow without questioning, what kind of mistakes we may have made, and how we can improve or change them (Graph 1). This simple flow shows the process of Barometer and any quantitative research in KONDA in general.

In this flow, we brought together each decision, step or issue we made consciously or implicitly and revealed the areas we will examine in depth. Revealing these also helped those in the team who were involved in different stages of the process and implemented different parts of the process to get to know the other stages better. Moreover, although we have the chance to intervene in some of these decisions, steps and issues, some are beyond our control, but we still thought it was important to name them. We made an interactive visualization to see it all together and go into detail if necessary (Graph 2).



Graph 1: Process in KONDA's quantitative research



All



Graph 2: Decisions in KONDA's quantitative research





### 1.3. Stage Two: Diagnosis

#### How wrong were we?

The research we published can be said to be in the middle of survey companies in terms of success.

Moreover, although we cannot predict, for example, the vote rate of İyi Parti, it seems that there is no other company that can predict it exactly. The 2.4 percent vote of the Yeniden Refah Party was a surprise not only for KONDA but for all companies.

We examined exactly how wrong we were: The numbers we measured reflect the real situation with a certain probability within a certain range of numbers. On May 6-7, a week before the election, we interviewed 3480 people to represent 60.7 million voters, and with this sample size, we aimed to measure the vote rates accurately with a 95 percent probability and within a certain confidence interval (within the margin of error). According to our calculation of confidence intervals, we have measured the vote rates of Sinan Oğan, the Ak Parti and the TIP correctly within the margin of error, and we have measured the other candidates and parties outside the statistically acceptable vote rate ranges.

#### Research method and sampling frame

To measure the voting preferences of voters in Turkey on May 14, we use a quantitative and probability-based research method (that is, each person's probability of being elected is equal to the others). Although there is a list containing the names and contact information of 60,721,745 voters for a list that we call the "sampling frame", which will allow us to select and access random voters, such a list is not available to us and we cannot use it because it contains personal information. It would be neither ethical nor legal. In this case, as an alternative, you can view all addresses, phone numbers of people, social media accounts, etc. in Turkey. It is necessary to turn to other lists containing In KONDA, we use the [election results](#) of the Supreme Electoral Council as another list containing the most detailed information about all voters in Turkey for the sampling frame. The results are listed poll by poll and we define the neighborhood/village as the smallest unit by combining polling stations and polling areas.

At this point we encounter the first source that may cause an error. For the May 14 elections, we used the results of the previous general election, 2018. This list simply reflects fewer voters than in 2023, with 56 million voters. There are changes whose impact we do not know such as the profiles of those who have died since 2018 and those who are now voters because they have come of age are different from each other, the places where they live have changed, there are approximately 2 million people who cannot vote according to the law, there are those who have acquired citizenship among Syrian immigrants and those who have purchased citizenship rights, etc. In summary, although the voters on the list we use as a sampling frame, the voters we want to represent, the people we meet when we go to the field, and those who go to the polls and vote on election day overlap to a great extent, they are not exactly the same.



### Sample:

Although simple random sampling is the most accurate method to keep the probability of randomly selecting any of the voters equal, the cluster sampling method is preferred due to its cost and impracticality. We use four criteria to cluster 49301 neighborhoods/villages in Turkey and create clusters from their combinations.

- 1) NUTS regions,<sup>1</sup>
- 2) Settlement type,
- 3) Political competition,
- 4) Education level.

We examined our sample and its structure, and according to these criteria, we saw that the election results in the neighborhoods we visited in the samples we took for the last four surveys we conducted before the May 14 election were very close to the election results in Turkey. In other words, we understand that the sample worked correctly and that we went to the right neighborhoods/villages. We announced just a week after May 14 that there was no fundamental problem in our sample, and therefore we did not see any obstacle to publishing the research we would conduct for May 28.

However, to examine the sample structure in more detail, we experimented completely random sampling without the political competition criterion and without the education level criterion. We have seen that when it is random, there is a risk of it becoming unrepresentative. The criterion of political competition, based on our subjective definitions, works in favor of some parties and against others. The effect of education level is quite low.

KONDA's sample structure has been based on the neighborhood/village unit since at least 2006. Although this logic was effective in those years when 30 percent of the voters lived in rural areas and the metropolitan neighborhoods did not grow that much, it now causes an imbalance such that the village with only 9 voters and the metropolitan neighborhood with 60 thousand voters are put on the same level. Even though this balance is observed during cluster definition and sample selection, that is, the probability of selection of the first one is calculated to be much lower than the probability of selection of the second one, 5 thousand metropolitan neighborhoods correspond to one-tenth of the neighborhoods/villages in terms of the number of units, but half of them in terms of the number of voters. We can predict that this imbalance will increase even more as metropolitanization increases. For this reason, we identified the sample structure for the upcoming elections as an area that needs improvement.

### Sampling:

---

<sup>1</sup> Since 2003, the European Union has adopted the geographical classification of geographical regions called **Nomenclature of Territorial Units for Statistics** (NUTS for short) for use in statistical data. According to this classification, there are 12 NUTS regions in Turkey at the 1st level, 26 sub-NUTS regions at the 2nd level and 81 provinces at the 3rd level. Since 2005, KONDA has been using this geographical distinction in both its sample design and analysis and reporting, instead of the 7 regions traditionally used in Turkey.

