Toward the end of 2021, the second winter of the coronavirus disease 2019 (COVID-19) pandemic, the number of COVID-19 cases began to increase dramatically in much of the world. In many countries, however, the majority of the adult population has received at least one dose of vaccination against severe acute respiratory syndrome coronavirus 2, which is seen as the main instrument to fight the pandemic. But how accurate are our estimates of COVID-19 vaccination coverage? There are examples recently of population surveys showing higher vaccination coverage than the official statistics indicate. In Germany, for example, the Robert Koch Institute (RKI) reported higher vaccination coverage on the basis of population surveys compared with what was being estimated on the basis of direct data coming from vaccination centers, hospitals, and physicians. The difference reported by the RKI amounts to 10 to 12 percentage points (Robert Koch-Institut 2021a). Other sources indicate similar or even higher discrepancies (Norddeutscher Rundfunk 2021).

Namely, the RKI’s official Digital Vaccine Coverage Monitor (Digitales Impfquotenmonitoring) collects data from vaccination centers, hospitals, and physicians’ practices. The RKI also collects data in the form of a rolling cross-sectional population survey called the Covim-Surveys (Robert Koch-Institut 2021a). At the beginning of October 2021, the RKI corrected its estimated vaccination coverage for adults from about 75.6 percent (fully vaccinated) to about 80 percent. This correction was based on new survey data coming out of the seventh Covimo survey, while the previous estimate had been based on the reporting of health care workers in the Digital Vaccine Coverage Monitor.
Monitor. This would translate to about 3.5 million more fully vaccinated adults than previously believed (Die Bundesregierung 2021). The reason for this discrepancy was cited as having to do with incomplete data coming from vaccination centers, hospitals, and physicians (Tagesschau 2021), an explanation echoed by the federal health minister at the time, Jens Spahn (Die Bundesregierung 2021). The new vaccine coverage figure based on the survey data was presented as more reliable and a more accurate estimate of the true vaccination coverage.

In the present study we argue, however, that this discrepancy is likely due to the sensitive nature of the question of one’s vaccination status and that social desirability and other problems related to question sensitivity are responsible for biasing vaccination estimates on the basis of survey data upward. It is a well-established empirical fact that when it comes to sensitive or socially loaded questions in surveys, misreporting by survey respondents occurs: negatively connoted or socially undesirable behaviors, attitudes, or traits are underreported, and positively connoted or social desirable ones are overreported (Krumpal 2013; Tourangeau and Yan 2007; Wolter 2012). Using an experimental survey setup, we investigate whether this also holds for survey questions on vaccination status. This is done by comparing estimates obtained from asking respondents directly about their vaccination status (as in the previously cited surveys) with estimates obtained using the so-called item count technique (ICT), a survey method used to provide complete anonymity to respondents to report potentially sensitive information (Droitcour et al. 1991; Glynn 2013). Using data from a large-scale experimental online survey in Germany (n = 2,641), we show that estimated prevalence rates of having received at least one vaccination against the coronavirus are indeed significantly lower, with a difference of about 10 percentage points, when asking indirectly using ICT. Hence, our conclusion is that conventional survey questions on COVID-19 vaccination status yield (upwardly) biased results and should be interpreted with care.

Although the motivation for this study was taken from a concrete example in Germany and with a focus on question sensitivity and social desirability bias, the findings surely have implications for other parts of the world, potentially wherever survey data are used to estimate vaccine coverage. In fact, institutions in many countries use surveys to obtain vaccine coverage estimates by asking individuals directly about their vaccine status, such as the U.S. Census Bureau (2021), the Centers for Disease Control and Prevention (n.d.), Statistics Canada (2021), and the CoviPrev survey program in France (Santé Publique France 2021). Research in the United States has also shown that surveys on COVID-19 vaccination coverage may be severely distorted by issues other than social desirability bias, like the unrepresentativeness of (large online) samples and/or poor survey quality (Bradley et al. 2021). A second study also demonstrated upwardly biased differences of 7 to 8 percentage points between administrative vaccination figures and survey estimates (Nguyen et al. 2021). On the other hand, another study from the United States showed that survey estimates are fairly close to administrative data. However, this study also showed that “when polls differed from the CDC [official administrative] rate, they often came in higher rather than lower” (Hatley 2021). Our study adds to this literature and points to a possible bias that has not been addressed yet. Furthermore, we present an application of ICT on a survey question on COVID-19 vaccination for the first time, as we are unaware of any other study existing on this.

