What difference makes a difference? – A meta-regression approach on the effectiveness conditions of incentives in self-administered surveys

Katrin Auspurg & Andreas Schneck (Goethe-University Frankfurt am Main)

Keywords:

Nonresponse, response rates, incentives, rational choice, exchange theory, meta-analysis

Abstract:

Nonresponse bias threatens the validity of inferences from sample results on the population. Increasing the response rate is in most surveys the only possible tool for reducing the risk of nonresponse bias. Especially incentives are one of the most recommended tools to increase response rates. In this study over 80 years of experimental evidence on the effectiveness of incentives in increasing survey participation is analyzed combining 365 trials from 205 studies in a meta-regression approach that allows to control for effect modifiers like different incentive values as well as type and timing of the incentive. In line with rational-choice and exchange theoretical assumptions incentives are proved to be an effective tool to increase survey response, but there is also a high variance between different incentive treatments. Unconditional monetary incentives are found to be the most effective setting which is able to increase the response rates up to 20 percentage points. Implications for survey research are discussed.

Acknowledgement: The authors want to thank the participants of the "Rational Choice Sociology: Theory and Empirical Applications" Workshop at Venice International University, December 2013 as well as Phil Edwards for providing further insights in his analyses. The research was funded by a Young-Scholar-Fund grant of the University of Konstanz.

Introduction

Survey research is supposed to draw a valid picture of various social phenomena. Due to budget, time restriction or missing access to the whole target population often only a small fraction of the target population can be surveyed. Nonetheless it's the goal of researchers to infer from sample parameters on the population. Valid inferences require that the responding sample differs solely by random from the target population: sample persons need to have the same (or at least a known) probability to respond to a survey request. If there are any systematic differences between responding and target population there exists non-response bias that threatens valid inference (cf. Groves 2009: 59).[1] For most surveys, especially if there is no or sparse information on the sampling-population, increasing the response rate is the only possibility to keep the risk of nonresponse bias at minimum.[2] Dropping response rates in the last decades increased the threat of nonresponse bias (Groves 2011; Aust and Schröder 2009; De Leeuw and De Heer 2002). To counter the problem of dropping response rates and nonresponse bias more effort is put into strategies increasing response rates (e.g. incentives, reminders, colored paper, etc.). According to Dillman (2007, 1978) and the AAPOR [American Association for Public Opinion Research] (2010), reminder and incentives are the most effective tools therein.

An extensive literature, including meta-analyses, broach the issue of incentives increasing response rates based on experimental studies. In previous meta-analyses incentives in general proved effective to increase response rates (Kanuk and Berenson 1975; Linsky 1975; Yu and Cooper 1983; Armstrong 1975; Church 1993; Edwards et al. 2002; Edwards et al. 2009; Boyle et al. 2012).[3] The meta-analysis of Edwards et al. (2005) summarized survey experiments testing

---

[1] See also the Public Opinion Quarterly special issue on "Nonresponse Bias in Household Surveys" 2006 (5).
[2] Although response rates are far from a perfect measure of nonresponse bias (Groves and Peytcheva 2008)
[3] Incentives also proved to be effective in other modes like telephone (Singer, Van Hoewyk, and Maher 2000) and cell phone surveys (Singer and Ye 2013).

monetary against nonmonetary incentives and provided first insights in the higher effectiveness of monetary incentives. (Edwards et al. 2009: 335). Increasing the incentive amounts was only found to be effective until an amount of about 1 US$ was reached (Edwards et al. 2005: 996).[4] In addition unconditional incentives that are provided independently of survey participation already with the survey request seem to be more effective than conditional incentives that are promised in the cover letter for the case of survey completion (Edwards et al. 2009: 347f.). However, both meta-analyses suffer from a low number of included trials (13 resp. 24) as well as excess heterogeneity of experimental treatments (e.g. caused by different values of incentives). All in all, there is a lack of research on the effects of different incentive treatments and in particular their interdependencies: Are for instance, conditional incentives the better the higher their value? Does a higher incentive value increases the effectivity also for non-monetary incentives?

Summarizing 365 trials from more than 200 studies we make three new contributions to the literature: First, we examine the effectiveness conditions of incentives in cross-sectional self-administered surveys (web and paper) including more different incentive- and study-settings than prior research. Second, we employ a meta-regression approach that allows to account for the substantial heterogeneity across study settings while at the same time controlling for possible publication bias that could inflate effect sizes. Third, using general theories on survey participation, like decision-orientated frameworks and theories on the exchange framework between researchers and respondents, the most effective and valid survey frames are specified not only empirically, but also analytically (cf. Schaeffer and Dykema 2011: 923), which allows more general conclusions for survey research.

[4] The term „incentive amount“ is only used for monetary incentives; in case of nonmonetary incentives the more general term „value“ is used instead.

