THE EFFECT OF QUESTION WORDING ON ATTITUDES TOWARD PRENATAL TESTING AND ABORTION

ELEANOR SINGER*
MICK P. COUPER

Abstract   At intervals since 1990, the General Social Survey (GSS) has asked a series of four questions inquiring into knowledge of genetic testing and attitudes toward prenatal testing and abortion, most recently in 2010. The questions about prenatal testing and abortion were framed in terms of "baby". But in the current anti-abortion climate, it seemed possible that the word "fetus" would carry more abstract, impersonal connotations than "baby" and might therefore lead to different responses, especially in the case of abortion. To resolve this issue, we designed the question-wording experiment reported in this research note. We found no significant differences by question wording for abortion preferences in the sample as a whole and small but significant differences for prenatal testing, in a direction opposite to that expected. However, question wording did make substantial differences in the responses of some demographic subgroups.

Introduction

At intervals since 1990, the General Social Survey (GSS; see http://www3. norc.org/GSS+Website/Documentation/) has asked a series of four questions inquiring into the general public's knowledge about genetic testing and attitudes toward prenatal testing and abortion, with the last administration in 2010.

From 1990 through 2010, the questions about prenatal testing and abortion were framed in terms of "baby." But after the 2010 results were released, some members of an NIH review panel questioned whether the answers might have been different had they been framed in terms of "fetus" rather than "baby." The word "baby" had been chosen originally on the assumption that "fetus" would be less familiar to respondents, and would therefore lead to more Don't Knows and No Answers. But, given the long-standing, and intensifying, national debate about abortion in the United States,[1] it seemed possible that the word "fetus," as a more abstract, impersonal term, might lead to more frequent expressions of preferences for prenatal testing and, especially, abortion.

Research by Simon and Jerit (2007) shows that the two terms—"baby" and "fetus"—are not neutral but carry ideological connotations, with pro-life groups more likely to use the term "baby" and pro-abortion groups the word "fetus." Their content analysis of briefs filed by opposing sides in *Stenberg v. Carhart* (2000), a US Supreme Court case involving the so-called partial-birth abortion (PBA) ban, found that "baby" made up more than 75 percent of total uses of either term by supporters of the ban, whereas "fetus" was preferred by roughly 60 percent in briefs opposing the ban. Simon and Jerit (2007) also carried out an experiment involving news reports identical except for deliberately varied usage of the two terms, and found that approval of the PBA ban was significantly reduced when subjects read an item containing consistent references to "fetus" rather than "baby."

The question-wording experiment described below tests whether framing questions about prenatal testing and abortion using these different terms leads to different responses and, potentially, creates different views of trends in public opinion on this issue. More broadly, the research has implications for how question wording can shape expressed opinion and, ultimately, policy.

Methods

The experiment was conducted as part of the Time-Sharing Experiments for the Social Sciences, or TESS (see www.tessexperiments.org; Mutz 2011). TESS, which is funded by the National Science Foundation, conducts general-population experiments on behalf of investigators throughout the social sciences. For data collection, TESS contracts with GfK Knowledge Networks (see www.knowledgenetworks.com), which conducts surveys using its KnowledgePanel®. The panel is a probability-based web panel based on an address-based sampling (ABS) frame, with recruitment by mail and telephone (see DiSogra, Callegaro, and Hendarwan 2009). Panel members without

---

1. See, for example, the following stories in the *New York Times:* Erik Eckholm, "Access to Abortion Falling as States Pass Restrictions," January 3, 2014; Jeremy W. Peters, "Parties Seize on Abortion Issues in Midterm Race," January 20, 2014; and Erik Eckholm, "Abortions Declining in US, Study Finds," February 3, 2014.

Internet access are provided with a laptop computer and free Internet access. The panel is designed to be representative of the US adult population. Surveys are conducted in English only.

SAMPLE AND RESPONSE RATE

The stand-alone survey on which the genetics questions were asked was administered by GfK Knowledge Networks from March 22 to April 2, 2012. Some 2,496 panelists 18 or older were randomly drawn from the KN panel; 1,570 responded to the invitation, yielding a completion rate (COMR; see AAPOR 2011, 37) for this final stage of 62.9 percent. The initial recruitment rate (RECR) for this study, reported by GfK Knowledge Networks, was 14.9 percent, and the profile rate (PROR) was 65.6 percent, for a cumulative response rate (CUMRR; see AAPOR 2011, 37) of 6.1 percent (14.9 x 65.6 x 62.9 percent; Callegaro and DiSogra 2008).