Background

COVID-19 vaccination has turned into a hotly debated topic not just in Germany but in many other countries as well. On one hand, vaccinations are widely seen as the only way out of the pandemic: they help protect oneself and others (especially those who are unable to be vaccinated themselves), and they reduce symptomatic and serious cases of COVID-19 so that the virus would lose its potency if everyone was vaccinated (Robert Koch-Institut 2021a). This is additionally pressing as patients with COVID-19 put a strain on health care workers, force planned procedures and operations to be postponed, and, in countries with public health care systems, drain taxpayer money. The magazine Der Spiegel noted that in Germany in November 2021, 90 percent of those being treated for COVID-19 in intensive care were unvaccinated and that each intensive care stay can cost upward of €100,000 (Friedmann and Winter 2021). According to official RKI data for autumn 2021, COVID-19-related hospitalization incidence rates were about seven times higher for the unvaccinated population aged 18 to 59 years, and about 5 times higher for those 60 and older (Robert Koch-Institut 2021b).

On the other hand, the first approved COVID-19 vaccines were developed in an unprecedented time frame. It took less than one year for the vaccine from BioNTech and Pfizer to be approved for emergency use, whereas previous vaccines normally needed at least several years before approval (Ball 2021). Furthermore, two of the vaccines are of the highly innovative messenger ribonucleic acid type (from BioNTech/ Pfizer and Moderna), which differ from conventional vaccines. For these reasons, many individuals are skeptical and distrustful of the vaccine, though the proportion of such individuals varies widely from country to country (Ebrahimi et al. 2021). Other sources of distrust, such as misinformation and fake news on social and conventional media platforms, also contribute to vaccine skepticism and the even more harmful full-on antivaccination movement (Islam et al. 2021).
Hence, it is not surprising that in some countries, the topic of vaccination has become a veritable powder keg, with the public as well as the media discussing whether vaccination and measures used by officials intended to increase pressure on unvaccinated individuals are creating a societal rift (Burkhardt 2021; Donocik 2021). For these reasons, we believe the question concerning one’s vaccination status is highly sensitive in times of a pandemic.

Sensitive survey questions, in turn, are an age-old problem in survey methodology. Tourangeau and Yan (2007) defined sensitive topics in terms of (1) intrusiveness, (2) threat of disclosure, and (3) social desirability. Intrusive topics are those that concern taboo issues. They can be thought of as those kinds of questions that are “inappropriate in everyday conversation or out of bounds for the government to ask” (Tourangeau and Yan 2007:860). It is plausible that the issue of one’s COVID-19 vaccination status has become so hotly debated that individuals may find questions concerning their vaccination status to be intrusive.

Threat of disclosure refers to questions that are sensitive because truthful answers could lead to negative consequences. Some obvious examples are drug use or infidelity, but one’s COVID-19 vaccination status may also fall into this category. That is, although there does not currently exist a COVID-19 vaccination mandate in Germany, individuals, especially in certain jobs (physicians, elderly and daycare workers, etc.), may fear consequences if their (negative) vaccination status were to be revealed.

Finally, if respondents perceive getting the vaccination as the socially “desirable” thing to do, whether or not they themselves agree with this social expectation, they may feel pressure to state that they are vaccinated when they are in fact not. Normally, it is assumed that individuals deviate from the truth to avoid disapproval from interviewers or any third parties present during the interview situation. In online surveys, respondents generally do not have to fear direct disapproval because there is no interviewer present. In such cases, cognitive dissonance (Festinger 1957), i.e., the feeling of discomfort experienced when one’s actions do not align with what one feels is expected, may also cause respondents to answer in a socially desirable way (Wolter 2012). In other words, it may cause discomfort for unvaccinated individuals to recognize the social pressure to get vaccinated.

Any one of these three explanations may cause some individuals in a survey to misreport their vaccination status. The fact that the topic of COVID-19 vaccination is highly sensitive, and because there exists a widely accepted social norm that the vaccination protects oneself and others, means the misreporting will be in the form of “overreporting,” likely biasing survey estimates of COVID-19 vaccination coverage upward.