[https://en.wikipedia.org/wiki/Nomenclature\\_of\\_Territorial\\_Units\\_for\\_Statistics](https://en.wikipedia.org/wiki/Nomenclature_of_Territorial_Units_for_Statistics)  
[https://tr.wikipedia.org/wiki/T%C3%BCrkiye%27nin\\_%C4%B0BBS%27si](https://tr.wikipedia.org/wiki/T%C3%BCrkiye%27nin_%C4%B0BBS%27si)  
[https://en.wikipedia.org/wiki/Nomenclature\\_of\\_Territorial\\_Units\\_for\\_Statistics](https://en.wikipedia.org/wiki/Nomenclature_of_Territorial_Units_for_Statistics)  
[https://tr.wikipedia.org/wiki/T%C3%BCrkiye%27nin\\_%C4%B0BBS%27si](https://tr.wikipedia.org/wiki/T%C3%BCrkiye%27nin_%C4%B0BBS%27si)



After modeling the sample universe we have defined by clustering and using the election results as the sampling frame, completely random neighborhoods/villages are selected among the clusters in proportion to the cluster size. However, at this point, certain restrictions arise regarding going to these neighborhoods and conducting surveys, based on both practicality, speed and cost. Although the sample balanced large neighborhoods and villages, our analysis found that the smallest villages were underrepresented: Although villages with fewer than 400 voters accounted for 5 percent of all voters, it makes up only 1.4 percent of those we interviewed in the last 1.5 years and this segment is much more inclined to vote for the Ak Parti than the general public.

Another restriction that comes into play when taking samples is the provinces visited. When we consider the provincial election, we say that we conducted surveys in an average of 29 provinces. In the five surveys we conducted around May 14, we revealed that in the provinces we visited, the votes of the Millet Alliance were higher than those of Turkey in general, while those of the Cumhuriyet Alliance were lower. This is also one of the possible sources of error: Even though we choose the neighborhoods and villages in these provinces that will represent the whole of Turkey, it may create a bias that we try to correct by weighting them later.

When we define provinces and take samples from the survey management system specially prepared for KONDA, we try not to disrupt the randomness by being as loyal as possible to the neighborhoods/villages that appear. However, due to location-related reasons such as the village road being closed by snow, lack of suitable transportation, a funeral in the neighborhood, or organizational reasons such as the obligation to complete the survey on a weekend, some of the initially selected neighborhoods were either prepared beforehand or an equivalent neighborhood in the same cluster was placed on the field day. or we can replace it with a village. In addition, we cancel and delete the data of the neighborhoods/villages visited by surveyors who are found to have filled out fake surveys. Our experience and brief review indicate that their impact is minimal and they are unlikely to be a source of error.

### **Size of the Sample:**

In such studies, which are based on the principle of taking a spoonful of soup to understand its taste, the size of the sample, that is, how many people will be interviewed, depends on the size of the population to be represented. It may vary depending on how much margin of error one is willing to tolerate in the resulting ratios and how precisely one wants to determine the likelihood of these measurements being accurate. The sample size of KONDA surveys that aim to represent the voter population or adult population in Turkey is also based on such a calculation. It is enough to meet 2,400 people to represent 60 million 721 thousand voters by agreeing to find rates 2 percentage points above or below the actual rate and claiming that there is a 95 percent probability that these rates will be in this range. Therefore, the number of 3480 in the research we conducted on May 6-7 and published the measurements is sufficient and the failure to match the election result is not due to this.



### **Survey form design:**

How the subject that is intended to be learned through research is addressed and how the questions are asked also affect the result. In order to learn the voting preference accurately, we apply standard methods such as not asking leading questions beforehand, not listing the parties but asking open-ended questions, and asking people to mark their votes on the ballot paper towards the election. However, this issue was not widely included in the biopsy examination, because it is not possible to perform any test for research before the election and numerically examine how the alternative will yield results. In future research, we aim to add questions just to examine this issue and test whether these questions have an effect on bias and social desirability.

### **Field Application:**

After the sample is taken and the survey form is prepared, that is, after it is clear where we will go and what we will ask, the field team steps in and the surveyors they assign knock on the doors, persuade people to participate in the survey, read the questions one by one in the doorway and complete 18 surveys in one day as standard. In this process, the effects of both personal skills and experiences and relationships emerge. We also apply age and gender quotas in the field. Otherwise, there is a risk that in a patriarchal society, the balance between men and women may be disturbed because women are more hesitant to talk to a stranger, or that pollsters may conduct more surveys with young people who are their peers.

We examined the survey numbers for field application. Although the target is 18 surveys, due to the skill of the surveyor, the guidance of the field team or some events beyond the control of the surveyor, they may conduct incomplete or excessive surveys and this may create a deviation. To examine the impact of this, we weighted the election results in the neighborhoods/villages in the sample according to these survey numbers and the data we collected in the field as if we had conducted 18 surveys in each neighborhood. We saw that the result did not change much either way and that this was not one of the major sources of error.

One of the biggest risks in research is that pollsters do not go door to door but fill out fake surveys, meaning they do not actually collect data from voters. Since a significant part of our field work was based on assigning, training, directing and controlling reliable people on the field day, we considered it very unlikely that this was the result of an error in selection and did not include it in detail in the biopsy. We did not revisit our three-pronged control process at this stage, as it was very difficult for the surveyor who did not conduct a survey but filled it out himself. Similarly, although it is an error-prone step, we did not consider it as a possible source of error since entering the answers in the surveys into the computer, that is, the data entry stage, is constantly checked.

