

Why do Polls Underestimate Mid-sized Parties in Indonesia?

Seth Soderborg

Harvard University, Cambridge, United States

Email: soderborg@g.harvard.edu*



p-ISSN: 2808-9529 (Printed)
e-ISSN: 2808-8816 (Online)

Jurnal Ilmu Sosial Indonesia (JISI)
<http://journal.uinjkt.ac.id/index.php/jisi>
VOL. 4, NO. 1 (2023)

Page: 1 – 13

Recommended Citation (APA Style):

Soderborg, S. (2023). Why do polls underestimate mid-sized parties in Indonesia? *Jurnal Ilmu Sosial Indonesia (JISI)*, 4(1), 1–13.
<https://doi.org/10.15408/jisi.v4i1.33365>

Available at:

<http://journal.uinjkt.ac.id/index.php/jisi/article/view/33365>

Article History:

Received 11 Maret 2023
Accepted 01 Mei 2023
Available online 28 Juni 2023

* Corresponding Author



This is an open access article under CC-BY-SA license
© Copyright Attribution-Share Alike 4.0 International
(CC BY-SA 4.0)

Abstract. Indonesian public opinion surveys consistently underestimate support for five mid-sized parties: PKB, PKS, NasDem, PAN, and PPP. This paper explores whether this consistent underestimation is the result of social desirability bias, late decisions or late mobilization, candidate-driven voter preferences, or sampling design. I find some evidence that parties with stronger mobilization networks and more candidate-centric strategies are more likely to be underestimated. A simulation study suggests that sampling design does not systematically disadvantage the parties in question. While social desirability bias is a possible cause of the downward bias, scholarly literature casts doubt on the social undesirability of expressed support for the parties. Further research should examine whether the types of voters who support the five underestimated parties might be harder to reach in surveys. Public opinion pollsters should consider noting in public statements that these mid-sized parties have tended to outperform their polls on election day.

Keywords: Public Opinion Surveys, Mid-Sized Parties, Social Desirability Bias, Pollster, Sampling Design, Indonesia.

Abstrak. Survei-survei opini publik Indonesia pra-pemilu secara konsisten menemukan dukungan kepada lima partai menengah: PKB, PKS, NasDem, PAN, dan PPP cenderung lebih rendah dibanding hasil aktual pemilu. Artikel ini menginvestigasi apakah prediksi yang lebih rendah dan konsisten ini merupakan akibat dari bias keinginan sosial (social desirability bias), pilihan elektoral atau mobilisasi yang telat, preferensi pemilih yang digerakkan oleh kandidat, atau desain pengambilan sampel. Naskah ilmiah ini menemukan bukti bahwa partai-partai dengan jaringan mobilisasi yang kuat dan strategi yang lebih berpusat pada kandidat cenderung kurang tercermin dinamika elektabilitasnya dalam survei. Studi ini menunjukkan bahwa desain sampling tidak secara sistematis berkontribusi atas kegagalan survei dalam memprediksi dukungan partai-partai menengah. Meskipun bias keinginan sosial adalah faktor yang memungkinkan penyebab angka elektabilitas yang terekam dalam survei cenderung rendah, literatur ilmiah sejauh ini meragukan bahwa bias tersebut merupakan faktor yang berpengaruh untuk kasus ini. Perlu studi lebih lanjut untuk mengkaji apakah sekelompok pemilih yang mendukung lima partai menengah yang mencatat bias dukungan lebih rendah di atas mungkin lebih sulit dijangkau dalam survei atau tidak. Lembaga-lembaga survei sebaiknya mempertimbangkan untuk menyampaikan kepada publik bahwa partai-partai menengah ini pada hari pemilihan cenderung mendapat dukungan lebih besar dibanding prediksi-prediksi survei pra-pemilu.

Kata Kunci: Survei Opini Publik, Partai Menengah, Bias Keinginan Sosial, Lembaga Survey, Desain Pengambilan Sampel, Indonesia.