Survey methodologists have come up with several propositions intended to counteract the problem of misreporting to sensitive survey questions. Among them, ICT (Droitcour et al. 1991; Glynn 2013) has recently gained attention and seems to outperform other techniques such as the randomized response technique (Fox and Tracy 1986) or the crosswise model (Yu, Tian, and Tang 2008). Recent meta-analyses have shown that ICT reduces misreporting compared with conventional direct questioning (DQ; Ehler, Wolter and Junkermann 2021; Li and van den Noortgate 2019). The idea behind ICT is to anonymize respondents’ answers to survey questions so that respondents are better protected and social desirability stimulations are reduced. This is achieved by an experimental setup in which respondents answer several items at once by the use of item lists and count their affirmative answers to the whole list. This conceals the individual answer to the sensitive item of interest. Details about the ICT procedure are presented in the next section.

Method

The data for this study stem from an online access panel survey among a quota sample of the German general population between the ages of 18 and 69 years. The field period lasted from September 10 to September 20, 2021. In total, 7,530 individuals were surveyed. The survey used quotas to ensure its representativeness in terms of sex, age, and federal state residence makeup.3 With the help of a randomized mechanism, individuals were chosen to take part in the vaccination experiment to be described later. 2641 individuals took part in this portion of the experimental survey.

To examine whether survey-based estimates of COVID-19 vaccination coverage are biased, we compare an estimate based on the conventional method of simply asking directly about the individual’s vaccination status (direct question, DQ) with an estimate obtained using ICT. ICT works by randomly assigning respondents into two groups. Each group is presented with several statements and asked to simply respond with the number of statements they agree with, or that apply to them, but not which statements. The first group, which is called the “short list” (SL) group, is presented with k harmless or “nonsensitive” statements. The second group is called the “long list” (LL) group, and they are presented with the same k nonsensitive items plus the sensitive item of interest. In our case, the sensitive item is whether or not an individual is vaccinated against the coronavirus. A simple difference in the average number of “yes” answers between the SL and LL

3It is important to note that our estimated vaccination coverage should not be compared with estimates from the official channels, as our sample differs in important ways, especially in terms of age: by not including respondents older than 69 years, our estimates may be systematically lower than the true prevalence (as older individuals are the most likely to be vaccinated). However, we can make comparisons within the study, specifically between the experimental groups to assess the bias due to social desirability. This means the results will be generalizable in terms of any overreporting in survey-based studies, but the absolute levels in percentage of vaccinated individuals should not be generalized.
Socius: Sociological Research for a Dynamic World

groups provides an estimate of the prevalence of the sensitive item (equation 1). The standard error takes into account the estimated sampling variances of both SL and LL groups (equation 2; Wolter and Laier 2014). Formally, the estimate, \( \hat{\pi}_{ICT} \), and the variance, Var(\( \hat{\pi}_{ICT} \)), are given by

\[
\hat{\pi}_{ICT} = \bar{X}_{LL} - \bar{X}_{SL}
\]

and

\[
\text{Var}(\hat{\pi}_{ICT}) = \text{Var}(\bar{X}_{LL}) + \text{Var}(\bar{X}_{SL})
\]

where \( \bar{X}_{LL} \) and \( \bar{X}_{SL} \) are the sample means of the LL group and SL group, respectively, and \( \text{Var}(\bar{X}_{LL}) \) and \( \text{Var}(\bar{X}_{SL}) \) are their sampling variances (squared standard errors) estimated using conventional formulae.

The key to ICT is that at an individual level, there is no way to tell whether an individual is vaccinated or not.\(^4\) But by comparing the means of the two groups, we get an estimate of the vaccination coverage at the aggregate level, because by randomizing the group assignment, the average number of “yes” answers for the \( k \) harmless questions is assumed to be equal. Hence, the difference in averages must therefore be the result of the sensitive item.

Table 1 presents the experimental design, the items used in the SL and LL group of the ICT procedure, and sample sizes of our study. In the DQ format, respondents were asked the question on COVID-19 vaccination directly. Allocation to the DQ group, the ICT SL group, and the ICT LL group was random.\(^5\) If our hypothesis is correct that unvaccinated individuals will tend to misreport their status because of social desirability pressures, then the estimated prevalence in the DQ group should be higher than in the ICT group, for which the method affords complete anonymity to report the truth.