### **Representation**

If our sampling logic is correct and each pollster completes the required number of surveys in the neighborhood or village they visit, we can accurately represent the voters in Turkey. So can we achieve this? In order to understand this, we check the data when we receive it in every research. In addition to TÜİK, the source we most frequently consult is the number of users announced by these media for social media usage rates, the World Health Organization for



alcohol consumption, the Ministry of Health for Covid vaccination rates, etc. We take these as reference points.

To verify the demographic representation of the research we conducted and published before the May 14 elections, we used the election statistics report published by the Supreme Electoral Board (YSK) approximately one month after the election and the most up-to-date demographic data of TÜİK.

When we compare it with the YSK data, there appears to be a small difference in gender distribution that is not enough to explain our inability to achieve the results. However, in terms of age distribution, it seems that we have not reached enough young people under the age of 29 and those over the age of 75. When we compare the participation rates by age and gender, there are differences in each group: Although there was an average of 86 percent participation, the people we interviewed in the research told us that they would participate in the ballot box at an average rate of 92 percent. But there is an imbalance: For example, although 97 percent of women between the ages of 18-34 said they would go to the polls, 87 percent did; On the other hand, although 97 percent of women between the ages of 35 and 49 said they would go, 91 percent did.

The difference between the age and gender distribution in TÜİK's 2022 data and the distribution in the research indicates that we meet less with the 33-48 age group than we should and more with the 49 and over age group, especially among men. Also, we interviewed more than it should be with people living in two- and three-person households. It turns out that we meet less with those living in households with 6 or more people.

### **Reasons for not being able to represent and acceptance rates**

The fact that the resulting data in a probability-based survey does not reflect the population may be due to the fact that some segments cannot be reached under equal conditions or some segments participate in the survey less or more. For example, the difficulty of entering gated housing estates or accessing remote mountain villages may lead to exclusion and unequal selection probabilities (non-coverage error). On the other hand, since single elderly women are more cautious than the rest of the population about opening the door to a pollster, or because a party's voters are generally more motivated and more willing to take part in a survey, even if they are accessible, there may be an inequality in their participation in the survey (non-response error). For example, it is not difficult to guess that the fact that 97 percent of women in the 18-34 age group told us that they would go to the polls, but 86 percent of them actually went, is not because these women made their intentions clear to us, but because young women who would not go to the polls anyway were not inclined to talk to us. This kind of inequality, especially among party voters, means that we can only measure the vote distribution of those who agreed to meet with us, not the electorate in general, and this may lead to inaccurate measurements.

Since being able to meet voters from different parties unequally is critical, we examine the differences in door-opening acceptance rates to see if such an error exists. If a pollster in neighborhoods/villages where the Ak Parti is strong has to knock on an average of 40 doors to complete 18 surveys during the day, and on the other hand, in places where the CHP is



strong, she knocks on 80 doors to reach the same number, this indicates that surveys are difficult to conduct and there may be differences in willingness to participate in the survey. In order to analyze exactly this, pollsters sometimes collect data on how many doors they knock on. We collected this data in the research we conducted in November and December 2022 and for the second round of the Presidential election. We are still examining with regression models what might have affected acceptance rates and have not yet reached a conclusion. However, at first impression, there are some differences between the parties. If there are real differences, this could be one source of error, and party voters may not be opening the door to our pollsters at the same rate, causing us to measure voting preferences differently.

### **Motivations of stating the voting preferences**

Let's assume that we reached people who exactly represent the voting population in Turkey and conducted a survey. Will these people accurately tell a stranger who is knocking on their door who they will vote for? Of course, there are many understandable reasons for them not to say so: such as fearing that their voting preference will be known if they are oppositionists, thinking that it will increase their chances of receiving social aid if they vote for the ruling party, giving the answer they want/expect to hear based on their impression of the pollster, and actually deciding at the last minute while going to the polls. There are well-established terms in the literature for such motivations: shy Tories, for those who used to hide their preference in the UK due to criticism in the media, who did not reveal their conservative party preference and kept it at the ballot box; Like the Bradley effect for a candidate who loses against a white mayoral candidate even though his verbal support is more acceptable because he is black. In the movie "Züğürt Ağa", the whole village says that they will vote for the lord in the headman election, but only one vote can be seen as a caricatured extreme example of such a situation.

It is enough to make a simple comparison to understand whether the voters told the pollsters their voting preferences correctly: The voting preferences they said they would use a week before the May 14 elections and the voting preferences they said they would use four days after the elections. Those who said they voted for the Ak Parti increased by 4 points, that is, approximately 2.5 million people. Those who said they voted for the Yeşil Sol Parti decreased by 2 points, or approximately 1.2 million people. While those who were undecided, said they would not vote, or did not answer the question at all were 15 percent before the election, those who said they did not go to the polls or voted blankly or did not answer the question at all after the election dropped to 8 percent. In other words, the conclusion of the election has created a change in the motivation of many people to vote.

The fact that some voters do not express their voting preferences accurately is actually a situation that KONDA has known and been aware of for years and has modeled its process accordingly. We are trying to present as accurate a picture as possible by asking many questions about politics and asking people to mark their votes on the ballot in addition to the standard election question a month or two before the election.





## Weighting

So far, we have discussed many possible deviations that may arise from sample structure, sample drawing, and practice in the field, and we have shown that there are deviations in both demographic representation and voting preferences. Even though we cannot reach the full representation we want when the field work is completed, it is possible to reduce the actual rates to the actual rates by using statistical methods and weighting. For example, if we want to represent a population that is 50 percent women and 50 percent men, but 25 percent of those we interview are women and 75 percent are men, when we double all the answers of women and reduce the answers of men by two thirds, we actually obtain data as if we have observed the necessary balance of men and women. We investigated many different weighting methods for biopsy work, evaluated which ones were applicable, and tested some on data.