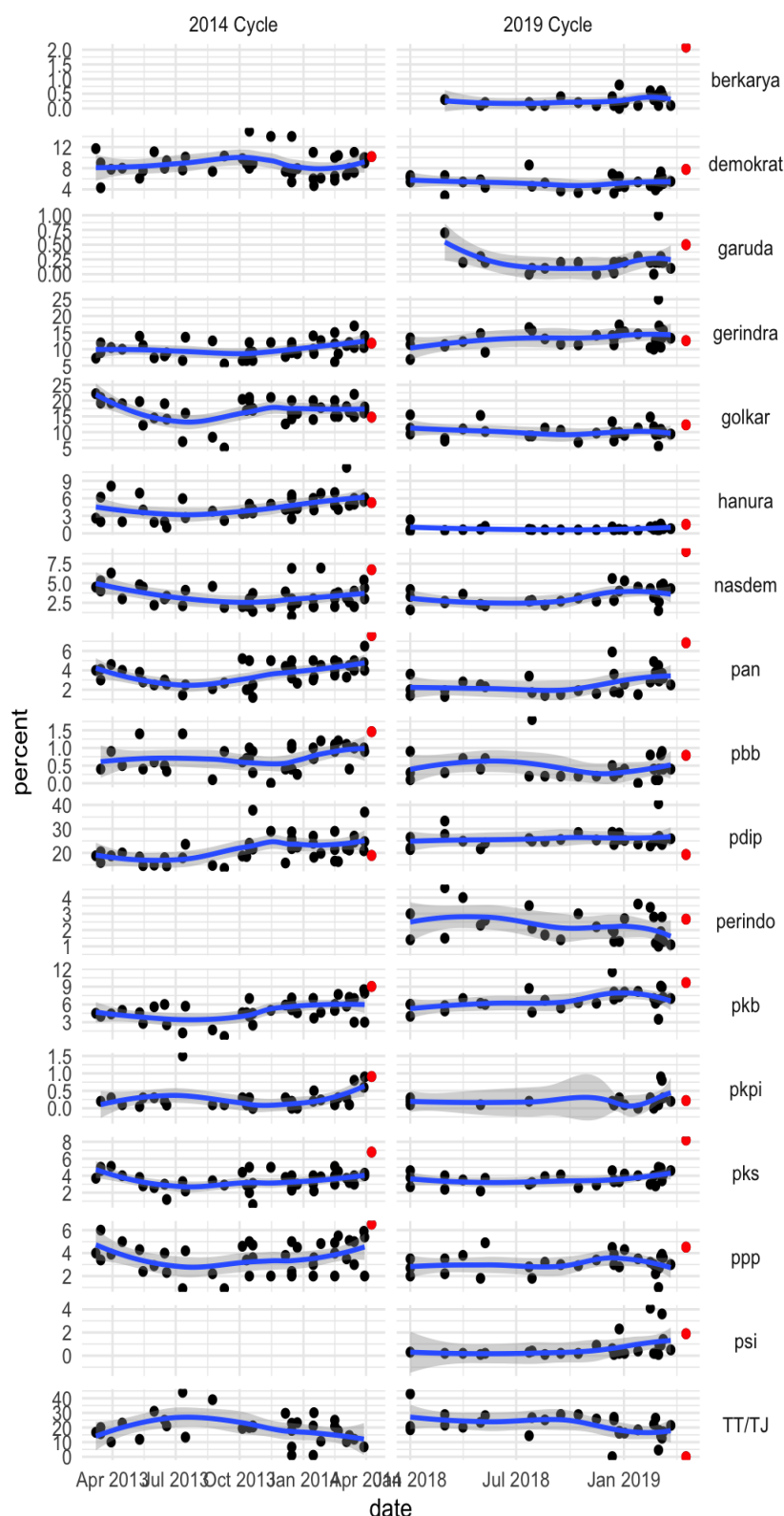
1. INTRODUCTION

Why do surveys of Indonesian voters consistently underestimate support for mid-sized parties while quite accurately estimating support for both larger and smaller political parties? In the 20 years since the beginning of polling in Indonesia, survey institutes have performed very well in absolute and relative terms, accurately estimating the results of presidential elections and correctly measuring the relative performance of the country's many political parties. This is not an easy task. The fact that pollsters consistently correctly identify the rank-ordering of parties only further adds to the question of why middling parties are usually underestimated—what mechanisms allow measurements of small and large party support to suffer only from ordinary survey error, while mid-sized parties all have consistently been measured with a downward bias?

To date, some scholarly and more general public attention has focused on one of the possible reasons: reluctance on the part of a specific kind of voter to answer survey questions. This explanation comes out of the fact that of the five parties who have been underestimated at least twice, four are Islamic parties. This explanation can surely account for at least some of the observed gap between survey results and election outcomes, but it cannot explain everything. In particular, an explanation focused on ideological voters' or opposition voters' reluctance to self-identify does not explain the consistent underestimation of support for the NasDem party, which is neither Islamic nor was, in the 2014 or 2019 election, a party of the opposition. In this paper, I consider the reluctant voter hypothesis as one of several possible explanations for the gap between survey results and election outcomes. In addition to the reluctant voter hypothesis, I consider four others: first, a tendency for genuine late-deciders to choose mid-sized parties, second, geographic correlations that interfere with the sampling design, third, survey model ballot design and the role of individual candidates, and fourth, gaps between the mobilizational capacity of the middling parties and others. A fifth explanation, class differences in reachability, requires data outside the scope of this paper and should be the subject of future work.

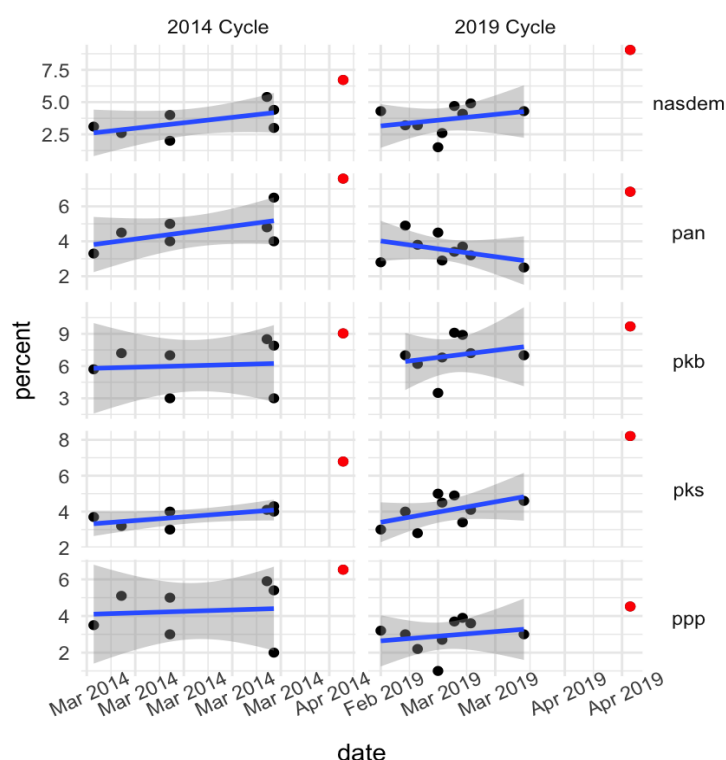
I find that social desirability bias is a less likely cause of the downward bias in estimated support than believed. I find some support for the hypothesis that late-deciding

voters (or late-mobilized voters) are more likely to vote for some parties than others. I also find some support for the hypothesis that candidate-centric parties are underestimated more than others. Finally, I present simulation-based evidence that sampling design does not interact with geographic vote distributions in to produce a downward bias in survey support. The possibility that important demographics are systematically hard to reach is not tested.



Middle Party Problems: There is a group of parties that has been underestimated in two consecutive elections.

When we zoom in on the parties with two consecutive underestimations, we see the following:



Each of the parties that has been consistently underestimated is mid-sized, with vote shares between five and ten percent of the national vote, and polling averages between three and seven percent of the vote. In addition to their size, some of the parties share important characteristics. Four of the five parties are Islamic parties. Two of the parties are often described as heavily reliant on candidate recruitment—PAN and NasDem. Three of the parties can draw on a mass organization (PAN and Muhammadiyah, PKB and PPP the NU). One of the parties has a disciplined internal cadre system—PKS. One of the parties has extensive media access—NasDem.

The polling error for PPP was not as serious in 2019 as it was in 2014. A few polls came within 0.2 percentage points of the right answer. A few polls also came very close to measuring PKB. For the others, PKS, PAN, and NasDem, note that no poll came within two percentage points of any of these parties' support, and that the bias was larger in 2019 than it had been in 2014.

This paper will look at the shared properties of these parties to examine whether any of them might play a role in the persistent underestimation of their support. I examine whether attitudes towards Islamic conservatism, candidate effects, or access to mobilization networks might explain the

gap. And I test whether the level and concentration of party support might interact with survey sampling designs in ways that systematically bias the results.

2. METHOD

Polling in Indonesia is a well-established industry, with a strong record of accuracy from a large group of pollsters. Using the multiparty error measure developed by Arzheimer and Evans, total survey error among Indonesian pollsters is lower than that measured in Germany and Brazil (Arzheimer and Evans 2014; Schnell and Noack 2014; Soderborg 2018).

Three elements of Indonesian politics create challenges for pollsters. First, there are many parties. Second, there are strong correlations between support for parties and class status that make it difficult to reach supporters of some parties. Some of these could result in moderate bias.

Observers of the regular bias sometimes conclude that Islamic party voters do not share their preferences with pollsters (detikNews 2012). This is connected to a more general concern that opposition voters might prefer not to indicate their preferences, at least when interacting with survey enumerators, a concern in many countries (Domínguez and McCann 1998). These concerns naturally lead to a hypothesis about the cause of consistent survey underestimates—social desirability bias. In this hypothesis, respondents who genuinely support at least some of the five parties prefer not to share their support with survey enumerators.

Does Social Desirability Bias Affect Surveyed Party Support? The most common concerns about social desirability bias relate to Islamic parties. In personal communications, Indonesian survey experts and politicians both express the belief that Islamic party voters are reluctant to share the fact of their support with survey enumerators. In other words, a social desirability bias is depressing measured support for the parties. Social desirability bias requires that respondents attempt to “minimize some socially undesirable characteristics,” with the values influencing the respondent coming from the respondent's internal value system, the interviewers' perception, or society as a whole (DeMaio 1984). Whether Islamic party voters should want to hide their preferences depends on the assumptions one makes about whether Islamic parties are generally disliked, disliked by survey enumerators, or whether support for Islamic parties is something a supporter should want to conceal for ego protecting reasons.