### Results

Estimates of the ICT SL and LL groups are reported in Table 2.\(^6\) Applying equations 1 and 2 yields the estimates shown in Table 3, along with the estimated vaccination rate for the DQ group. According to the latter, about 85 percent of the respondents have received at least one vaccination dose against the coronavirus. The ICT estimate is only 75 percent. The difference of about 10 percentage points is statistically distinguishable from zero at \( \alpha = .03 \) (\( z = 2.13 \)). We can state therefore that COVID-19 vaccination coverage on the basis of survey data using conventional DQ is susceptible to overreporting bias, and estimates based on this method are likely biased upward. There is much reason to believe that the

\( ^4 \)Obviously, this holds only if not all items of the LL are answered in the affirmative or in the negative (floor and ceiling effects). To avoid this, the nonkey items are designed in a way that they contain both low- and high-prevalence items, making it unlikely that floor and ceiling effects occur.

\( ^5 \)The survey contained further experimental groups devoted to other research problems. These cases are not included in the analysis in this article.

\( ^6 \)R and Stata code and data to replicate the analysis can be found at https://osf.io/rjf5e/.

---

**Table 1. Experimental Design.**

<table>
<thead>
<tr>
<th>Group</th>
<th>Items</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>DQ</td>
<td>“I have received at least one dose of vaccination against COVID-19.” (yes/no)</td>
<td>526</td>
</tr>
<tr>
<td>ICT short list</td>
<td>“I was treated by a physician last week.”</td>
<td>1,085</td>
</tr>
<tr>
<td>ICT long list</td>
<td>“I have never been diagnosed with cancer.”</td>
<td>1,030</td>
</tr>
<tr>
<td>ICT long list</td>
<td>“Private health insurance should be abolished.”</td>
<td></td>
</tr>
<tr>
<td>ICT long list</td>
<td>“Alternative medicine like homeopathy is helpful.”</td>
<td></td>
</tr>
<tr>
<td>ICT long list</td>
<td>“I have received at least one dose of vaccination against COVID-19.”</td>
<td></td>
</tr>
</tbody>
</table>

Note: DQ = direct questioning; ICT = item count technique. Translated from German.

**Table 2. ICT Estimates.**

<table>
<thead>
<tr>
<th>Group</th>
<th>Mean</th>
<th>Standard Error</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICT short list</td>
<td>1.988</td>
<td>.029</td>
<td>1,085</td>
</tr>
<tr>
<td>ICT long list</td>
<td>2.738</td>
<td>.033</td>
<td>1,030</td>
</tr>
</tbody>
</table>

Note: ICT = item count technique.
aforementioned previous surveys (some of them conducted by the RKI) also suffer from this bias.7

We also carried out a subgroup analysis of the DQ-ICT difference by region (western vs. eastern Germany), gender, and education. Because of the limited statistical power of the ICT estimates, we present only some basic bivariate comparisons, shown in Table 4. The results are striking. Whereas for western Germans, men, and respondents without a university education, the DQ-ICT differences are lower than 10 percentage points and below conventional levels of statistical significance, we see large differences for the other groups, namely among eastern Germans, women, and highly educated respondents. This is plausible in the light of a social desirability interpretation: those respondents who tend to have a lower vaccination rate and/or those who feel a high pressure from social desirability norms (probably the highly educated in particular) are more prone to misreport their own vaccination status. Moreover, the findings of the subgroup comparisons can be taken as a further robustness check of the main message of our article, as we find upwardly inflated estimates of COVID-19 vaccination in all subgroups if DQ is applied.

Table 3. Estimated Vaccination Coverage.

<table>
<thead>
<tr>
<th>Question Type</th>
<th>Estimated Vaccination Coverage</th>
<th>Standard Error</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct questioning</td>
<td>84.8%</td>
<td>1.6</td>
<td>526</td>
</tr>
<tr>
<td>Item count technique</td>
<td>75.0%</td>
<td>4.3</td>
<td>2,115</td>
</tr>
</tbody>
</table>

Table 4. Subgroup Analysis: Reported Coronavirus Disease 2019 Vaccination by Question Format.