Theoretical outline

From the perspective of rational choice theory respondents choose actions, like participating in a survey, if the benefits of this action (e.g. an interesting survey topic) exceeds its costs (esp. opportunity costs, the time that is needed for the survey completion Dillman 1978: 14). The higher the benefits, the more likely the costs of the action (survey participation) are outrun for a large number of sample persons. Hence the first two assumptions are:

H_{1a}: Survey participation is higher if there are any incentives; H_{1b}: The higher the incentive-value awarded, the higher the odds of response.

In regard to the type of incentive (monetary or not) the character of money as a "general reinforcer" (Skinner 1953: 79) is crucial. Money allows respondents to fulfill their own needs instead of being limited to the gift (e.g. a ball-pen or a pan-scraper; Little and Engelbrecht 1990)) itself. With nonmonetary incentives there should always be some risk that the sample person is not (or no more) in need of it.

H_2: Monetary incentives are more effective than nonmonetary incentives.

Following these arguments one can also expect a positive interaction of monetary incentives with the incentive value. If respondents are not in need of the nonmonetary incentive (for instance, as they already have enough ball-pens or pan-scrapers), the incentive has no subjective value for the respondent, irrespective of its objective value; in other words, it is more the concrete incentive than its monetary value that determines the utility for the respondent.[5] Therefore, following rational choice theories in particular the utility of monetary incentives should increase with the amount awarded.

H_3: There is a positive interaction between monetary incentives and the incentive amount.

[5] The respondent could only profit if he or she sells the incentive or is able to trade it, which means however transaction costs that have to outrun the utility, which is – given typical values of incentives – usually not the case.

From a narrow rational choice perspective in addition conditional incentives should be more effective than unconditional ones. This is the case because respondents can minimize costs (save time) and maximize benefits by taking the incentive without survey participation; hence one should expect higher response rates with conditional than unconditional incentives. However prior studies already found this hypothesis not supported by empirical data.

Using an extended rational choice version including social norms one can indeed assume that unconditional incentives work better than conditional ones (Groves, Cialdini, and Couper 1992: 480). The explanation is the norm of reciprocity (Gouldner 1960; Mauss 1967) which relies on normative expectations to reciprocate gifts. Exchange theories (Blau 2008; Homans 1958) allow to include those norms in a rational choice framework. Blau (2008), for instance, assumes a bargaining relation between actors (like researchers and survey participants) which is stabilized but not completely driven by norms of reciprocity (Blau 2008: 92). The compliance to social norms can be understood as an additional utility term as following norms allows to avoid social sanctions as well as moral costs (Slote 1985: 165). Moral costs occur if the sample person internalized the norm of reciprocity but declines to participate (and therefore has a bad conscience). Especially in one-shot interactions there is experimental evidence that individuals follow norms of social cooperation also in anonymous and therefore less sanctioned one-shot interactions (see e.g, for a review of lab experiments: Diekmann 2004: 491).

Following Blau, in case of unconditional incentives which are provided in advance and independent of survey participation the exchange relationship is framed as social exchange creating "unspecified obligations" (Blau 2008: 93) in the respondent to complete the survey. In the beginning exchange relationship an unconditional incentive furthermore serves as a "symbol of trust" (Dillman 1978: 16) because the researcher offered a `credit' already in advance, while otherwise (with conditional incentives) the respondent has to trust that he or she is really rewarded after survey completion. In contrast conditional incentives aren't diffuse obligations but an ordinary payment for completing the survey (framing an economic relationship in the terminology of Blau). To sum up, there are additional immaterial utility terms that are bound to unconditional incentives:

H_4: Unconditional incentives are more effective than conditional incentives.

Social exchange situations are usually characterized by low-valued exchanges in a beginning exchange relationship (Blau 2008: 94). Values are then more and more increased in course of the stabilizing exchange relationship. Due to the character of (cross-sectional) surveys as one-shot interactions, only low values will be exchanged and expected. In case of economic exchange rewards that equalize the effort are expected; incentives should hence at least reach values that approximate labor-market wages dependent on the time and effort needed for survey participation. Using higher values than usual at the beginning of a social exchange situation could erode the initial symbol of trust shifting the social exchange situation to an economic one:

H_5: There is a negative interaction between unconditional incentives and the incentive value.

So far it was assumed that individuals have a full awareness of the incentive and its value. However, given literature on limited cognitive abilities and satisficing behavior of respondents this is an unrealistic assumption (cf. the theory of bounded rationality Simon 1983; for evidence see e.g. Krosnick, Narayan, and Smith 1996). Not all respondents may notice the incentive, in particular if it is only mentioned in the cover letter. Unconditional incentives like attaching a bill to the questionnaire are physically present and hence probably more salient to sample persons than pure promises that incentives are awarded after the completion of the survey. Given that, one can expect an interaction between the "salience" of incentives and their value for respondents (see the leverage-salience theory of Groves, Singer, and Corning (2000):

H_6: There is a positive interaction between unconditional and monetary incentives.