To the question of whether Islamic party support is something that the public might react negatively to, the literature on Islamic parties has

tended to emphasize a shift towards both moderations of the parties and acceptance of the parties into the mainstream. This likely minimizes the degree to which Islamic party support is something worth keeping secret. Tanuwidjaja argued that, in fact, the views held by Islamic parties that might once have been worth hiding are now held shared by non-Islamic parties (2010). This in fact continued a trend noted by Baswedan, who emphasized the “Islam friendliness” of nominally secular parties like Golkar evidenced by their recruitment of Islamic student association activists (2004). While both of these authors had vested interests in the claim that Islamic parties were becoming more mainstream, the claim is relevant for the question of whether Islamic party support is something respondents would prefer to keep hidden. By the mid-2010s, the more radical proposals associated with the PKS, the most conservative Islamic parties, had been shelved in favor of a moderation strategy (Tanuwidjaja 2012). If social desirability bias reduces respondents’ willingness to express support for Islamic parties, it is likely not because the parties are currently perceived as radical.

Polling Errors and Underestimation of Conservative Positions: Much of the literature on countries other than Indonesia that focuses on polling misses is interested in whether candidates’ levels of support are consistently underestimated when the candidate is considered more right-wing (Prosser and Mellon 2018), or whether polls consistently err when female or ethnic minority candidates are on the ticket. These are cases of social desirability bias creating a gap between the stated preferences given by survey results and the revealed preferences expressed in secret ballots.

Several cases in which this supposedly happened are more complicated than they seem. Donald Trump’s victory in the US presidential election happened in spite of the polls accurately predicting that he would win fewer votes than Hillary Clinton. In fact, US presidential election polling in 2016 was more accurate from a total survey error standpoint than it had been in 2012, and about as accurate as surveys in the US have been since 1972 (Kennedy et al. 2018; Silver 2018). Some of the more commonly assumed sources of bias work differently from expectations, when they do exist. Scholars have found that female candidates in the United States were usually underestimated, not overestimated (Stout and Kline 2011), while support for black American candidates was only overestimated (due to social desirability bias) under a narrow set of conditions (Stout and Kline 2015).

Another incident in which a conservative position won an unexpected victory against the polls was, in fact, not the case. Brexit polling produced

consistently mixed results, with many polls (especially those that incorporated some online respondents) estimating majority support for the “Leave” option rather than the “Remain” option (Gelman 2016). Moreover, some of the places where theories of right-wing underestimates would be expected to operate, like France in races involving the Le Pen family, do not exhibit this tendency. Right-wing populist Marine Le Pen’s support was overestimated, not underestimated, in both rounds of the 2017 and 2022 French presidential elections.

This does not mean that right-wing underestimates never occur, only that they are less common than is sometimes claimed. Among the countries where a right-wing figure claims to have been consistently underestimated by polling is Brazil, where Jair Bolsonaro’s late rise in the polls and decisive election victory in 2018 led to constant claims later on (as his popularity declined) that pollsters had never managed to accurately measure his support. The data relevant to this claim are straightforward. Most pollsters had Bolsonaro well behind his opponents until the final month prior to the first round, when support apparently coalesced around him (Schreiber 2022). However, even as the polls had Bolsonaro well ahead of his opponents, with 40 percent (Datafolha) or 41 percent (Ibope), with his nearest opponent at 25 percent, they did underestimate his 46 percent take in the first round (UOL Eleições 2018). Bolsonaro was, in fact, underestimated, though the five-point average gap was much smaller than Bolsonaro later claimed.

In the 2022 presidential election, Brazil’s pollsters again consistently underestimated support for then-incumbent President Jair Bolsonaro. Major pollster Ipec estimated a 51 percent Lula vote and 37 percent Bolsonaro vote in the first round—the final result was 48 – 43 (Cerqueira et al. 2022). Datafolha, which along with Ipec is usually considered Brazil’s leading pollster, estimated Bolsonaro support at 36 percent. In the second round, pollsters expected a 52 – 48 (Datafolha and Qaest) or 54 – 46 (Ipec) race (Gomes 2022). The second round results were 50.9 to 49.1. One surveyor that performed very well in the first round, Atlas Intel, performed much worse in the second round (53.4 – 46.6) and issued a public apology. Two races in a row, Brazilian surveyors had underestimated support for a right-wing populist figure.

Brazilian surveyors conduct large surveys, and many samples were quite large—with minimum samples of 2,000 and many surveyors reaching samples in excess of 4,000 respondents. The absolute performance was quite strong—Datafolha (the country’s leading pollster) was off by just one percentage point. What is notable, though, is that nearly every pollster was off in the same

direction in both rounds, and those pollsters whose first-round estimates were close the final second-round results tended to move a bit further away from the final results as they approached the second round. In the case of this large emerging democracy, polling consistently underestimated support for a controversial figure. The problem occurred across two consecutive presidential cycles (defeating efforts to mitigate it). Brazil provides a true example of an unambiguous polling bias against a right-wing candidate. Is this a common outcome?

Scholars who write about the global wave of populism have sometimes suggested that right-wing populist parties frequently win in surprise upsets. However, in places where right-wing parties have won, they have not necessarily been underestimated. Austria's conservative VPO is often moderately underestimated in polls, but its far right FPO is usually slightly overestimated. In Belgium, the far-right Vlaams Belang was underestimated by an average of four percentage points in the run-up to the 2019 federal elections, but was overestimated in 2014. The conservative but less extreme N-VA was accurately estimated in both 2014 and 2019. One exception to this trend was Germany, where the far-right AfD in surged in late polls to seven percent, and ultimately won twelve percent (Rebecca Staudenmaier 2017). In 2021, however, German pre-election polls slightly overestimated support for the AfD.

There are indeed notable incidents when polling errors led to underestimates of support for far-right candidates. However, when examined over a large number of countries and elections, there is little evidence to support the claim that, in general, right-wing or far-right candidates or parties are systematically underestimated. Rather, underestimates affect a variety of parties in different ways. Error, not bias.

Literature Conclusion: I take seriously Prosser and Mellon's conclusion in their review of polling errors: "there is little evidence that voters lying about their vote intention (so-called 'shy' voters) is a substantial cause of polling error. Instead, polling errors have most commonly resulted from problems with representative samples and weighting, undecided voters breaking in one direction, and to a lesser extent late swings and turnout models." Given the lack of polling error in presidential elections in Indonesia, and the reasons given in the literature

arguing that support for Islamic parties is a fairly mainstream attitude, and the existence of a polling bias for a non-Islamic party, I am inclined to agree with these authors and look for the source of the bias in the breaking of "undecided" voters in specific directions, sampling challenges, and the effects of last-minute mobilization.

3. RESULT AND DISCUSSION

This paper uses two datasets. The first is a collection of 82 top-line results from nationally representative public opinion polls. This dataset records the date of the poll, the pollster, the sample size, the percent of support for each party, and the share of respondents who did not answer the party support question.

The second dataset is a list of village-level vote returns from 61,385 of Indonesia's 82,881 villages. These were the full list of election returns available as of 30 May 2019, when the vote returns were collected from the KPU website by the author using a script written by Nicholas Kuipers.