## Demographic weighting

At KONDA, for a while now, in order to ensure demographic representation, we have been using the weighting method called raking, first according to one demographic characteristic, then weighting the resulting distribution according to the next demographic characteristic, and repeating until the resulting distributions agree with the TURKSTAT data we use as reference. When we applied different weightings to the data, which we created from combinations of features such as gender, age, region, settlement code, employment status, we compared the resulting voting preferences with the data from the field and the real election result. We would expect demographic weighting to partially correct the deficiencies in clusters such as young women, those over 75, and those living in large households. However, it has little effect on the data coming from the field, it does not bring it much closer to the election result, and we did not see this as one of the possible sources of error.

We also tried different weighting methods from the data. For example, instead of considering demographic characteristics separately, we applied the cell-based weighting method, which takes all of them into account and fine-tunes them. We made weightings according to the YSK's election participation rates by age and gender and by TÜİK's age and gender distribution in the regions. These weightings did not create a fundamental change in vote preferences and did not bring us close enough to the real election result. Therefore, it seems difficult to claim that the error stemmed from not being able to predict whether voters would participate in the election or not.

When performing demographic weighting, we assume that the answers of women who participated in the survey were similar to those of women who did not participate in the sample in which we balanced the male-female ratios. However, they may have different profiles and different voting preferences; Some studies of panel studies show that in some demographic groups, voters of one party are more likely to participate in the survey. For example, among voters over the age of 75, whom we have less access to, if Ak Parti supporters participate in the survey more and CHP supporters participate less, weighting may cause Ak Parti supporters to appear more in the survey than they do and CHP supporters to appear less. If weighting ensures that demographic clusters are represented in the right proportions, we would expect weighting to move us closer to the election result, and if there is an imbalance, to move us further away. Our initial calculations show that age weighting brings it closer, albeit slightly,



while gender weighting brings it further away. Therefore, this could be one of the sources of the error. We will continue to conduct this analysis in other studies using other demographic characteristics.

### Political weighting

Some voters declare a preference that differs from the actual vote. As a matter of fact, in all our research, the ruling party is said to have received a much higher rate than it has received. Based on this, although we do not yet know the election results, we do not take the vote preferences for the next election as they declare. To close this gap between statement and behavior, we also apply political weighting using their answers to other questions about politics. Even though the result of this weighting is much closer to the election results than it was before weighting, it is still not a measurement that falls within the margin of error.

Another potential source of error seems to be choosing to use spoken voting preferences rather than voting preferences marked on the ballot. Among those marked on the compass, MHP, YRP and Zafer Party have a slightly higher rate, the CHP has a slightly lower rate, and these are closer to the election result. At the same time, since the rate of those who do not mark is lower, it is more likely to give more accurate results. However, the difference between what is said in the Presidential vote preference and the ballot paper is much less.

Those who said they had voted for the İyi Parti for a long time were much lower than they actually were, and in order to calibrate the real election result, it was necessary to give a coefficient of more than twice in the political weighting. It did not seem possible to intervene without any data regarding this situation, which was a symptom indicating that there might be a problem in political weighting. However, it is stated in the literature that giving weights below 0.5 or above 2 for any element will give unhealthy results. One solution to this is to trim the coefficients and not go beyond these limits. First, since only İyi Parti was outside these limits, we reduced its coefficient to 2 and calculated new coefficients. When we made this calculation, the voting rates of the Ak Parti and CHP increased even though they should have decreased. Thereupon, we kept the coefficients of these two parties whose coefficients were below 1 constant and applied the recalculation of coefficients only to the other parties. Although this method brought the vote rate of the İyi Parti very close to the election result, it could not prevent the Ak Parti, the CHP and the YSP from remaining higher than they should have been.

In summary, making some decisions differently regarding political weighting could have enabled us to obtain slightly closer results. We make the biggest intervention on the data we obtain throughout the entire process through political weighting, and it does not seem possible to give up on this. Therefore, reducing as much as possible the impact of all decisions and steps that may prevent us from measuring accurately beforehand and that may have contributed to the error, even if only slightly, may enable less intervention at this stage. In addition, rapid changes in the political landscape,. Due to changes such as alliances between parties, divisions, mergers, the increase in the number of parties participating in the elections, it is necessary to try different methods and make different analyses.





### **Diagnosis phase in summary**

Although we did not complete the diagnosis phase completely, we completed it to a large extent, and a clear picture began to emerge as to why we could not make the selection, and what wrong decisions we made throughout the process, or the steps we took incorrectly or incompletely. The review, at least by the method of elimination, enabled us to understand which decisions and steps did not cause the error.

Almost every decision and step we have pointed out does not alone explain our mistake, but only a small part of it. If we correct everything we can in decisions and actions, we hope that the need for political weighting will be greatly reduced, and at least the research results we publish will come closer to the election results and remain within the acceptable range/margin of error.

### **Factors Outside Our Control**

There are also some events that occurred specifically for this election or factors that any polling company encounters in any election in general but are not under our control, and they may partially explain our failure to win the election. Since what we can do to fix/improve these is limited, we have briefly listed them here so that you can understand them and not miss them.

- The fact that we conducted the research we published on May 6-7 in a very fast election period, a week before the election, and voters' preferences can change within a week,
- Muharrem İnce withdrew his candidacy for the presidency three days before the elections,
- Voters with certain demographic characteristics do not participate in surveys at all but still go to the polls.