Data and Method

Inclusion criteria and reached sample: The meta-dataset includes 365 trial effect sizes from 205 studies reported in 156 publications.[6] In total 205.216 sample persons in incentive test groups and 105.229 sample persons in the non-incentive control groups are included. All experimental trials had to meet the following inclusion criteria: a control group with no incentive treatment, a self-administered survey mode (either web or postal), an explicitly reported incentive-value and information on the number of respondents and nonrespondents.[7] Relevant publications were identified in the Cochrane Review of Edwards et al. (2009), which covers publications until 2008. Furthermore the dataset was extended by relevant publications in preceding meta-analyses (as cited in the introduction). In order to identify more recent studies, and thereby particularly web survey studies, an extensive literature search was conducted to cover the period from 2008-2013.[8]

To tackle the issue of heterogeneity, numerous variables on the level of publication, study, and trial were coded, like the incentive value, publication year, survey population and topic (for descriptive statistics on all variables used in this study see Table A1 in the appendix).[9] The comparability of incentive values across countries and study years was assured by transforming the incentive-value in US$ as well as inflation adjustment.[10]

Meta-regression techniques: Meta-analyses summarize empirical evidence from experiments, similar to narrative reports of the state of research (cf. Cooper 2010:4) computing a

---

[6] A publication may consist of several studies and each study of several experimental groups, while each study is defined by having only one control group.
[7] Self-administered survey modes are in particular relevant for the topic at hand because of their lower response rates as well as the lacking possibility of a convincing interviewer who is able to "convert" nonrespondents.
[8] To cover as many publications as possible different web-search engines (Google Scholar, Sociological Abstracts, Web of Knowledge and WISO for studies in German language) were used (cf. Falagas et al. 2008). Additionally the Web Survey Bibliography (WebSM) was used to track relevant studies.
[9] E.g. the publication year, study population, incentive value, etc.
[10] The incentive-value was inflation-adjusted using the CPI (Consumer Price Index) of the survey-country at the year of survey using auxiliary World Bank datasets (International Monetary Fund and International Financial Statistics 2012a, b) For lottery incentives the expected value was computed (in case of conditional incentives the number of respondents was used as a proxy).

mean effect size. A major drawback of meta-analyses is their sensitivity towards heterogeneity (in the analysis at hand due to the varying incentive settings or study conditions). In case of excess heterogeneity the results are only statistical artefacts comparing ‘apples with oranges’ and do not allow valid inference on treatment effects (Eysenck 1984:57; Sharpe 1997).

Meta-regression-models are able to examine this heterogeneity by controlling for „potential effect modifiers“ (Deeks, Higgins, and Altman 2008:284). For the analysis at hand a WLS-MRA (weighted-least-square meta-regression analysis) model was estimated due to better statistical performance compared to fixed or random effects MRA models (Stanley and Doucouliagos 2013). Similar to a fixed–effects MRA the inverse variance of the effect size is used to weight the effect sizes in the WLS-MRA, whereas the SEs are larger in WLS-MRA models which reduces the risk of a biased estimator (Stanley and Doucouliagos 2013: 19)

$$\ln(\mathrm{OR_i}) = \beta_0 + \beta_1 \mathrm{SE_i} + \beta_2 \mathrm{Z_i} + \mathrm{u}_i \quad (1)$$

To measure the effect of incentives (the dependent variable) logarithmic odds ratio (OR) were used. OR have better statistical properties than other measures like risk ratios (the reason is their independence of the marginal frequencies which assures that they are unbounded even for high control response rates Cook 2002:1433). The intercept (β_0) term in equation (1) represents the PET-estimate (precision-effect test) of a genuine underlying empirical effect after controlling for possible publication bias holding all moderators constant. Publication bias is caused by selectively publishing of only significant findings (Dickersin 2005: 13). Especially small and less precise studies are often not published but thrown in the scientific “file drawer” as they often do not reach statistical significance (Rosenthal 1979). Technically this censoring of insignificant studies leads to an overrepresentation of small but statistical significant effects and hence to an upwardly biased overall effect. In equation (1) the β_1 coefficient (FAT-estimate; funnel-asymmetry test) tests for publication bias by including the SE of the effect sizes (i) as a measure of study precision (Stanley and Doucouliagos 2014: 64).[11] In the presence of publication bias the FAT-coefficient would have a significant (positive) influence on the effect size. Small studies (with a high SE) should exhibit larger incentive effects than more precise studies. In order to obtain unbiased estimates of the baseline incentive effect the hybrid estimator proposed by (Stanley and Doucouliagos 2014: 71) is used that replaces the SEs in equation 1 in case of a significant PET estimate with the variance of the effect sizes (PEESE; Precision-effect estimate with SE). The coefficient vector (Z_i) includes all effect modifiers, whereas u_i represents the WLS error term.