3.1. The "*Tidak Tahu/Tidak Jawab* Challenge

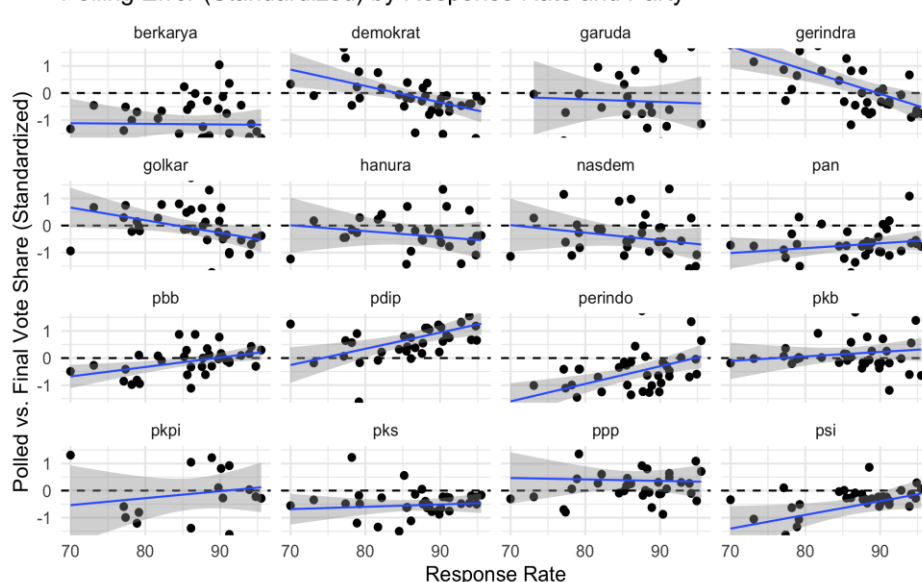
When survey design textbooks talk about "fundamental error," part of what they mean is that the questions on a survey and the context in which they are administered are not the same as the choices the surveys want to measure. Nor are they the same as the context in which those choices are made. In election surveys, the context difference is quite large. Where a surveyor interacts with a randomly chosen respondent on the phone or in the respondent's home, election day means a trip to the voting booth and a fixed set of choices. Even when a survey uses a ballot simulation—as most Indonesian survey firms do—there is always one option available to a survey respondent that is not available to the voter: "I don't know / I refuse to answer," combined in Indonesian electoral surveys as "*tidak tahu/tidak jawab*." Not voting or spoiling a ballot (*golput*) is different from refusing or not knowing, because unvoted ballots do not count. The simple gap between counting survey responses and counting votes ensures that in almost all cases, there will be some gaps between the survey estimated percentage of voters who will choose which party, and the final tally. Much of the survey gap is likely embedded in this group.

There are four main ways the TT/TJ respondents could move the realized result away from the survey results. First, it is at least theoretically possible that none of the TT/TJ respondents vote, while all people with a preference do. One reason to consider this possibility is that the rate of TT/TJ responses in late-cycle surveys often approximates the rate of non-voting by eligible voters. If this were happening, we would observe a specific and distinctive trend: all parties' vote shares would increase in proportion to their surveys vote intention. That would be the mechanical effect of removing the TT/TJ respondents from the sample, which is the mathematical equivalent of a situation in which a TT/TJ response is a perfect predictor of non-voting, while a party vote intention is a perfect predictor of voting for that party. However, this is not what occurs. The middling parties win a larger share of the votes than detected in the survey while the largest parties, especially PDI-P, win a smaller share.

This indicates that TT/TJ respondents should not be assumed to be non-voters. This also indicates that TT/TJ respondents do not ultimately vote for the parties in the same proportions as the rest of the survey responses. Instead, TT/TJ respondents may have specific partisan tendencies.

That TT/TJ respondents lean in certain partisan directions is the broad theoretical statement of a common view in the Indonesian politics literature. It is also the second explanation rooted in the behavior of respondents who do not indicate a preference. A common view among Indonesian pollsters is that supporters of Islamic parties, and supporters of candidates closer to conservative Islam, prefer not to reveal their preferences to surveyors. This hypothesis implies that PKB, PKS, PAN, and PPP should all consistently outperform their polls. And, in fact, they do. A second implication of this theory is that these four parties should receive higher shares of responses in surveys with fewer TT/TJ respondents. In other words, the presence of TT/TJ respondents hides Islamic party respondents such that Islamic party vote share should increase as the percentage of TT/TJ respondents declines. Moreover, surveys with smaller TT/TJ shares should be more accurate measures of Islamic party support, even after controlling for time (the share of TT/TJ respondents tends to reach its nadir in the surveys closest to election day).

Polling Error (Standardized) by Response Rate and Party



There is some support for this hypothesis, as support for two parties, PKS and PAN, does not change as the share of TT/TJ voters approaches zero. It would be reasonable to conclude that supporters of these parties may indeed be more likely to choose not to share their preferences with survey enumerators. Note that this does not give any indication of the reason for doing so. These voters might be concealing a preference they already have, or they might in fact be late-mobilized voters. The control for time rules this second explanation out partially, but it is hardly enough to fully remove that possibility from our list of explanations. That said, the lack of the same correlation for the other Islamic parties is telling. There is also an important demographic component to this particular party story. For a number of reasons, educated Indonesians are both more likely to support these two parties and more difficult to survey. They are also, on average, more likely to refuse to answer political questions. This adds to the complexity of interpreting the previous result. It is possible that supporters of these two parties suppress their political preferences when speaking with enumerators; it is also possible that it is simply very difficult to get supporters of these parties into samples.

So far, I have considered a straightforward hypothesis of TT/TJ respondents as perfect non-voters or perfectly proportional voters and concluded that it is incorrect. Some of the explanation for the survey gap must be related to the behavior of voters represented in surveys by the TT/TJ respondents. I have also considered the popular hypothesis that the TT/TJ category contains many supporters of conservative Islamic parties who are uncomfortable sharing their preferences with enumerators. This has some support in the cases of PAN and PKS, as we observe no changes in

support for these parties as the share of TT/TJ respondents changes. However, this approach has two main limitations. First, it is not clearly supported in the (less biased) cases of PKB and PPP and, second, it does not explain why NasDem has experienced the same gap.

One possibility is that there is something about the party label question that fails to interest respondents. This is quite relevant under Indonesia's open-list proportional system, where voters see the names of the party's candidates on the ballot and can directly choose them. In most surveys, the main question used to measure party support shows a model ballot with only the party logos, not the names of candidates as would appear on the final ballot. It might be the case that the absence of notable figures from the ballot reduces surveyed vote intention more for some parties than for others.

This is what Burhanuddin Muhtadi thinks occurred with NasDem in 2019 (Personal Communication). NasDem is perhaps the most aggressive recruiter in Indonesian politics, having built its entire apparatus around pulling popular candidates from other parties and offering them incentives and teams to come over to their side. It has been quite effective. If the lack of candidates on the ballot simulation matters for any party, it ought to matter for NasDem. In the survey described to me by Burhanuddin, Indikator found that when they used the candidate list, NasDem polled both better and closer to their final result than with party logos alone.