## 1.4. Third Phase: Treatment

After completing the first phase, namely mapping and planning, in July, we spent September, October and November with the diagnosis phase. Although there remain some decisions and steps of the process that we can analyze and have not yet diagnosed, we have largely completed it. Some of the sources of error arise from processes that are under our control. We will implement corrections and improvements in the coming months and continue to keep you informed.

### Suggestions for Correction/Improvement

#### Sample and Sampling

- Testing using different data sources to update/calibrate the sampling frame over the years,
- Making it based on ballot box areas rather than neighborhood/village based,
- Updating the data used in the sampling criteria, trying to use different criteria such as socio-economic development index,
- Trying to use the 2nd level NUTS region as a criterion instead of the 1st level.

#### Survey form

- Adding questions that will reduce bias to the form and examining their effects, and examining the effects of those that have been asked so far.

#### Field

- To examine the possible impact of increasing the number of provinces or changing the provinces, to conduct field trials,
- Investigating what we can do to reach missing demographic groups such as young or older women,
- To investigate the need and method of making the practice of collecting acceptance rate data permanent.

#### Weighting

- To make analyzes for Outlier, inconsistent subject etc. and remove some subjects from the data according to the results,
- To extract data from neighborhoods/villages where surveys were conducted below or above a certain number,
- In addition to demographic weight, political weight and trimming, removing inconsistent, missing/excessive surveys and applying all the experiments we made throughout the biopsy together and examining the effect,
- Experimenting with combinations of old and new questions for political weighting.



## 1.5. Why Were the Polls Wrong? Evaluation of the KONDA Biopsy Report

*Prof. Dr. Nebi Sümer*  
*Sabancı University*

Although all polling companies correctly predicted the Barack Obama-Mitt Romney race in the 2012 American presidential election, almost all reputable polling companies were wrong about the Donald Trump-Hillary Clinton race. A similar situation occurred in Turkey in the May 14, 2023 elections. The vast majority of polling companies even institutionalized companies like KONDA that have made a name for themselves with their accurate predictions, were wrong, especially in their Presidential predictions. KONDA launched an in-depth systematic analysis process immediately after the election to investigate the possible reasons for this error and to determine new strategies in its future surveys. The "biopsy" report prepared for the completed part of this ongoing process, in which I was also a part, should be read as a self-assessment interim report. I evaluated this report, which is based on numerous analyses, by critical reading from the perspective of the referee who evaluated the article. However, first let me share my general views on the much-talked-about "margin of error".

Political public opinion research consists of an effort to generalize the findings observed (obtained) from a sample to the universe by using basic objective methods in the field of measurement-evaluation science. Scientific criteria are used to understand to what extent the data observed from the sample represent its equivalent (i.e. universe parameters) in the universe (voter). Therefore, we define survey-observed behavior as the true behavior variance plus the error variance ( $T\text{-true} + E\text{-error}$ ). The less this unintentional error (variance), the more accurately a behavior observed from a representative sample reflects the real behavior in the universe. If the task was to measure Turkey's average height, the error variance would probably be limited to the sensitivity of the tape measure used and the good performance of the people doing the measuring, and the observed behavior would correspond to the average height in the universe. But when it comes to predicting voting behavior, the number of factors that will contribute to error variance is almost endless! Factors such as choosing the right sample to asking the right questions, the extent to which each person contacted is willing to answer, to what extent the answer reflects the person's current daily thought, social desirability (norm) or vote in the election, the impact of political events that took place between the date of the survey and the election, part of the error variance. Even factors such as how much consistency there is between attitude (expressed opinion) and actual behavior (votes cast in the ballot box) in the country can influence the process. Even if each factor does not contribute significantly to the error variance on its own, it weakens the representativeness of the observed behavior to the real behavior by enlarging the error variance through cumulative contribution.

Survey companies try to estimate all this potential error variance within what they call the "margin of error", usually within the 95 percent confidence interval. In other words, they claim to the public that if we conduct this survey 100 times, the results will be in the range we gave 95 times, and the voter's vote will most likely be in this range. The only way to realize this claim is to objectively identify the sources that contribute to error variance and conduct a survey with a research design and field application that will minimize their effects. This requires



systematic testing of the psychometric quality of the measurement, that is, the reliability (consistency within and across different time periods) and validity (power of predicting actual voter behavior) of the surveys.

In multi-party systems, when the margins of error calculated for a single party or candidate are taken into account for all candidates, the differences between the lower limit of one and the upper limit of the other actually include a very high range of variation. For this reason, margins of error are not enough to show the truly accurate prediction range. In general, commenting on trends observed from many different surveys is more reliable than looking at just one or two surveys. Therefore, the most important finding for companies conducting political surveys is the consistency of the pattern of change among their repeated surveys, that is, the pattern of explainable change. A series of polls showing the gradual increase or decline in the leadership of a party/candidate is seen as a concrete sign of a real trend, free from margins of error. Since punctuated changes represent a major transformation, they should be viewed with skepticism, considering whether they have a real counterpart or to what extent they are an illusion of compound margins of error.

For example, in the three surveys conducted by KONDA before May 6-7, Kemal Kılıçdaroğlu's votes were 44.6, 45.0 and 45.5 percent (1-2 April, 15-16 April, 29-30 April). Although these are within the confidence interval according to the 14 May results, the fact that it left the confidence interval with 49.3 percent on May 6-7 should be interpreted carefully as a leaping change. Although this change is partially explained by the decline in Muharrem İnce's vote, when looking at the size of the deviation, it is noteworthy that the margin of multiple error has increased to a critical level, especially in the last week.