The inclusion of multiple treatment groups compared to only one control group violates the assumption of independent observations (Borenstein 2009:238f.). To nevertheless assure unbiased estimates cluster robust SEs were used (Rogers 1994). The modified Breusch-Pagan test for unbalanced panels (Baltagi and Li 1990) indicated no random-effects[12] multi-level structure (χ^2 = 0.219; p = 0.639), therefore the simple WLS-MRA with clustered SEs is still efficient.[13]

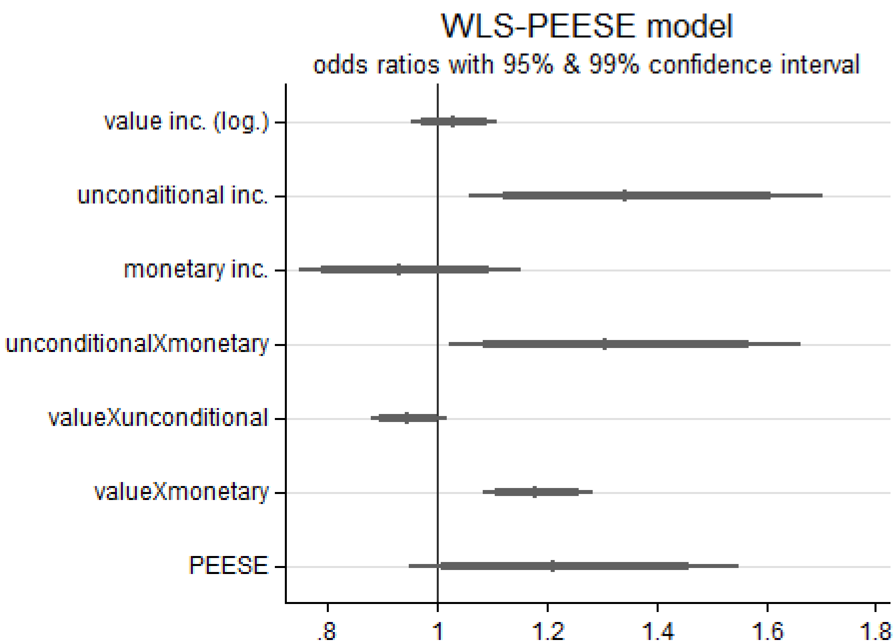
Results

For ease of interpretation, regression coefficients are displayed graphically. The underlying regression models are provided in the appendix (Table A2). Overall incentives seemed to be an effective tool to increase response rates (H_{1a}). Because of the significant PET-estimate PEESE-Models are reported. In the PEESE-model the positive incentive effect was found to be smaller compared to the FAT-PET but still clearly pronounced (OR: 1.211; 95%-CI: 1.005, 1.460; cf. Figure 1). With an assumed baseline response rate of 50% a conditional monetary incentives worth 1$ would increase the response rate by 4.8 percentage points up to 54.8% (for a graphical display

[11] The WLS publication bias test used (excluding all covariates addressing heterogeneity) is equivalent to the Eggers-test (Egger et al. 1997) routinely used.
[12] Fixed effects multi-level models are not appropriate due to the high proportion of studies with only one trial.
[13] There was no indication for random effects in the FAT-PET models as well as the models controlling for respondents burden.

see Figure 2).[14] The small difference between the PET and the PEESE-estimates may be caused by the substantial, although insignificant FAT-test, which presumes the absence of publication bias.[15] These results were robust controlling for respondents burden which was measured by survey length (with multiple imputation of missing values). The results in Figure 2 are based on the PEESE-model not controlling for respondents burden in order to include as much trial information in the model as possible (345 vs. 220 trials).[16]

Figure 1 WLS-MRA Model



[14] The odds ratios are transformed to risk differences (with control response rates on the x-axis) for ease of interpretation (Zhang and Yu 1998).

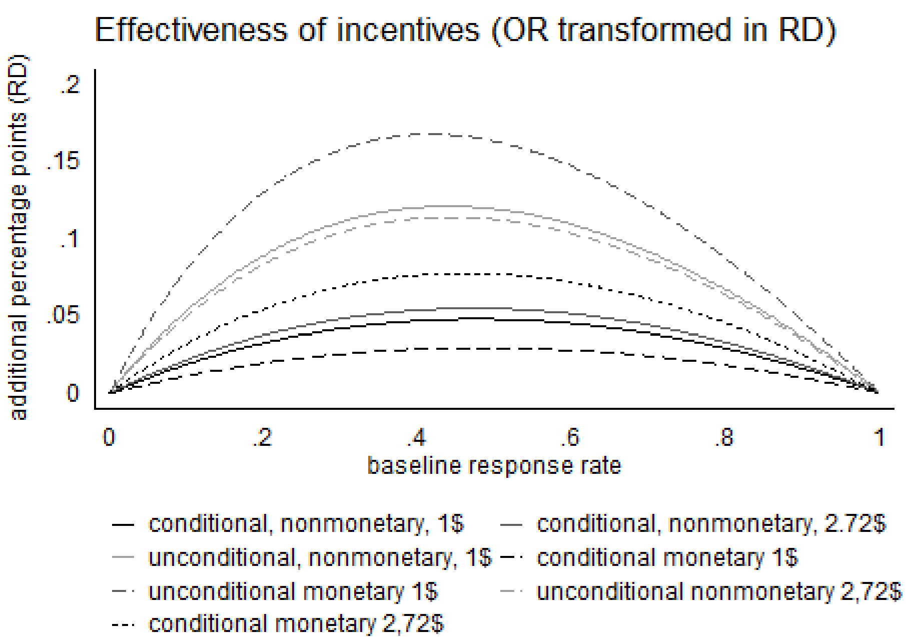
[15] In opposite to our presumptions the effect of small studies is on average smaller (OR: 0.545; 95%-CI: 0.285, 1.042), which may be caused by less accurate or more explorative incentive treatments in smaller, and therefore more inexpensive studies

[16] Nonetheless the results of the PEESE model controlling for respondents burden as well as both FAT-PET-models are reported in Table A2 the online appendix.