One way of measuring whether there is some connection between candidate list survey questions and accurate measures of party vote share would be to identify the parties that are more and less candidate-driven. This can be done by calculating the share of a party's votes won by the party label relative to the share won by specific candidates. A candidate-list focused hypothesis would imply that parties that rely more on individuals should be underestimated more by surveys than parties that rely less on individuals.

3.2. Candidate Effects.

We can also check the role of candidate effects on overall voter support by analyzing how support for parties changes in the period after candidate list announcement. The more candidate-driven a party's support, the larger the change in support should be observed when comparing party support shortly before and shortly after the announcement of the final candidate list.

To do this, I compare the polling average of parties in the three months before and after the release of the final candidate list. This approach assumes that any effects of candidate names in polling will occur after the release of the final candidate list, rather than in the period following the release of the initial candidate list (the DCS).

Party	2014 pre-DCT	2014 post-DCT	2019 pre-DCT	2019 post-DCT
TT/TJ	20.1	18.8	22.1	18.2
DEMOKRAT	7.0	6.8	6.1	4.8
GERINDRA	9.2	10.1	15.1	13.2
GOLKAR	15.0	16.5	9.6	9.9
HANURA	4.3	5.7	0.6	0.8
NASDEM	3.2	3.5	2.6	3.8
PAN	3.7	4.0	2.2	3.1
PBB	0.5	0.9	0.7	0.3
PDIP	22.8	19.9	25.2	27.5
PKB	5.7	5.7	6.7	8.0
PKPI	0.1	0.2		
PKS	2.9	3.6	3.4	3.4
PPP	3.6	4.3	2.6	3.6
BERKARYA			0.1	0.2
GARUDA			0.1	0.1
PERINDO			2.4	2.4
PSI			0.3	0.6

While not dispositive, a large change would be consistent with the idea that a party's support is closely related to candidates, and that their brand is a smaller part of their appeal than for other parties. While it would not account for the polling gap as a whole (since this approach uses polling results), the presence of a large jump would be a sign that candidate-related effects could be one source of a gap between polled and realized support.

In 2014, "don't know" responses declined slightly after the release of the candidate list, and Hanura, Golkar, and Gerindra experienced the largest increases in support. Hanura's increase of 1.5 percentage points is notable because it represents a 30 percent increase in support for the party. PDI-P experienced a three percentage-point decline. PKS and PPP registered moderate increases after the release of the candidate list. Notably, however, the two parties currently viewed as the most candidate-driven—PAN and NasDem (Indikator Politik 2013)—did not experience a national jump after the release of the 2014 DCT.

The pattern was different in 2018. In that year, the release of the DCT corresponded to a larger four percentage-point drop in TT/TJ respondent share, but party support was more variable before and after the list was made public, relative to 2014. Support for Demokrat and Gerindra was lower than before the list was released. PDI-P support was higher post-release,

even further away from the final result the party obtained. Of the parties that would ultimately be underestimated in 2019, all received higher shares of respondent votes after the candidate list was released. NasDem polled 1.5 times higher, PAN had increased its support by 30 percent, PKB was polling within one percentage point of its final result, and PPP was up by nearly 50 percent relative to previous polls. PKS registered no change, with both pre- and post-list percentages of 3.6—well below their final tally.

Since many of the parties that were underestimated in 2019 experienced large increases in measured support following release of the candidate list, we should consider candidate effects as part of the reason for surveys' underestimate of these parties' support. The pattern of post-list changes in 2014 does not align as neatly with this hypothesis, as the largest changes occurred among parties that were not underestimated. We therefore cannot conclude that candidate effects account for most of the underestimates. We also, however, cannot discount candidate effects as a contributor to the underestimate. Surveyors should make note of parties that experience large increases following release of the candidate list, and be prepared for those parties to exceed expectations (unless the party increasing is PDI-P).

3.3. Mobilization

Finally, another cause of differences between survey results and election results might be due to mobilization effects. The final surveys of the election season are generally in the field until a bit more than one week before the election, leaving a week or more between the bulk of final surveys and election day. In that week, parties and candidates are at their busiest. Candidates make their final push in this time. And the notorious dawn attack, in which candidates spread tens of thousands of envelopes of cash, occurs in this period after the final survey. This means that no surveys capture the impact of these frenzied late efforts.

If any of the parties are better than their rivals at mobilization, then this could account for some of the gaps between surveyed vote intention and realized vote share. Parties with more effective mobilization apparatus might bring more TT/TJ respondents into their fold, or they might capture a portion of the vote intention earlier given to another party, or some combination of both. This period of invisibility is a source of fundamental error that cannot easily be dealt with.

One simple way of accounting for late mobilization would be to assume that trends in the final months of the campaign season will continue until election day. Of the underestimated parties in 2019, PKS, PKB had positive poll trends in the final

three months, but NasDem, PAN and PPP did not. Simple trend extrapolation would not have eliminated the polling error. This mobilization measure can only detect mobilization in the period of time when surveys are being fielded. Pushes in the final two weeks are not detectable.

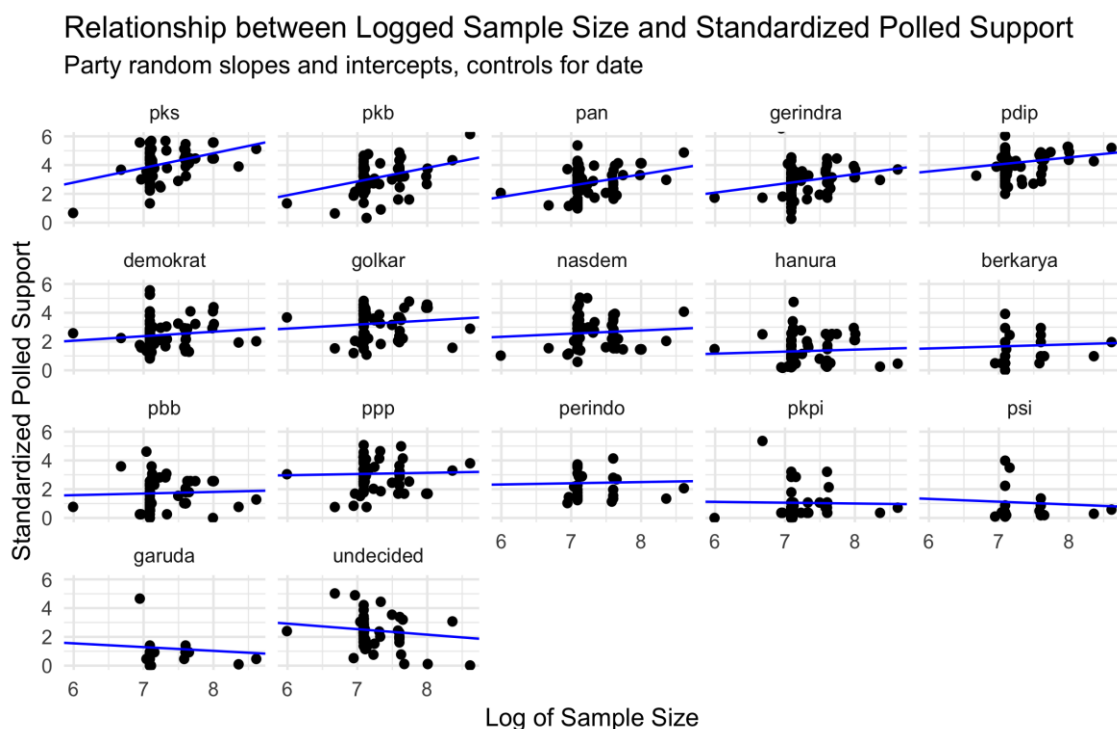
A future study focused on late mobilization could examine measures of local electoral resources, like the presence of a large number of provincial DPRD members or density of social organization membership, correlate with survey overperformance. This would be one way to account for parties' differential capacity to mobilize voters at the end of the electoral cycle.