One of the most fundamental determinants of the margin of error is sample size. In large samples, the margin of error is smaller. However, in large samples such as the one used by KONDA, the margin of error calculated for the entire sample is not valid for subgroups and parties. Since small parties are represented by a smaller sample, their margin of error is much higher. In other words, in multi-party systems, the margin of error in the 95 percent confidence interval turns into a meaningless value, especially for small parties. Since the margin of error is calculated for the entire sample, it is much higher than for the overall sample for subgroups (e.g. only young people, women) or small parties (e.g. parties with less than 5 percent of the votes), and the bias will be higher in these groups.

Another important factor affecting the margin of error is the variance width. In homogeneous groups where the variance is low, the margin of error is relatively lower. In surveys, small party supporters and sub-groups (for example, Kurds for Turkey) are both more difficult to reach and less likely to receive a response when reached (non-response bias). For this reason, an attempt is made to correct (get closer to the universe parameter) by weighting to balance the representation of small parties or sub-sections. Since high weighting increases the variance in the distribution within the party, it increases the error range for small parties much more, resulting in a bias called the "pattern effect" in statistics. For this reason, in multiple political systems, especially in small parties, margins of error and naturally deviations from the prediction are expected to be higher. As stated in the biopsy report, we can relatively reduce



the weighting variance inflation by trimming by taking the square roots. However, its contribution to reducing error variance is limited.

Demographic weighting is mandatory up to a certain level and, as stated in the report, it increases the predictive power, even if it is small. However, it is not possible to correct sample statistics that deviate greatly from the population parameters by weighting. Even with low deviations, weighting has a limited correction effect. In KONDA surveys, since the deviation from the population parameters in basic demographic characteristics such as age, gender and education level is generally low, the effect of weighting on changing the general pattern is low. However, it would be useful to pay particular attention to education level in demographic weightings and to try weighting methods in demographic subgroups. For example, the weightings for education level and age should be made separately for men and women and the effect should be examined. Potential laterality arises more from political weighting. High political weighting is used because the probability of reaching major party supporters and receiving a response is disproportionately greater than the probability of reaching small party supporters and receiving a response. High levels of political weighting may increase the intra-party variance, thus widening the margin of error, and may lead to biased estimates because the equivalent in the universe parameters is not fully known. Therefore, it is necessary to develop sampling methods that will reduce the need for weighting and full access to the determined sampling frame.

Another factor that increases the margin of error may arise from the proportional distribution of swing votes across highly unstable parties. It should be investigated to what extent the proportional distribution of undecided voters among parties, which rises to almost 30 percent before the election periods, causes forecast deviation. The proportional distribution of swing voters, who are generally in the middle of the right-left spectrum in terms of political attitudes, whose level of participation in the election is lower than those who express their voting preference, and who have a higher tendency to follow the majority (herd psychology) when approaching the election, may be an important deviation factor. Particularly in surveys with high political weighting, the proportional distribution of swing voters can be a significant cause of deviation. In KONDA surveys, even in the week before the election, the rate of swing voters was 7.8 percent in the 6-7 May survey, , and 2.2 percent said they would not vote; In other words, a total of 10 percent of the votes are distributed proportionally. If the previous distribution is already biased, the proportional distribution will further increase the bias. Weighting strategies based on models that estimate the vote preference of swing voters, taking into account political attitude questions and their place in the right-left spectrum question, should be tried.

With his “biopsy” study, Konda examines each measurable and statistically analyzable process from the sampling process to the analysis of data, with in-depth comparative analysis, that may contribute to error variance. Thus, it reviews the psychometric quality of the research method and measurement it uses. Below I have evaluated each stage in the biopsy in the same order.



### **1.5.1. Research method and sampling frame**

A randomly selected sample from a universe whose parameters are fully known can represent the universe. In Turkey, which is experiencing rapid socio-demographic change and internal migration, as stated in the report, the sample drawn according to the parameters of 56 million voters in 2018 does not fully overlap with the 61 million voters in 2023. Not only the level of overlap in numbers, but also the differences resulting from the diversification of the type of voters (immigrants added, those granted citizenship, etc.) and demographic changes in the five-year period (decrease in village population and increase in median age, etc.) may also have caused relative bias in the level of sample representation. Another factor that is not taken into account in the biopsy but may contribute to bias related to sample representation is that foreign votes are not included in the sample circle. The real universe is actually about 64 million, including 3.4 million voters abroad. Participation in elections abroad, at 52 percent, is much lower than the participation in Turkey (87 percent). The distribution of votes abroad among parties is not parallel to the distribution within Turkey. For example, while the Ak Parti's vote rate abroad is approximately 44 percent, it is 35 percent in the country. With this disproportionate participation and non-parallel voting preferences, 5 million of them voted for the first time and four million of them voted abroad in the last five years. More than 9 million (14 percent) living voters, two hundred thousand of whom became citizens and became voters, were not represented.

Apart from this, although voters in villages with fewer than 400 voters constitute 5 percent of the total electorate, only 1.4 percent of the people interviewed come from this segment, which seems to have further biased the probabilistic random selection. In summary, the unbalanced representation of overseas and village voters who are likely to vote for the Ak Parti may have created relative bias in the surveys. Increasing the representativeness of the universe parameter through demographic and political weighting of the sample defined by layers based on regions and neighborhood/village units through probability random sampling also has its own limitations. As a result, the lack of sample representation may have contributed to the fact that the votes of the Millet Alliance were higher and the votes of the Cumhur Alliance were lower in the provinces visited in the five surveys conducted before May 14, compared to the general population of Turkey.