In line with H_{1b} the incentive value awarded had a positive but non-significant effect on the odds of response (OR: 1.028; 95%-CI: 0.971, 1,089). Contrary to H_2 awarding monetary incentives wasn't an effective tool at all and even showed a small negative, but insignificant effect (OR: 0.929; 95%-CI: 0.788, 1.094). There was, however, evidence for H_3 that monetary incentives exhibit a positive interaction effect with the incentive value (OR: 1.180; 95%-CI: 1.106, 1.259). Increasing the incentive amount by 1 log unit (e.g. from 0 to 1 log unit = 2.72$) would yield an increased response rate from 5.5 up to 7.7 percentage points. Consistent with H_4 unconditional incentives increase the odds of response substantially (OR: 1.343; 95%-CI: 1.121, 1.609). Applied to our example the 50% baseline response rate would be increased by 11.9 percentage points. In case of unconditional incentives the value of the incentive (H_5) had only a slight significant negative effect (OR: 0.945; 95%-CI: 0.894, 1.000). For unconditional incentives, increasing the incentive by 1 log unit would reduce the positive incentive effect from 11.9 to 11.3 percentage points. The significant interaction effect of unconditional and monetary incentives (H_6) increased the odds of response substantially (OR: 1.304; 95%-CI: 1.085, 1.568). In our example monetary unconditional incentives would increase the response rate in total by 16.3 percentage points up to 66.3% (while nonmonetary unconditional effects would reach a smaller increase by 11.9 percentage points). All in all a 2.72$ unconditional monetary incentive would be able to raise the assumed 50% baseline response rate up to 69.3%.

The results are stable for different model specifications (From FAT-PET to PEESE, with and without controlling for respondents burden). Only the interaction of unconditional and monetary incentives throughout weakened and turned non-significant when controlling for respondents burden. Additional analyses (not shown, but available on request) suggest that the full and reduced sample differ only slightly. Therefore the weakened interaction effect between unconditional and monetary incentives is not caused by the drastic reduction in the number of cases but due to the control variable itself. The significant negative effect of respondents' burden supports the rational-choice assumption that survey costs (esp. opportunity costs) matter. Looking at the other control variables there were no differences in the effectiveness of incentives across country, topic, or population of the survey. Differences in the effectivity occurred only for the number of reminders (which decrease the effectiveness of incentives) and web surveys (which increase the effectiveness compared to postal surveys).

Figure 2 Plot of Effectiveness of Incentives



Discussion

This paper examined the effectiveness conditions of incentives tested in 365 experimental trials reported in 205 studies conducted in more than 80 years of survey research from 1930 up to now). The meta-regression approach and the theoretical framework allowed for a more detailed identification of conditions where incentives should work best than prior studies. All in all the results are in line with a wider version of the rational-choice framework that is extended by social exchange theories and in particular norms of reciprocity: even for the one-shot interactions given with the cross-sectional surveys included in this study, sample persons were found to be stronger indebted and therefore more likely to fulfil the survey request if unconditional incentives were provided. In particular the combination of low-valued, monetary and unconditional incentives was found to increase response rates. Surprisingly, monetary incentives did not show a *per se* higher effectiveness than non-monetary ones. But the slight negative effect of monetary incentives is equalized at 1.46$ proving that higher values of monetary incentives outrun the effectivity of nonmonetary incentives.

What do these results mean for survey practice? In terms of cost efficiency the results are bad news, because the higher effectiveness of unconditional incentives means sunk costs for sample persons who still decide not to respond and hence charge the overall research budget if there are many immune “hardcore”-nonrespondents. This may lead to a dilemma between increasing response rates and saving costs. A still very effective alternative tool are low valued unconditional monetary incentives. Using monetary incentives with higher incentive amounts is an economical solution especially if the expected baseline response rates are low.

Notwithstanding its merits, the study has some limitations. The analyses at hand provide no analysis of incentive efficiency in terms of getting more respondents per dollar of the overall research budget spent (in case of unconditional incentives also nonrespondents get incentives which means sunk costs). Further research could include fixed survey costs (printed questionnaires, postage, or programming costs) to get more meaningful estimates on the *cost*-efficiency of incentives. Also the extent to which nonresponse bias is reduced by the additional respondents remains an open question: as already indicated, low response rates amplify the divergence of parameter estimates between respondents and nonrespondents (Groves 2009: 59). Increasing responses is only an effective tool for reducing nonresponse bias if additional respondents that both differ in regard to characteristics under study and are not missing at random in samples reached without incentives are convinced to participate. If not, increasing response rates might even be more harm than good. Future studies should try to use more direct indicators for nonresponse bias, which would, however, require that already the primary studies include measurements with known reference points in the target population. As long as this information is not available one has to conclude based on the results combined in this meta-analysis that it is the combination of unconditional, monetary and low-valued incentives that offers the most effective (and also very cost-efficient) way to increase survey response.