A large portion of the survey bias for the five parties probably comes from TT/TJ respondents. There is evidence that eventual PAN and PKS voters do not exit the TT/TJ condition when surveyed, but it is unclear whether this is related to preference hiding, late mobilization, or difficulty reaching respondents. Four of the five underestimated parties experienced large increases in support after the 2019 candidate list was released. This is one sign that candidate effects may be related to the consistent underestimate of certain parties. No single cause appears to have been responsible for the survey bias, but candidate-centrism is a sign that bias might occur. So, too, is a lack of correlation between the percentage of TT/TJ respondents and party voters.

3.4. Problem of Sampling Design

One reason the surveys are consistently able to measure the relative strength of the parties is also closely related to statistical power. While the actual point estimates of party support are somewhat underpowered, the relative sizes of any given pair of parties can be more than adequately measured using surveys of the roughly 1,200 respondent size that most firms use. Taken together, this means that, from statistical principles alone, we should expect that the relative ranking of the parties is quite accurate, even though arithmetic margin of error contains, in some cases, the full range of support of both parties in a comparison pair. At the same time, the overall levels of support are more difficult to measure.

Another reason to consider the role of sampling design in affecting measured support for parties is that there is, for some parties, a relationship between survey sample size and polled support.



In fact, the three parties with the strongest relationship between sample size and polled support are among the five underestimated parties. NasDem and PPP, however, do not display this relationship. Gerindra and PDI-P also performed better in larger surveys, but were accurately estimated and overestimated, respectively. Some of the correlation may be due to surveyor house effects. However, the existence of any sample size correlation at all, even after controlling for timing, is worth paying attention to. If samples of different sizes produce consistently different results for specific parties, then there is good reason to suspect that sampling design is interacting in some way with patterns of support to introduce error.

In the range of votes won by the middle parties, there is the possibility that patterns of vote concentration in specific areas are uneven enough that there is systematic underestimation because the primary sampling units generate integer bias. Integer bias means that the expected number of a party's voters in a given primary sampling unit might consistently be on the lower half of the fraction range of integers, such that when the number is rounded, it goes to the next lower integer rather than the next higher integer. If this is the case in enough locations, it could produce a downward bias. Normally, overestimates in one place might be balanced by underestimates in another. However, with the right patterns of geographic concentration it is possible for the downward integer rounding to predominate. This could occur when support is concentrated in a few areas but never reaches a very high level. Mid-sized parties with concentrated support might be especially vulnerable to this bias. Key to this is the fact that samples are limited to 10

respondents per primary sampling unit. Larger PSUs are less likely to experience this bias, but more likely to miss areas of high (or low) support, thus trading one source of potential bias for another.

3.5. Vote Concentration and Sampling Issues

To test whether concentration might be interacting with sample size to produce underestimates, I first calculated the concentration of votes at the village level using the Hirschman-Herfindahl index of market concentration, which calculates the degree to which votes for a given party are contributed by a few villages, or are instead the result of relatively even performance across many villages.

In 2019, three parliamentary parties were far more concentrated by this measure than the others: PKB, PKS, and PPP. Of these, PKB was twice as concentrated as PKS, and four times more concentrated than the average party. These three parties are also among the set of parties that is consistently underestimated.

While concentration measures militate mildly against the integer error hypothesis, they do provide additional support for explanations related to local mobilization capacity. Higher levels of concentration imply the presence of social and organizational networks that feed the parties. PKB, which is by far the most concentrated parliamentary party, relies on geographically concentrated NU chapters. PKS, which is only half as concentrated as PKB, also leverages tight networks. NU's other parties, PPP and PBB, are also among the above median concentrated party.

The problem with concentration as a straightforward cause of possible survey error is that three of the most concentrated parties, PKPI, PSI, and PBB, were accurately measured by surveys. These are small parties that did not reach the parliamentary threshold, but it is still notable that higher concentration did not automatically produce survey error. The integer bias hypothesis predicts that smaller parties might still be estimated correctly, but it is worth noting that while many of the underestimated parties were concentrated, it was not the case that all concentrated parties were underestimated, nor were all underestimated parties concentrated. NasDem was one of the least concentrated of all parties. And PAN was below median levels of concentration.

3.6. Integer Bias

The integer hypothesis is that under certain conditions, patterns of vote concentration can interact with the size of primary sampling units to moderately underestimate support. A global measure of vote concentration might not be the right tool for measuring this kind of concentration pattern. A better approach would directly model the sampling design in interaction with the underlying population.

To do this, I first ran a simple test where I rounded realized village level vote share and compared the rounded result to the actual realized votes. This tends to underestimate votes, but might do so differentially for different parties. If integer error is an issue, it might show up here. Interestingly, the parties most affected by rounding are PDI-P and PKB. Gerindra, Golkar, and NasDem come next. A simple integer error based on the realized vote share cannot account for the underestimate issue. If it did, PDI-P would be underestimated rather than overestimated. PKB's high integer error rate is consistent with the hypothesis, but it is the only one of the underestimated with high integer error. Once again, this test

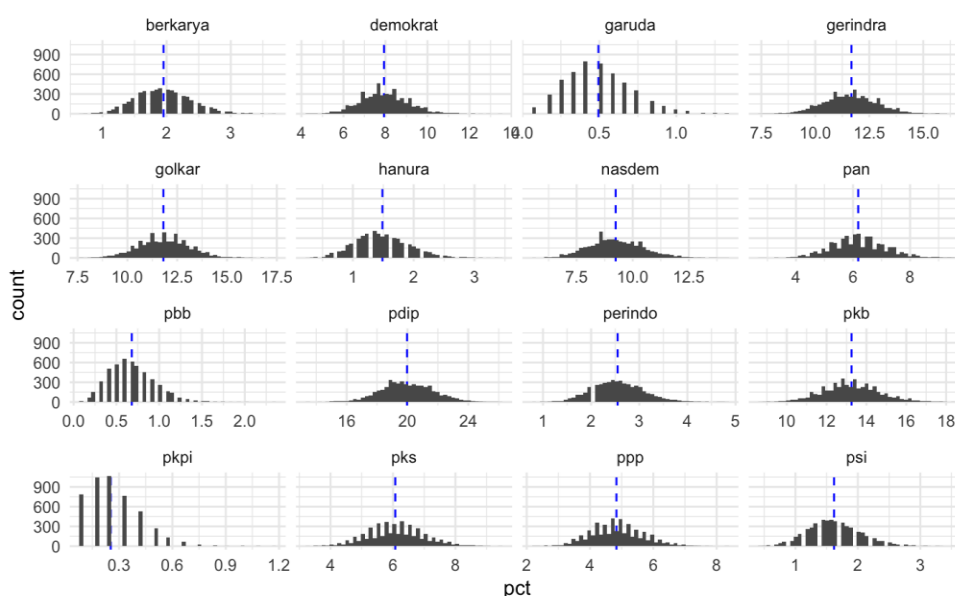
A better test of integer error uses repeated resampling of data to determine whether the pattern of realized votes is in some way interacting with sampling design to harm the survey. To do this, I use a dataset of realized election returns aggregated up to the village level. I exclude votes cast abroad, as these would not be included in a survey. For this analysis, the baseline value against which sampled results should be evaluated is the baseline set by

the realized vote dataset, not the official election returns for the whole country. With this simulation, I can directly measure the interaction of the two-stage sampling process with the underlying population. The question to be answered here is whether the two-stage sampling process used in most surveys interacts in some way with patterns of support to systematically over- or underestimate support for specific parties.