How generalizable are the results obtained from this panel? Research by GfK Knowledge Networks (Dennis 2010) comparing this kind of sample with telephone RDD samples suggests that they are equally representative (see also Yeager et al. 2011). Smith (2003), conducting an explicit comparison of GSS and KN data, found that while the two agreed in most comparisons, there were some systematic differences. In particular, because of the way response options are presented, levels of Don't Know (DK) responses tend to be consistently higher on the GSS, and questions about potentially sensitive behaviors tend to elicit higher levels of agreement on the KN than on the GSS. The latter finding, also reported by Dennis, Li, and Chatt (2004) and Dennis and Li (2007), is attributed by them to social-desirability effects of the presence of an interviewer on the GSS. Our analyses (not shown; table available from authors) indicate that with the two exceptions noted above, responses to the genetic testing questions on the 2010 General Social Survey and on the 2012 KN survey carried out for TESS are very similar, suggesting that the findings of this experiment should apply to the GSS and other surveys.[2]

QUESTIONS

The four questions about genetic testing asked on the TESS questionnaire in 2012 (as well as on the GSS, most recently in 2010; see Smith et al. [2011]) are as follows:

GENETEST: How much would you say you have heard or read about genetic testing—a great deal, something but not very much, or nothing at all?

GENEGOOD: Some people say that genetic testing is a wonderful advance. Others think it may cause trouble. Based on what you know, do

2. It is, of course, possible that offsetting forces in the two studies may yield similar distributions.

you think genetic testing will do more good than harm or more harm than good? (More good than harm/More harm than good)

GENESELF: Today, tests are being developed that make it possible to detect serious genetic defects (before a baby is born/in the fetus during pregnancy). But so far, it is impossible either to treat or to correct most of them. If (you/your partner) were pregnant, would you want (her) to have a test to find out if the (baby/fetus) has any serious genetic defects? (Yes/No)

GENEABRT: Suppose a test shows the (baby/fetus) has a serious genetic defect. Would you, yourself, want (your partner) to have an abortion if a test shows the (baby/fetus) has a serious genetic defect? (Yes/No)

A randomly selected half of the respondents were administered the four questions with the "baby" wording for the last two questions, while the balance were administered the four questions with the "fetus" wording. Examination of the distributions of demographic variables used as controls indicates that the randomization was successful—none differed significantly, and almost all differed by no more than a percentage point between the two halves. Following the last question, a random half of the respondents were asked, "We would very much appreciate your telling us why you chose this answer." The other half were asked, "Earlier, you said you (would/would not) want to have a test to find out if the (baby/fetus) has any serious genetic defects. We would very much appreciate your telling us why you chose this answer." They were probed once if they did not provide an answer.

The KN panel provides both selection and post-stratification weights. We examined both the weighted and unweighted distributions. The results are similar, except that the unweighted differences are somewhat smaller, and fewer of the differences were statistically significant. We present the weighted analyses here (unweighted analyses are available upon request).

Results

Distributions of responses to both versions of the question asking whether the respondent would want prenatal testing for a genetic defect are shown in table 1.

Responses are weighted and exclude DKs and refusals (less than 1 percent after a single probe). The difference resulting from question wording is significant at the 5 percent level (chi-square = 4.33, d.f. = 1, $p = 0.037$). However, the direction of the difference is opposite to that suggested by Simon and Jerit's (2007) research: fewer respondents indicate a preference for prenatal testing when the question asks about testing a fetus rather than a baby. This is not due to a greater percentage of DKs or refusals to the question asking about a fetus: the percentage of respondents giving nonsubstantive

Table 1. Preference for Prenatal Testing by Question Wording (weighted distributions)

Preference	Baby (%)	Fetus (%)	Total (%)
Yes	64.1	59.0	61.5
No	35.9	41.0	38.5
Total	100.0	100.0	100.0
(*n*)	(771)	(792)	(1,563)

Chi-square = 4.33, d.f. = 1, *p* = 0.037

Table 2. Preference for Abortion by Question Wording (weighted distributions)

Preference	Baby (%)	Fetus (%)	Total (%)
Yes	40.2	42.1	41.2
No	59.8	57.9	58.8
Total	100.0	100.0	100.0
(*n*)	(765)	(788)	(1,553)

Chi-square = 0.523, d.f. = 1, *p* = 0.47

answers is 0.52 to the question asking about "baby," versus 0.38 to the question asking about "fetus."