References

AAPOR [American Association for Public Opinion Research]. 2010. *Best practices for survey and public opinion research*. http://www.aapor.org/Best_Practices1.htm (last access: 25.07.2014).

Armstrong, J. Scott. 1975. "Monetary incentives in mail surveys." *The Public Opinion Quarterly* 39 (1):111-116.

Aust, Folkert, and Helmut Schröder. 2009. "Sinkende Stichprobenausschöpfung in der Umfrageforschung – ein Bericht aus der Praxis." In *Umfragen - Herausforderungen und Grenzen*, edited by Martin Weichbold, Johann Bacher and Christof Wolf, 195-212. VS Verlag für Sozialwissenschaften.

Baltagi, Badi H., and Qi Li. 1990. "A lagrange multiplier test for the error components model with incomplete panels." *Econometric Reviews* 9 (1):103-107.

Blau, Peter M. 2008. *Exchange and power in social life*. 12. print. ed. New Brunswick: Transaction.

Borenstein, Michael. 2009. *Introduction to meta-analysis*. Chichester, U.K.: John Wiley & Sons.

Boyle, Terry, Jane Heyworth, Jenny Landrigan, Robin Mina, and Lin Fritschi. 2012. "The effect of lottery scratch tickets and donation offers on response fraction: A study and meta-analysis." *Field Methods* 24 (1):112-132.

Church, Allan H. 1993. "Estimating the effect of incentives on mail survey response rates: A meta-analysis." *Public Opinion Quarterly* 57 (1):62-79.

Cook, Thomas D. 2002. "Advanced statistics: Up with odds ratios! A case for odds ratios when outcomes are common." *Academic Emergency Medicine* 9 (12):1430-1434.

Cooper, Harris M. 2010. *Research synthesis and meta-analysis a step-by-step approach*. 4. ed, *Applied social research methods series 2*. Los Angeles u.a.: Sage.

De Leeuw, Edith D, and Wim De Heer. 2002. "Trends in household survey nonresponse: A longitudinal and international comparison." In *Survey nonresponse*, edited by Robert M. Groves, Don A. Dillman, John L. Eltinge and Little Roderick J. A., 41-54. New York: Wiley.

Deeks, Jonathan J., Julian P. T. Higgins, and Douglas G. Altman. 2008. "Analysing data and undertaking meta-analyses." In *Cochrane handbook for systematic reviews of interventions*, 243-296. John Wiley & Sons, Ltd.

Dickersin, Kay. 2005. "Publication bias: Recognizing the problem, understanding its origins and scope, and preventing harm." In *Publication bias in meta-analysis: Prevention, assessment and adjustments*, edited by Hannah R. Rothstein, Alexander J. Sutton and Michael Borenstein, 11-33. Oxford: Blackwell Science.

Diekmann, Andreas. 2004. "The power of reciprocity: fairness, reciprocity, and stakes in variants of the dictator game." *Journal of Conflict Resolution* 48 (4):487-505.

Dillman, Don A. 1978. *Mail and telephone surveys the total design method*. New York u.a.: Wiley.

Dillman, Don A. 2007. *Mail and internet surveys the tailored design method*. 2. ed. Hoboken, NJ: Wiley.

Edwards, Phil, Rachel Cooper, Ian Roberts, and Chris Frost. 2005. "Meta-analysis of randomised trials of monetary incentives and response to mailed questionnaires." *Journal of Epidemiology and Community Health* 59 (11):987-999.

Edwards, Phil, Ian Roberts, Mike Clarke, Carolyn DiGuiseppi, Sarah Pratap, Reinhard Wentz, and Irene Kwan. 2002. "Increasing response rates to postal questionnaires: systematic review." *BMJ* 324 (7347):1183.

Edwards, Phil, Ian Roberts, Mike J. Clarke, Carolyn DiGuiseppi, Reinhard Wentz, Irene Kwan, Rachel Cooper, Lambert M. Felix, and Sarah Pratap. 2009. "Methods to increase response to postal and electronic questionnaires." *Cochrane Database of Systematic Reviews* (3).

Egger, Matthias, George D. Smith, Martin Schneider, and Christoph Minder. 1997. "Bias in meta-analysis detected by a simple, graphical test." *BMJ* 315 (7109):629-634.

Eysenck, H.J. 1984. "Meta-analysis: An abuse of research integration." *The Journal of Special Education* 18 (1):41-59.

Falagas, Matthew E., Eleni I. Pitsouni, George A. Malietzis, and Georgios Pappas. 2008. "Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses." *The FASEB Journal* 22 (2):338-342.

Gouldner, Alvin W. 1960. "The norm of reciprocity: A preliminary statement." *American Sociological Review* 25 (2):161-178.