For this procedure, I imitate the sampling design used in most face to face surveys by randomly selecting 120 villages, with odds of a village being included set as the village's share of the total national vote (equivalent to a population weighting under assumptions of equal turnout propensity). I then sample 10 lines in each selected and calculate the share of votes received by each party nationally. This process is repeated 5,000 times.

The simulation differs from the survey method in a few important ways. First, it is not a survey. Second, it does not further subdivide the village into RTs, as is done in the real surveys, nor does it incorporate a gender quota. Post-survey weighting is not part of this method. Finally, the dataset being used does not include results from every TPS. This means that the baseline used to evaluate the simulations' accuracy is different from the actual national result. It is possible that the missing data might be correlated with characteristics that would matter for the outcome. Despite these differences, this approach is a useful way to check whether properties of sampling design might affect the ability of surveys to detect support for specific parties.

The simulation approach finds that parties are well-measured by this sampling design. Mean party support across the simulations is extremely close to the realized vote share (relationship of actual CI to reported margin of error). Parties exhibit largely normal distributions.



There is some mild skewness visible in the figure above. I check whether skewness measures are higher among parties that are over- or underestimated. The results show that the levels of skewness are well within the range of normally distributed data. Higher levels of skewness are not concentrated among parties that were underestimated. Kurtosis measures, which check whether the distribution is driven more by extreme events and can indicate whether there is a higher chance of an individual sample missing the true value, are also consistent with normal distributions and not correlated with a party being underestimated. There is little evidence to support the idea that integer error is at work when sampling from the realized vote dataset. If the sample size is increased to 2,000, the only change that occurs is a narrowing of the standard deviation of results for all parties.

Neither the 1,200- nor the 2,000-respondent simulation contains the survey results within two standard deviations of the mean. From simulation alone, it is not possible to obtain the results actually observed in the surveys at the sample sizes used in real surveys.

A mechanical integer error cannot alone explain the pattern of party underestimates. A few things make the test described in this section different from real-world surveys. First, there are no TT/TJ when sampling from the realized votes. Second, there are no respondents who refuse to be interviewed. Third, since probability weights for villages were assigned based on vote totals, there are no differences between the population weights that would be used in a survey and the actual weight of the village in the final vote total—another possible source of error. Despite these differences, the lack of bias in the simulated samples indicates that the challenge of measuring support for the five underestimated parties lies in the realm of non-response, late mobilization, candidate effects, or concealment.

These measures show that because of their patterns of concentrated support, a few parties are at somewhat greater risk of being underestimated than others. However, this simulation procedure does not support the claim that integer error or an interaction between PSUs and the sampling procedure consistently biases surveys against the underestimated parties.

4. CONCLUSION

This paper explores some possible causes of a consistent survey bias affecting five Indonesian political parties. In it, I consider causes rooted in voter preferences and activation—concealment, late mobilization, and a preference for candidates over

parties—as well as causes rooted in sampling design.

A large portion of the survey bias for the five parties probably comes from TT/TJ respondents. There is evidence that eventual PAN and PKS voters do not exit the TT/TJ condition when surveyed, but it is unclear whether this is related to preference hiding, late mobilization, or difficulty reaching respondents. Four of the five underestimated parties experienced large increases in support after the 2019 candidate list was released. This is one sign that candidate effects may be related to the consistent underestimate of certain parties. No single cause appears to have been responsible for the survey bias, but candidate-centrism is a sign that bias might occur. So, too, is a lack of correlation between the percentage of TT/TJ respondents and party voters.

There is a notable correlation after controls between sample size and polled party support. While it does not consistently map onto the underestimate parties, it does signal that under some circumstances, sampling design may be affecting polled support. The sample size-support correlation is not the result of larger samples correcting downward integer bias. Evidence from simulated resampling following the procedures used in most Indonesian surveys suggests that integer bias does not systematically affect any party.

The evidence suggests that the underestimates are related to respondents. The underlying cause may be respondent behavior, late mobilization of sincere “don’t know” voters, or concealment. There are few reasons, however, to believe concealment is common. There is also the possibility that differential response rates, especially the low response rate among more educated voters, might be an important source of bias. The data available for this paper do not allow an analysis of whether the more-educated respondents who vote for parties like PAN and PKS are harder to reach.

Recommendations: Good surveyors should consider not only the facts in their surveys, but also the ways that media might create unwarranted narratives on the basis of survey results. For example, in the 2017 DKI election, surveyors never gave a majority to Ahok, but reporters described his 41 percent result as making him an easy winner. Similarly, in 2019, the focus of most coverage of Islamic parties was whether they would cross the parliamentary threshold. Now that there have been two consecutive elections in which surveys underestimated these parties, it would be worthwhile for surveyors to communicate that polled results for five parties have consistently underestimated their final level of support. Media treats nearly all surveys as predictions, despite the

fact that surveys are not predictions. One way to handle this, and to avoid being blamed for claims that one has not made, is to separately discuss the results of the surveys and what those might mean for final election tallies. At minimum, a note that, for example, a PPP polling three percent is likely to cross the threshold, should probably appear in the slides.

Another option for pollsters would be to enter the prediction game. That is, to present both the results of the survey and their implications for some model of how survey results translate into election outcomes. In doing this, one simple adjustment might be to add between 1.5 and three percentage points to the underestimated parties while pulling two percentage points from PDI-P. This is the least complex adjustment, but it has an excellent track record. More complicated adjustments could assign portions of the TT/TJ votes to the various parties differentially, using either historical correlations between TT/TJ share and polling bias, or through a demographic model of TT/TJ voters. An additional approach might attempt to measure late mobilization by extrapolating forward time trends from the final month of polling. If a party improved over that month, it might be reasonable to assume that improvement reflects effective mobilization, and extend the trend forward to election day. Each of these approaches involves moving from the realm of present measurement to that of prediction. Prediction has its own challenges and its own risks. Despite these risks, it may be worth pollsters engaging in some prediction so that they can more easily demonstrate the differences between a survey result and a prediction about the future.