In future surveys, the sampling frame based on neighborhood/school voter ballot box distribution based on updated voter lists may increase the representativeness of the sample's population.

### **1.5.2. Sampling width**

The sample size above 3000 used by KONDA is above the sample size (2400) determined with a 2 percent margin of error based on the 95 percent confidence interval traditionally used in proportional samples. It is possible to relatively increase the predictive power of the measurement by reducing the margin of error. However, this may not be feasible as it would increase the cost exponentially by increasing the minimum sample width required. For example, when the margin of error is reduced from 2 percent to 1 percent, the required sample increases approximately fourfold (9588 people). In summary, although it is possible to reduce the sampling error by obtaining statistics closer to the population parameter as the sample size increases, large samples are not feasible in terms of cost and continuity.



Therefore, instead of increasing the sample size, it is a better strategy to use methods that will increase measurement power (validity) and take measures to reduce bias.

### **1.5.3. Survey form design**

Factors such as the design of the survey form, its length, the order of the questions/items, the question type (yes/no style, Likert type multi-range, open-ended, etc.), the ability of expression and understandability level to be equivalent for all participants affect the measurement power. For example, as a result of the order effect, people whose political preferences are asked at the beginning of the survey may answer the attitude questions at the end of the survey consistent with the political preference they previously expressed, rather than the way they actually feel. Balancing the order of questions is a simple way to prevent order effects. It could be tested in future surveys. Participants with low education levels may have difficulty answering Likert-type questions. A lot of statements and social desirability biases come into play in surveys. The most obvious bias I observe in Likert-type questions asked in political surveys is the bias towards concentrating on the extremes or the middle. In scales with five intervals, the probability of marking 3 is very high, and in scales with 10 intervals (for example, in the frequently used political tendency question), the probability of marking 5, 6 and the extremes is very high. The social desirability effect, which we can define as the tendency to respond according to social norms rather than real thoughts, is more common especially in collectivist cultures such as Turkey. An example reflecting the social desirability effect is the difference between declared participation in elections stated in surveys and actual participation. Although approximately 95 percent of the people interviewed in KONDA surveys said they would vote in the election, actual participation remained at 87 percent. The probability of a person who answers the political questions of the pollster who comes to the door and tells his political preference to say no to the question "Will you vote in the election" is very low due to the social desirability effect.

Asking concrete behavioral questions instead of open-ended questions increases the likelihood of getting the "correct" answer. A situation seen in KONDA surveys that supports this observation is that the rate of "undecided" participants in questions asking about political party preference is very high in open-ended questions, but is lower in measurements made on the ballot paper. This indicates that the likelihood of getting answers to concrete behavioral questions may also be high. In the next stage, the consistency level of the answers given to the questions in different surveys can be examined. Including behavioral questions in surveys that leave as little room for interpretation as possible can increase measurement objectivity by reducing the social desirability effect. In the continuation of biopsy studies, a set of questions with high reliability and validity can be determined through comparative experimental studies on the survey/question format. In summary, sufficient analysis has not yet been made regarding the questionnaire format in biopsy. It is necessary to conduct comparative in-survey research with experimental methods to better understand the errors that may arise from the survey format. For example, future surveys can be tested by including alternative questions.

### **1.5.4. Application in the field**

In field practice, inter-interviewer consistency and 18 survey completion rates may be factors that increase potential error variance. It is understood that pollsters often fail to reach the 18





target in small neighborhoods and villages, further reducing the representation of voters living in villages. Although analyzes made according to the gender of the interviewer do not generally create a statistically significant difference, they need to be examined in more detail. Approximately 55 percent of interviewers interview participants of the same sex and 45 percent interview participants of the opposite sex. The rejection rate of male interviewers (21 percent) is higher than that of female interviewers (17 percent). Similarly, the no response rate for male pollsters (53 percent) is higher than for female pollsters (47 percent). In addition to measures to be taken to standardize field practice, it would be useful to systematically examine the level of consistency between field practices in dimensions other than gender. Since the 18 survey is the smallest unit of the sample area, it may be more accurate to analyze the units completed with 5 or fewer surveys, which may be outliers, from the sample.

### **1.5.5. Representation**

I have summarized general observations about the representativeness of the sample environment above. Analyzes regarding demographic representation show that KONDA samples reflect the general demographic characteristics of the society at a generally consistent level. However, young people under the age of 29 and those over the age of 75 appear to be less represented due to higher rejection rates and difficulties in reaching them. When compared with TUIK statistics, it was seen that the 33-48 age group was included in the KONDA samples less than it should have been, and the number of men aged 49 and above was included in the KONDA samples more than it should have been. In general, it seems that nuclear families living in metropolises are relatively more represented due to ease of access, while those living in rural areas and the elderly are less represented. Since gender and age weightings alone cannot fully correct this situation, probability cluster sampling methods can be tried to ensure access to under-represented demographic groups.