Groves, Robert M. 2009. *Survey methodology*. 2nd ed, *Wiley series in survey methodology*. Hoboken, N.J.: Wiley.

Groves, Robert M. 2011. "Three eras of survey research." *Public Opinion Quarterly* 75 (5):861-871.

Groves, Robert M., Robert B. Cialdini, and Mick P. Couper. 1992. "Understanding the decision to participate in a survey." *The Public Opinion Quarterly* 56 (4):475-495.

Groves, Robert M., and Emilia Peytcheva. 2008. "The impact of nonresponse rates on nonresponse bias: a meta-analysis." *Public Opinion Quarterly* 72 (2):167-189.

Groves, Robert M., Eleanor Singer, and Amy Corning. 2000. "Leverage-saliency theory of survey participation: Description and an illustration." *The Public Opinion Quarterly* 64 (3):299-308.

Homans, George C. . 1958. "Social-behavior as exchange." *American Journal of Sociology* 63 (6):597-606.

International Monetary Fund, and International Financial Statistics. 2012a. Consumer Price Index (2005 = 100). edited by World Bank. http://data.worldbank.org/indicator/FP.CPI.TOTL?display=default (last access 25.07.2014).

International Monetary Fund, and International Financial Statistics. 2012b. Official exchange rate (LCU per US$, period average). edited by World Bank. http://data.worldbank.org/indicator/PA.NUS.FCRF/countries?display=default (last access 25.07.2014).

Kanuk, Lesue, and Conrad Berenson. 1975. "Mail surveys and response rates: A literature review." *Journal of Marketing Research* 12 (4):440-453.

Krosnick, Jon A., Sowmya Narayan, and Wendy R. Smith. 1996. "Satisficing in surveys: Initial evidence." *New Directions for Evaluation* 1996 (70):29-44.

Linsky, A.S. 1975. "Stimulating responses to mailed questionnaires: A review." *Public Opinion Quarterly* 39 (1):82-101.

Little, Eldon L., and Edwin G. Engelbrecht. 1990. "The use of incentives to increase mail survey response rates in a business environment: A field experiment." *Journal of Direct Marketing* 4 (4):46-49.

Mauss, Marcel. 1967. *The gift forms and functions of exchange in archaic societies*. New York: Norton.

Rogers, William. 1994. "Regression standard errors in clustered samples." *Stata Technical Bulletin* 3 (13):19-23.

Rosenthal, Robert. 1979. "The file drawer problem and tolerance for null results." *Psychological Bulletin* 86 (3):638-641.

Schaeffer, Nora Cate, and Jennifer Dykema. 2011. "Questions for surveys: Current trends and future directions." *Public Opinion Quarterly* 75 (5):909-961.

Sharpe, Donald. 1997. "Of apples and oranges, file drawers and garbage: Why validity issues in meta-analysis will not go away." *Clinical Psychology Review* 17 (8):881-901.

Simon, Herbert A. 1983. *Reason in Human Affairs*, *Harry Camp lectures at Stanford University*. Stanford, Calif.: Stanford University Press.

Singer, Eleanor, John Van Hoewyk, and Mary P. Maher. 2000. "Experiments with incentives in telephone surveys." *Public Opinion Quarterly* 64 (2):171-188.

Singer, Eleanor, and Cong Ye. 2013. "The use and effects of incentives in surveys." *The ANNALS of the American Academy of Political and Social Science* 645 (1):112-141.

Skinner, Burrhus Frederic. 1953. *Science and human behavior*: Simon and Schuster.

Slote, Michael. 1985. "Utilitarianism, moral dilemmas, and moral cost." *American Philosophical Quarterly* 22 (2):161-168.

Stanley, T.D., and Hristos Doucouliagos. 2013. Better than random: weighted least squares meta-regression analysis. In *Economics Series*: Deakin University, Faculty of Business and Law, School of Accounting, Economics and Finance.

Stanley, T.D., and Hristos Doucouliagos. 2014. "Meta-regression approximations to reduce publication selection bias." *Research Synthesis Methods* 5 (1):60-78.

Yu, Julie, and Harris Cooper. 1983. "A quantitative review of research design effects on response rates to questionnaires." *Journal of Marketing Research* 20 (1):36-44.

Zhang, Jun, and Kai F. Yu. 1998. "What's the relative risk? A method of correcting the odds ratio in cohort studies of common outcomes." *JAMA* 280 (19):1690-1691.