Responses to the question asking about preferences for abortion in case a prenatal test reveals a fetal defect are shown in table 2. Responses are again weighted and exclude DKs and refusals; as in the case of preferences for prenatal testing, percentages of nonsubstantive responses are very low: 1.28 for "baby" and 0.91 for "fetus." The difference between question wordings is not significant (chi-square = 0.523, d.f. = 1, *p* = 0.47).

We next examined the effect of question wording on preferences for prenatal testing and abortion, controlling for demographic characteristics (gender, race, age, education, religion, and ideology) that clearly predate opinions about these issues (the results of these analyses are available in online appendix table 1). We also examined the interactions of demographic characteristics and question wording to detect whether the effect of question wording differed across subgroups. All interactions were tested simultaneously in a single model for each question of interest (the results of these analyses are available in online appendix table 2).

Introducing controls for demographic characteristics does not eliminate the significant effect of question wording on genetic testing preferences, but the effect appears to vary by demographic characteristics. Specifically, the interactions with education (Wald chi-square = 9.67, d.f. = 2, *p* = 0.008) and religion (Wald chi-square = 17.4, d.f. = 2, *p* < .0001) are statistically significant.

College graduates are significantly less likely to say they would want prenatal testing if the question refers to a fetus rather than a baby, as are those describing their religion as "other."

We performed a similar analysis for abortion preferences in case of fetal defect (the results of these analyses are available in online appendix table 2). Unlike prenatal testing, question wording does not make a significant zero-order difference in preferences for abortion. However, most demographic characteristics are significantly related to abortion preferences, and a number of interactions between demographic characteristics and question wording also have significant effects: race (Wald chi-square = 3.87, d.f. = 1, p = 0.049), age (Wald chi-square = 12.81, d.f. = 3, p = .0051), and ideology (Wald chi-square = 8.98, d.f. = 2, p = .011). When "baby" is used, there is a significant difference by race, with nonwhites less likely to prefer abortion in case of fetal defect than whites; no such difference is found when "fetus" is used. On the other hand, "fetus" shows significant differences by age, with 18–25-year-olds especially unlikely to express a preference for abortion; no such effect is found for "baby." Finally, those who describe themselves as "conservative" seem unaffected by question wording, whereas "liberals" and "moderates" are significantly affected, but in opposite directions (more support for abortion by liberals when "baby" is used, less support by moderates when "baby" is used). In short, we see no consistent pattern across the interactions: sometimes "baby" has larger effects than "fetus," and sometimes the reverse is true; the effects are not consistent across GENESELF and GENEABRT.

For both prenatal testing and abortion, the overall effects of question wording are small, as demonstrated in tables 1 and 2. In the case of prenatal testing, the difference is five percentage points, statistically significant but in a direction opposite to that hypothesized; for abortion, the difference is two percentage points, in the direction hypothesized but not statistically significant. For both questions, examining the pseudo-R^2 statistics indicates that the effect of question wording is extremely small compared to that of the demographic characteristics, regardless of whether weighted or unweighted versions are used.

But question wording does make an appreciable difference for specific subgroups, as noted above. To illustrate the practical impact of statistically significant changes made by the wording of the question for a particular subgroup, table 3 shows the percentage of different educational groups who say they would opt for prenatal testing if the question were worded in terms of "baby" or "fetus."[3] Many more college graduates express a preference for prenatal testing when the word used is "baby" rather than "fetus," whereas question wording makes virtually no difference for those with a high school

---

3. Table 3 is based on the weighted three-way cross-tabulations. Similar results are obtained if we use the predicted proportions derived from the full model with interactions described above.