### **1.5.6. Motivations for stating voting preference**

The motivation for voicing one's voting preference in surveys should be evaluated together with the extent to which the expressed preference is consistent with the votes cast at the ballot box. In countries where emotional polarization is high and opposition is costly, the motivation to vote involves social psychological processes that cannot be explained by statistics alone. It is known that the motivation to vote in surveys is lower in disadvantaged social groups, minorities and opposition groups. For this reason, "undecided" and "no answer" options are expected to be higher in these groups. After all, polls measure voting intent. While the declaration of intent is a relatively normative, rational statement, voting behavior is a behavior that also includes a kind of social identity declaration, especially in polarized societies where political competition is high. "Can not go to the ballot box" is a situation seen especially when party supporters with opposing political views are forced to vote for the same candidate.

Vote shifts occur between similar groups, not between opposing political groups. For example, the probability of a conservative who is hurt by high inflation and economic crisis to vote for the Ak Parti, YRP or MHP is much higher than the probability of voting for the opposing group. For example, YRP's 2.9 percent vote, which no polling company could accurately predict, can be seen as an attempt to change voting preferences without compromising social identity and





basic belonging. For this reason, it is more possible for alliances based on similar views and social affiliations to protect their votes. The Cumh ue Alliance, in which the parties in the alliance entered the elections under their own emblems, benefited from this advantage. However, while CHP and İyi Parti, which are based on a similar socio-demographic base in the Millet Alliance, ran in the elections with their own parties, four right-wing parties, all of which were to the right of the İyi Parti in terms of both the demographics of the voter base on which they relied and their political views, entered the elections under the CHP. Although this situation partially brought votes to the presidential candidate of the Millet Alliance, it seems that its contribution to the CHP remained minimal due to the difficulty of switching to the opposing identity ("not being able to vote"). Ahmet Davutođlu, Chairman of the Gelecek Party, one of the Millet Alliance parties, explains this as follows in his interview on August 1, 2023:

*"My last choice was to enter the election from the CHP lists. I'm being frank. Before that I tried everything. I said to the three parties, 'Let's enter together.' I made an offer to İyi Parti and said, 'These right-wing voters will not vote for CHP, let's be together.' "I tried all the ways."*

The vote share of other parties within the 25 percent received by the CHP, which was around 27 percent in the KONDA surveys before the election, is unknown, but it is a matter of curiosity whether the 2 points between 27 percent and 25 percent is the difference between the normative statement in the survey and "could not vote in the election" is among the unmeasurable factors.

The fact that the two main alliances entered the elections with traditional political polar leaders may have caused the reinforcement of traditional voting behaviors under the influence of polarization in the electorate and, as a result, the strengthening of the tendency to follow the majority and regression to the mean in swing voters.

Regardless of these speculations, even if the vote preference is unknown, KONDA's attempt to understand the political orientation of the voters by asking questions about political attitudes is a very appropriate attempt to understand the motivation of the vote preference. These questions, which are very useful for correctly interpreting possible patterns of change in voting preferences, serve as a kind of validity criterion.

### **1.5.7. Weighting**

As I mentioned above, weighting is a frequently used correction (calibration) method when there is a difference between the sample and population parameters whose statistics are known. Surveys generally use demographic weighting. However, due to the multi-party structure in Turkey, weighting based on political competition is widely used. Analyzes conducted within the scope of the biopsy showed that the effect of demographic weighting on the results and party vote rates was very limited. Since KONDA samples do not deviate much from the population parameters in terms of demographics, it is expected that the weighting will have a limited effect. However, demographic weighting cannot fully correct for possible bias in political preferences.

In Turkey, the relationship between demographic variables, especially education and age, and political attitudes is much higher than in Western democracies. Especially in Turkey,



education level is a factor that should be weighed carefully, as it is also a strong indicator of socio-demographic level. As the education level increases, the probability of voting for the CHP increases, while the rate of voting for the Ak Parti decreases. This ratio becomes even more complex when we look at the joint effect of education level with age and gender. For example, if the preference of low-educated women over the age of 70 to vote for the Ak Parti is maximum and older voters are under-represented in the polls, weighting cannot correct this situation.

One of the respected scientists on this subject. Prof. Dr. David Dutwin argues that one of the most important factors why pollsters were wrong in the 2016 elections was that education level was not taken into account. In the USA, whites with college degrees make up 30 percent of all registered voters, while whites with less than high school education make up 44 percent of voters. While only 38 percent of white college graduates voted for Trump, Trump's vote increased to 64 percent among white voters with a high school education or less. It was envisaged that calculating the results by weighting education among registered white voters would reduce the margin of error.

#### **1.5.8. Conclusion**

Diagnostic stage analyzes performed within the scope of biopsy provide rich data to objectively evaluate potential sources of error in KONDA surveys. The analysis process should continue by testing the alternative approaches suggested in future survey applications and the suggestions presented in the report with experimental/comparative methods. Testing the psychometric quality of questionnaires should remain an established practice. In this way, the survey format and content can be further developed. The results obtained after applying the recommendations made for the treatment phase, which we discussed in the analysis phase, in future surveys will also be a test for the validity of the evaluations made.

A survey company's ability to make accurate predictions depends on its ability to conduct and evaluate evidence-based, objective surveys. For this reason, it would be very useful for survey companies conducting political research to use the integrated scientist-practitioner model, which is accepted in psychology vocational education in Western countries, in the survey application and evaluation process. Each survey exercise should be designed as both an exciting forecasting venture and a scientific research project.

**KONDA**  
ARAŐTIRMA VE DANIŐMANLIK

Yıldız Mah. Őair Naifi Sok. No:2  
BeŐiktaŐ / İstanbul  
bilgi@konda.com.tr  
+90 212 275 17 66 (pbx)  
+90 212 275 17 68 (fax)