Appendix

Table A1 Descriptive Results Trials

	Mean	N	sd
value of incentive (log.)	.805	365	1.352
unconditional incentive	.652	365	0.477
monetary incentive	.726	365	0.447
highest lottery price	47	365	289.136
Survey country		365	
Northern America	*0.655*	*239*	
Europe	*0.195*	*71*	
Australia/Oceania	*0.123*	*45*	
Asia	*0.027*	*10*	
Net sample (ref. gross sample)	0.430	365	0.496
Survey population		365	
Population	*0.345*	*126*	
Health workers	*0.164*	*60*	
Customers	*0.093*	*34*	
Educational population	*0.096*	*35*	
Others	*0.301*	*110*	
Survey topic		347	
Market research	*0.323*	*112*	
Social	*0.259*	*90*	
Health	*0.320*	*111*	
Others	*0.078*	*27*	
Experimental randomization		365	
random	*0.737*	*269*	
Non random	*0.063*	*23*	
Not reported	*0.200*	*73*	
Internet survey (ref. paper and pencil)	0.148	365	0.356
Trial year (uncentered)	1994.107	365	13.035
Number of reminders	1.342	363	1.402
Length of questionnaire (in pages)	9.703	235	8.156

Table A2 WLS-Models (log OR)

	WLS-FAT-PET	WLS-PEESE	WLS-FAT-PET 2[17]	WLS-PEESE 2
value of incentive (log.)	0.0337	0.0278	0.0698	0.0668
	(0.0284)	(0.0293)	(0.0472)	(0.0474)
unconditional incentive	0.293**	0.295**	0.470***	0.477***
	(0.0891)	(0.0916)	(0.129)	(0.131)
monetary incentive	-0.0753	-0.0740	0.0413	0.0474
	(0.0813)	(0.0831)	(0.105)	(0.106)
unconditionalXmonetary	0.277**	0.266**	0.0817	0.0682
	(0.0927)	(0.0934)	(0.117)	(0.118)
valueXunconditional	-0.0573*	-0.0561*	-0.0990**	-0.0992**
	(0.0279)	(0.0284)	(0.0374)	(0.0373)
valueXmonetary	0.161***	0.165***	0.122**	0.123**
	(0.0316)	(0.0329)	(0.0442)	(0.0451)
highest lottery price[18]	0.000136	0.000124	-0.0000400	-0.0000367
	(0.000167)	(0.000164)	(0.000203)	(0.000201)
Survey country (ref. Northern America)				
Europe	0.0266	0.0269	-0.0304	-0.0306
	(0.0575)	(0.0583)	(0.0762)	(0.0766)
Australia/ Oceania	-0.0811	-0.104	0.0646	0.0505
	(0.0824)	(0.0833)	(0.101)	(0.0986)
Asia	0.240*	0.245*	0.175	0.177
	(0.113)	(0.114)	(0.116)	(0.116)
net sample	0.00978	0.00694	0.0371	0.0344
(ref. gross sample)	(0.0423)	(0.0426)	(0.0430)	(0.0428)
Survey population (ref. general population)				
Health workers	-0.00326	-0.0150	-0.0904	-0.0979
	(0.0688)	(0.0695)	(0.0895)	(0.0903)
Customers	-0.0918	-0.0993	-0.0543	-0.0533
	(0.104)	(0.103)	(0.0883)	(0.0882)
Educational population[19]	-0.00758	0.00107	0.110	0.119
	(0.0745)	(0.0750)	(0.0888)	(0.0914)
Others	-0.00990	-0.0123	0.00310	0.00371
	(0.0595)	(0.0598)	(0.0709)	(0.0719)
Survey topic (ref. market research)				
Social	-0.0499	-0.0453	0.169*	0.178*

[17] Model with respondents burden
[18] Non-lottery incentives were coded as 0
[19] University/ school teachers/ students

	(0.0554)	(0.0563)	(0.0785)	(0.0781)
Health	-0.0825	-0.0766	0.00976	0.0167
	(0.0616)	(0.0632)	(0.0733)	(0.0745)
Others	-0.102	-0.100	-0.103	-0.0970
	(0.0809)	(0.0816)	(0.0753)	(0.0747)
Not reported	-0.0902	-0.103	-0.350	-0.360
	(0.211)	(0.215)	(0.211)	(0.208)
Experimental randomization (ref. random)				
Nonrandom	0.190	0.196	0.222	0.222
	(0.101)	(0.101)	(0.166)	(0.162)
Not reported	-0.0368	-0.0450	-0.0984	-0.102
	(0.0512)	(0.0505)	(0.0526)	(0.0525)
Internet survey	0.226**	0.217*	0.166	0.155
(ref. paper and pencil)	(0.0860)	(0.0871)	(0.0958)	(0.0952)
Mean centered trial year	-0.00660*	-0.00643*	-0.00106	-0.000891
	(0.00258)	(0.00264)	(0.00181)	(0.00184)
Number of reminders	-0.0393*	-0.0396*	-0.0353	-0.0358
	(0.0176)	(0.0178)	(0.0188)	(0.0193)
SE (FAT)	-0.607		-0.422	
	(0.329)		(0.279)	
Var		-0.915		-0.677
		(0.532)		(0.472)
Length of questionnaire			-0.0137***	-0.0136***
(in pages)			(0.00292)	(0.00295)
PET/ PEESE	0.244*	0.192*	0.180	0.135
	(0.0970)	(0.0947)	(0.124)	(0.125)
Observations	345	345	220	220
R^2	0.557	0.553	0.555	0.553
Adjusted R^2	0.523	0.518	0.495	0.493

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$