<table>
<thead>
<tr>
<th></th>
<th>DQ</th>
<th>ICT</th>
<th>Difference</th>
<th>p (Difference)</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Western Germany</td>
<td>86.3%</td>
<td>78.4%</td>
<td>7.9</td>
<td>.106</td>
<td>2,256</td>
</tr>
<tr>
<td>Eastern Germany</td>
<td>75.3%</td>
<td>55.0%</td>
<td>20.4</td>
<td>.112</td>
<td>385</td>
</tr>
<tr>
<td>Male</td>
<td>88.9%</td>
<td>81.6%</td>
<td>7.3</td>
<td>.258</td>
<td>1,341</td>
</tr>
<tr>
<td>Female</td>
<td>81.0%</td>
<td>67.9%</td>
<td>13.1</td>
<td>.044</td>
<td>1,300</td>
</tr>
<tr>
<td>Lower education</td>
<td>83.6%</td>
<td>79.8%</td>
<td>3.8</td>
<td>.479</td>
<td>1,954</td>
</tr>
<tr>
<td>Higher education</td>
<td>88.5%</td>
<td>61.7%</td>
<td>26.8</td>
<td>.002</td>
<td>687</td>
</tr>
</tbody>
</table>

Note: Western Germany includes the city of Berlin. Higher education is made up of a university or university of applied science diploma. Values in parentheses are standard errors. DQ = direct questioning; ICT = item count technique.

Discussion

Different methods tend to produce different results. The fact that COVID-19 vaccination coverage estimates differ depending on the method of data collection should come as no surprise to those in the scientific community. But such discrepancies are troublesome not only because they make it more difficult to develop prognoses and plan policy but also because they can undercut trust in governments and institutions, which is already at a premium in many regions in the ongoing pandemic.8

Our analysis of the use of survey data in estimating vaccine coverage underlines those difficulties: although surveys may be useful in the context of many other types of vaccines, we argue that the topic of COVID-19 vaccination in late 2021 is too sensitive to rely solely on survey data for coverage rates. Although it cannot be ruled out that official

7For robustness analysis, we also performed a test for design effects as proposed by Blair and Imai (2012), with the result that there are no possibly biasing design effects. We further applied alternative maximum likelihood and nonlinear least squares ICT estimators as proposed by Blair and Imai as well. These estimators did not perform better (with respect to statistical efficiency and convergence) than the difference-in-means estimator presented above.

8In Germany, the RKI was heavily criticized for correcting its vaccination coverage estimates. The newspaper Bild printed the headline “Again and Again False Numbers from the RKI: The Corona-Chaos-Agency” (Bild 2021). The article continues, “The [RKI] is THE central institution of the Federal Government to monitor diseases. Since the beginning of the pandemic, the government has followed THEIR recommendations concerning Corona policies and depends on THEIR figures.” The Twitter comments of this headline, for example, are filled with accusations that the government and its institutions are corrupt and deceitful, demonstrating how such discrepancies can fuel distrust in institutions and conspiracy theories in general (https://twitter.com/BILD/status/1446511665256833048, last accessed November 17, 2021).
statistics based on reporting by hospitals and physicians are not also biased (perhaps because of incomplete or erroneous reporting), we showed in this investigation that COVID-19 vaccination coverage on the basis of survey data is likely biased upward by social desirability. Providing individuals with an anonymous way to report their unvaccinated status resulted in an estimated vaccination coverage that was significantly lower than the one based on the conventional method of DQ.

However, there are some important limitations to note. First, this article is written within the context of Germany in the fall of 2021. The local situation may change in the future, and it may be that the topic of COVID-19 vaccination is not as sensitive in other parts of the world. In some countries, vaccination coverage is extremely high. For example, in Portugal and Singapore, nearly 90 percent of adults are fully vaccinated as of November 2021. This means that in those countries, the question of vaccine status is likely much less sensitive. After all, as Tourangeau and Yan (2007:860) noted, “a question about voting is not sensitive for a respondent who voted,” so a question about one’s vaccination status is likely not sensitive for someone who is vaccinated.