Table 3. Preference for Prenatal Testing by Education and Question Wording (weighted distributions)

Education	Baby		Fetus	
	%	(n)	%	(n)
High school or less	55.4	(329)	57.7	(344)
Some college	62.8	(225)	57.1	(231)
College graduate	78.5	(217)	62.8	(218)
Total	64.1	(771)	59.0	(792)
Chi-square tests for differences by education	$\chi^2 = 30.64$, d.f. = 2, $p < .0001$		$\chi^2 = 1.88$, d.f. = 2, $p = 0.39$	

education or less, and only a small difference for those with some college education.[4]

We reran the analyses above with several restrictions on the sample to see whether the results would change significantly. Because questions about testing and abortion might be more relevant to people of childbearing age, we restricted the analyses to those under 45. We also restricted the analysis of responses to the abortion question to those who indicated that they would want prenatal genetic testing.

The only restriction that made a significant difference in the results was limiting the analysis to those indicating a preference for prenatal testing. Such respondents were much more likely to express a preference for abortion than those who had indicated they did not want prenatal testing, and they also made up a sizable fraction (about two-thirds) of the total number of respondents. However, there is no indication that respondents were confused because they were asked about their preferences for abortion even though they had expressed no desire for prenatal testing. Most respondents (94 percent) gave what we considered to be "consistent" answers to this sequence of questions; that is, they said they wanted neither testing nor abortion, or both testing and abortion, or testing but no abortion. Only 6 percent gave what we considered to be inconsistent responses; that is, expressing no desire for prenatal testing but a preference for abortion if a prenatal test showed a defect

4. Some readers of an earlier version of this article—all college-educated women—suggested that the term "baby" conjures up a more vivid image than the word "fetus," making respondents more likely to want to protect babies than fetuses by having prenatal testing. However, that logic does not explain why only college-educated women would be so affected, nor does it explain why question wording does not make a difference with respect to preferences for abortion. We found no evidence for this explanation in respondents' answers to the open-ended question about why they had indicated a preference for prenatal testing. With one exception—high school graduates were much less likely than those with more education to say they wanted testing in order to have a choice when "baby" rather than "fetus" was used— reasons given were very similar for "fetus" and "baby" at every educational level.

in the fetus.[5] Consistency (by our definition) differed by question wording, with those asked about "baby" significantly more likely to give consistent responses than those asked about "fetus" (95.3 versus 92.4 percent [chi-square = 5.82, d.f. = 1, $p = 0.016$]).

As already noted, the word "fetus" had originally not been used on the GSS because of a belief that it would be more difficult to understand than "baby" and that this effect would be more pronounced among older and less well-educated respondents. In order to test this hypothesis, we examined item nonresponse to the two question wordings (fetus versus baby) on the 2012 TESS survey. There was no consistent pattern of item nonresponse by question wording for either weighted or unweighted responses, even with controls added for education and age.

Following Draisma and Dijkstra's (2004) suggestion that longer response times may signal difficulties with a question, we also examined response times for the two versions of GENESELF and GENEABRT. For GENESELF, the mean times were 20.54 seconds for baby and 20.65 seconds for fetus. For GENEABRT, the mean times were 14.23 seconds for baby and 14.82 for fetus. Neither of these differences is significant. Thus, we found no support for the hypothesis that one word is more or less easily comprehended, at least at the present time.

We also attempted to assess the validity of the two terms by comparing the pseudo *R*-squares for models predicting preferences for prenatal testing and abortion from alternative question wordings plus demographic characteristics. For both questions, pseudo *R*-squares were slightly higher when "baby" was used (0.076 versus 0.066 for prenatal testing and 0.097 versus 0.066 for abortion). We also examined the correlation between answers to GENEGOOD and GENESELF, and between GENESELF and GENEABRT, by question wording. The correlations were slightly higher between the second pair of questions than the first (around 0.4 versus 0.3), but question wording made virtually no difference in either set of correlations. Thus, there is no evidence from any of these analyses that "fetus" should be substituted for "baby" in future administrations of these questions.

Discussion and Conclusions

This question-wording experiment was designed to test whether using the term "fetus" rather than "baby" would alter public preferences about prenatal testing for genetic defects and for abortion if a test revealed such defects. We found a statistically significant but small difference in responses to the question about

---

5. Answers to an open-ended question probing the reasons for these responses indicated that many of those with a preference for abortion but not testing said they did not want to know about possible prenatal defects because they would not have an abortion under any circumstances but would be seriously conflicted if they knew ahead of time that their child would be born with a severe genetic defect.

genetic testing, with those asked about a "fetus" five percentage points *less* likely to say they would want such a test. No significant differences were found in responses to the question about abortion. As suggested in footnote 3, it seems possible that "fetus," as a more abstract term, may actually have reduced preferences for testing, while leaving expressed preferences for abortion unaffected. Although question wording made relatively small differences in the sample as a whole, those differences became substantial in some subgroups. Furthermore, the effect of question wording differed for the two questions.