As many pollsters suspect, the balance of the evidence suggests that the underestimate challenge is connected more to respondents than to other aspects of the survey process. Given this, it would be worth exploring which questions types generate the highest share of partisan identification or partisan vote intention. Many pollsters already ask multiple versions of this question, and the formats used in the voting intention follow international best practices in limiting the appeal of outside options. It remains notable that as the election grows close, the rate of TT/TJ responses tends to converge towards the share of the electorate that does not vote. This is not to say that all TT/TJ respondents do not vote—indeed, this paper has presented evidence that they tend to show up for some parties and not for others—but it is meant to flag the fact that Indonesia probably does not have the partisan intent under-identification noticed in the Americas. It may be worth further exploring probe format questions, in which voters who initially decline to answer are encouraged to make a choice. The differences between an encouraged and an un-

encouraged format question would be a powerful additional tool for identifying where TT/TJ respondents and their population equivalents are going on election day (Baker and Renno 2019).

Although this paper had less access to the data necessary to demonstrate candidate effects at the respondent level, the combination of post-DCT jumps, correlation between candidate-driven voting and survey underestimates, and accounts of surveyors whose constituency-level surveys with candidate names outperformed constituency surveys without candidate names, all suggest that candidate names should be used as early as possible in the survey cycle. Moreover, questions with candidate names should be reported over questions without them whenever possible. If there are significant differences within surveys with and without candidate names, those ought to be reported.

In comparative perspective, Indonesian pollsters perform very well. The downward bias in estimated support for the five parties discussed in this paper has not led to inaccurate ordering of parties' relative performance. It is important to keep this in mind. PKB, PKS, NasDem, PAN, and PPP are likely to continue to outperform their polls, but their relative strength is likely accurately measured already. If any of these parties jumps after the release of the candidate list, we can expect that party to outperform its polls. If any of these parties' support remains constant while the share of TT/TJ respondents changes, we can expect that more of the TT/TJ respondents will end up with them. With or without these signs, pollsters should communicate the difference between a survey result and a prediction.

5. REFERENCES

- Arzheimer, Kai, and Jocelyn Evans. 2014. "A New Multinomial Accuracy Measure for Polling Bias." *Political Analysis* 22(01): 31–44.
- Baker, Andy, and Lucio Renno. 2019. "Nonpartisans as False Negatives: The Mismeasurement of Party Identification in Public Opinion Surveys." *The Journal of Politics* 81(3): 906–22.
- Baswedan, Anies. 2004. "Political Islam in Indonesia: Present and Future Trajectory." *Asian Survey* 44(5): 669–90.
- Cerqueira, Carolina, Daniel Reis, Danilo Moliterno, and Marcello Sapio. 2022. "Pesquisas Erram e Divergem Dos Resultados Das Urnas." *CNN Brasil*.
<https://www.cnnbrasil.com.br/politica/resultados-das-urnas-divergem-de-pesquisas-eleitorais/>.
- DeMaio, Theresa J. 1984. "Social Desirability and Survey Measurement: A Review." In *Surveying*

- Subjective Phenomena*, eds. Charles Turner and Elizabeth Martin., 257.
- detikNews. 2012. "Partai Islam Disurvei Merosot, PPP: Itu Kesalahan Besar!" *detiknews*. <https://news.detik.com/berita/d-2062225/partai-islam-disurvei-merosot-ppp-itu-kesalahan-besar> (June 12, 2023).
- Domínguez, Jorge I., and James A. McCann. 1998. *Democratizing Mexico: Public Opinion and Electoral Choices*. JHU Press.
- Fernandes, Arya. 2018. "Politik Identitas Dalam Pemilu 2019: Proyeksi Dan Efektivitas." *CSIS*.
- Gelman, Andrew. 2016. "Brexit Polling: What Went Wrong? | Statistical Modeling, Causal Inference, and Social Science." *Statistical Modeling, Causal Inference, and Social Science*. <https://statmodeling.stat.columbia.edu/2016/06/24/brexit-polling-what-went-wrong/> (June 11, 2023).
- Gomes, Israel. 2022. "Quais Pesquisas Erraram e Acertaram Resultado Do 2º Turno Entre Lula e Bolsonaro?" *O Povo*. <https://www.opovo.com.br/eleicoes-2022/2022/10/31/quais-pesquisas-erraram-e-acertaram-resultado-do-2-turno-entre-lula-e-bolsonaro.html>.
- Indikator Politik. 2013. "Efek Popularitas Calon Legislatif Terhadap Elektabilitas Partai Jelang Pemilu 2014." Presented at the Jakarta. http://www.indikator.co.id/uploads/20131014105703.20130730171101.INDIKATOR_Rilis_Survei_Efek_Calon_Terhadap_Partai.pdf.
- Kennedy, Courtney et al. 2018. "An Evaluation of the 2016 Election Polls in the United States." *Public Opinion Quarterly* 82(1): 1–33.
- Mietzner, Marcus, Burhanuddin Muhtadi, and Rizka Halida. 2018. "Entrepreneurs of Grievance: Drivers and Effects of Indonesia's Islamist Mobilization." *Bijdragen tot de taal-, land- en volkenkunde / Journal of the Humanities and Social Sciences of Southeast Asia* 174(2–3): 159–87.
- Prosser, Christopher, and Jonathan Mellon. 2018. "The Twilight of the Polls? A Review of Trends in Polling Accuracy and the Causes of Polling Misses." *Government and Opposition* 53(4): 757–90.
- Rebecca Staudenmaier. 2017. "German Election: Poll Puts Populist AfD in Third Place." *Deutsche Welle*. <https://www.dw.com/en/german-election-far-right-afd-third-in-latest-poll/a-40551590> (June 11, 2023).
- Schnell, Rainer, and Marcel Noack. 2014. "The Accuracy of Pre-Election Polling of German General Elections." *Methods, Data, Analyses* 8(1): 5–24.
- Schreiber, Maria. 2022. "Pesquisas Eleitorais 'erraram Tudo Em 2018' Como Diz Bolsonaro?" *Estado de Minas*. https://www.em.com.br/app/noticia/politica/2022/09/26/interna_politica,1398068/pesquisas-eleitorais-erraram-tudo-em-2018-como-diz-bolsonaro.shtml.
- Silver, Nate. 2018. "The Polls Are All Right." *FiveThirtyEight*. <https://fivethirtyeight.com/features/the-polls-are-all-right/> (June 11, 2023).
- Soderborg, Seth. 2018. "Predicting Multiparty Elections with Limited Data: A Constituency-Level Model of Indonesian Elections." Working Paper.
- Stout, Christopher T., and Reuben Kline. 2011. "I'm Not Voting for Her: Polling Discrepancies and Female Candidates." *Political Behavior* 33(3): 479–503.
- . 2015. "Racial Salience, Viability, and the Wilder Effect: Evaluating Polling Accuracy for Black Candidates." *Public Opinion Quarterly* 79(4): 994–1014.
- Tanuwidjaja, Sunny. 2010. "Political Islam and Islamic Parties in Indonesia: Critically Assessing the Evidence of Islam's Political Decline." *Contemporary Southeast Asia* 32(1): 29–49.
- . 2012. "PKS in Post-Reformasi Indonesia: Catching the Catch-All and Moderation Wave." *South East Asia Research* 20(4): 533–49.
- UOL Eleições. 2018. "Pesquisas Eleitorais: Intenção de Voto No 1 Turno." *UOL Eleições 2018*. <https://noticias.uol.com.br/politica/eleicoes/2018/pesquisas-eleitorais/brasil/1-turno/>.