Another limitation of the study is that we cannot compare our survey estimates with the survey estimates of the RKI or any other institute, because our survey was a nonprobability sample of a particular group of Internet users. However, the randomized experimental design means that we can indeed compare estimates between groups (DQ and ICT) within the study. This also entails that although we have an unbiased estimate of the sample treatment effect (with high internal validity resulting from a strict experimental setup), we have no guarantee that the population treatment effect is the same. This relates both to the German population and to other countries. As we noted earlier, question sensitivity probably varies across different populations, and hence the treatment effect of applying ICT to survey questions on COVID-19 vaccination is likely to also vary. Moreover, we cannot make any statements about the validity of estimates on the basis of data collected directly from health officials and workers, such as those collected in the RKI’s Digital Vaccine Coverage Monitor. It is possible that incomplete data transferred from vaccination centers, hospitals, and physicians may lead to other forms of bias (especially if the quality of the reporting is dependent on other unobserved factors).

Another issue is statistical power. If we follow the advice of Blair, Coppock, and Moor (2020), then our sample size implies a power of less than 80 percent for a given DQ-ICT difference of 10 percentage points. This is a notorious problem of ICT procedures, which always come with highly inflated standard errors compared with conventional estimates. With respect to future studies, we strongly recommend ensuring sufficiently large sample sizes on one hand and using more advanced ICT setups that help boost the statistical efficiency of the estimates on the other hand (for some propositions, see Aronow et al. 2015).

More work is needed to discover further sources of bias (e.g., reaching nonnative speakers in surveys) to get a better idea of the true vaccination coverage. Until then, discrepancies in statistics will continue to exist for well-known reasons such as sampling error, survey biases, systematic under- and over-reporting by health organizations, and others. The ultimate goal should be to work toward understanding as many sources of bias and inaccuracy as possible in order to provide the general public with honest and transparent information and avoid confusion and the potential for misrepresentation of statistics.

Data

Data and replication code for the analysis can be found at https://osf.io/rj5se/.

Funding

This study is part of the project Sensitive Questions and Social Desirability: Theory and Methods, funded by the German Research Foundation (grant WO 2242/1-1 to F.W. and MA 3267/7-1 to J.M.). The publication of this article was funded by Chemnitz University of Technology and by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 491193532.

ORCID iD

Jochen Mayerl https://orcid.org/0000-0002-4599-9976

References


Author Biographies

Felix Wolter is a postdoctoral researcher and lecturer at the Cluster of Excellence “The Politics of Inequality” and the Sociology Department at the University of Konstanz. His research interests include survey methodology with a specialization in sensitive survey questions and factorial survey experiments, social inequality and social stratification, and rational choice and game theory.
Recent publications have appeared in *Public Opinion Quarterly* and *Research & Politics*.

**Jochen Mayerl** is a professor of sociology with a focus on empirical social research in the Faculty of Behavioural and Social Sciences at Chemnitz University of Technology. He specializes in the study of attitude-behavior relations, social context analysis, survey methodology, and structural equation modeling. His research has appeared in *PLoS ONE*, *European Societies* and *Environment & Behavior*.

**Henrik K. Andersen** studied social sciences (BA) and empirical political and social research (MA) at the University of Stuttgart. He is currently a research assistant in the Institute of Sociology at the Chemnitz University of Technology. Before that, he was a research assistant in the Department of Social Sciences at Technische Universität Kaiserslautern. His substantive interests are in social desirability bias and more recently the examination and explanation of intraindividual changes in environmental behavior. His methodological focus is on the application of fixed-effects panel regression in the structural equation modeling framework.

**Theresa Wieland** studied sociology and worked as a research assistant at the University of Mainz. She is now an independent doctoral fellow at the Cluster of Excellence “The Politics of Inequality” at the University of Konstanz. Her research interests are environmental and social inequalities, influences on public attitudes, and international research.

**Justus Junkermann** studied sociology at the University of Leipzig (BA and MA). He is currently a research assistant at the Institute for Medical Sociology at Martin Luther University of Halle-Wittenberg. Before that, he was a research assistant at the Institute for Sociology at Johannes Gutenberg University Mainz. His interests are agent-based modeling, social desirability research, survey methodology, and medical sociology.