In an effort to understand whether differences in comprehension produced these effects, we looked at item nonresponse (admittedly a weak proxy for comprehension) to the questions about testing and abortion for both question wordings. Item nonresponse to these questions was very low and did not differ by question wording, even in subgroups defined by education and age, nor did response times differ between question wordings. Several attempts to assess the relative validity of these terms likewise failed to show one as superior to the other. The large majority (94 percent) of respondents gave consistent responses to the questions about testing and abortion, though consistency differed somewhat by question wording.

Based on the results of the present study, it seems impossible to say that one of these terms is better than the other. However, respondents seem to use the term "baby" more often spontaneously, even when the question is framed in terms of "fetus": 14.9 percent of open-ended responses included the term "baby," while only 3.2 percent of open-ended responses used the term "fetus." Use of "fetus" did not, as might have been expected, significantly increase preferences for abortion. For comparability in an analysis of trends, therefore, the word "baby" should be retained in future administrations of these questions. But, given the difference that question wording makes for the responses of specific subgroups, analysts (and producers) of poll data should pay particular attention to the wording used in poll questions, especially those trying to shape the debate or influence attitudes.

Supplementary Data

Supplementary data are freely available online at http://poq.oxfordjournals.org.

References

American Association for Public Opinion Research (AAPOR). 2011. *Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys.* 7th ed. Deerfield, IL: AAPOR.

Callegaro, Mario, and Charles DiSogra. 2008. "Computing Response Metrics for Online Panels." *Public Opinion Quarterly* 72:1008–32.

Dennis, J. Michael. 2010. "KnowledgePanel®: Processes & Procedures Contributing to Sample Representativeness & Tests for Self-Selection Bias." GfK Knowledge Networks. Available at http://www.knowledgenetworks.com/ganp/reviewer-info.html.

Dennis, J. Michael, and Rick Li. 2007. "Results of a Within-Panel Web Survey Experiment of Mode Effects." Paper presented at the American Association for Public Opinion Research, Anaheim, CA, USA.

Dennis, J. Michael, Rick Li, and Cindy Chatt. 2004. "Benchmarking Knowledge Networks' Web-Enabled Panel Survey of Selected GSS Questions against In-Person Interview." Knowledge Networks Report, February.

DiSogra, Charles, Mario Callegaro, and Erlina Hendarwan. 2009. "Recruiting Probability-Based Web Panel Members Using an Address-Based Sample Frame: Results from a Pilot Study Conducted by Knowledge Networks." Proceedings of the Joint Statistical Meetings, Survey Research Method Section, 5270–83.

Draisma, Stasja, and Wil Dijkstra. 2004. "Response Latency and (Para) Linguistic Expressions as Indicators of Response Errors." In *Methods for Testing and Evaluating Survey Questionnaires*, edited by Stanley Presser, Jennifer M. Rothgeb, Mick P. Couper, Judith T. Lessler, Elizabeth Martin, Jean Martin, and Eleanor Singer, 131–48. Hoboken, NJ: Wiley.

Mutz, Diana C. 2011. *Population-Based Survey Experiments*. Princeton, NJ: Princeton University Press.

Simon, Adam F., and Jennifer Jerit. 2007. "Toward a Theory Relating Political Discourse, Media, and Public Opinion." *Journal of Communication* 57:254–71.

Smith, Tom W. 2003. "An Experimental Comparison of Knowledge Networks and the GSS." *International Journal of Public Opinion Research* 15:167–79.

Smith, Tom W., Peter Marsden, Michael Hout, and Jibum Kim. 2011. *General Social Surveys, 1972–2010: Cumulative Codebook*. Chicago: National Opinion Research Center.

Yeager, David S., Jon A. Krosnick, LinChiat Chang, Harold S. Javitz, Matthew S. Levendusky, Alberto Simpser, and Rui Wang. 2011. "Comparing the Accuracy of RDD Telephone Surveys and Internet Surveys Conducted with Probability and Non-Probability Samples." *Public Opinion Quarterly* 75:709–47.