Predicting Response Rates Once Again

Daniel Heimgartner[†] & Kay W. Axhausen[†]

[†]Institute for Transport Planning and Systems, ETH Zurich

August 30, 2024

Abstract

We investigate the relation between a survey's response burden and response rate, differentiating recruitment efforts and incentives paid. The results indicate that the effect of response burden is more negative than previously expected. Recruitment shifts the response curve upwards, while incentives flatten it. Surveys beyond 2‘000 points appear overly burdensome, sustaining high response rates only through recruitment coupled with incentives. Without incentives, the level effect of recruitment quickly vanishes. Contrary to previous findings, we can not identify a negative time-trend. The data, functions and workflow underlying this analysis are organized as an R-package to foster a collective effort towards understanding response rates.

Keywords: predicting, measurement, rating, survey instruments, response burden, response rate

1 Questions

Counting *response burden scores* (Schmid and Axhausen, 2019b) is not a very popular task at the Institute for Transport Planning and Systems (IVT, ETH Zurich). One or the other former PhD student got away without reporting them. However, the collective effort over the years yielded a unique dataset allowing us to understand response rates as a function of recruitment efforts and incentives. Other research institutes might therefore be encouraged to also contribute. This paper briefly elaborates the methodology once again and introduces the **responseRateAnalysis** R-package (https://github.com/dheimgartner/responseRateAnalysis) with its helper functions to easily replicate the results. Since the last update (Schmid and Axhausen, 2019a) a handful students left not until the final response burden score of their surveys was counted. Thanks to them, we feel that the time is right to update response rates results once again.

The main purpose of this paper is to replicate the analysis, update the estimates based on the enlarged sample and to make the analysis easily replicable: The R-package can be installed as explained below but we hope that other research groups rather clone the repository and contribute with their scored survey instruments according to the guidelines outlined on github so that we or they can predict response rates from time to time again.

```
> devtools::install_github("dheimgartner/responseRateAnalyis")
```

2 Methods

The main data collection effort consists of scoring the survey instruments according to Table 1.

In comparison to the previous publication (Schmid and Axhausen, 2019b), 14 additional surveys were added by IVT members. A new category (no recruitment but with incentive payments) can now be distinguished. However, only five studies fall in this category. The current state of the database is attached to the package `response_rates` and its variables documented (`?response_rates`). The full sample underlying this report can be found in Appendix A.

The distribution of the response burden scores (RBS) for the four recruitment (R) and incentive (I) categories are visualized in Fig. 1. The following abbreviations are used throughout the text: For example, *RxI* stands for the category where respondents were recruited as well as incentives paid. An

Table 1. Response burden: Points by question type and action

Item	Points
Question or transition (up to 3 lines)	2.0
Each additional line	1.0
Closed yes/no answers	1.0
Simple numerical answer (e.g., year of birth)	1.0
Rating with up to 5 possibilities	2.0
Rating with more than 5 possibilties	3.0
Left, middle, right rating	2.0
Scales with 3 and more grades	2.0
Best of ranking with cards	4.0
Second and each additional best ranking	3.0
Answer to sub-questions of up to 5 words	1.0
Answer to sub-question of up to 2 lines	2.0
a) Rsponse to half-open question with less than 8 possibilities	2.0
Each additional one	2.0
b) Response to half-open question with at least 8 possibilities	4.0
Each additional one	3.0
Answer to "please specify"	2.0
First answer to an open question	6.0
Each additional answer to the open question	3.0
Mixing showcards	6.0
Giving/showing a card to the respondent	1.0
Per response category on a showcard	1.0
Filter	0.5
Branching	0.5
For each stated choice question with 2 alternatives	2.0
For each stated choice question with 3 alternatives	3.0
For each variable of the stated choice situation and each question	1.0

Based on *Gesellschaft für Sozialforschung (GfS)*, Zürich, 2006 (updated).

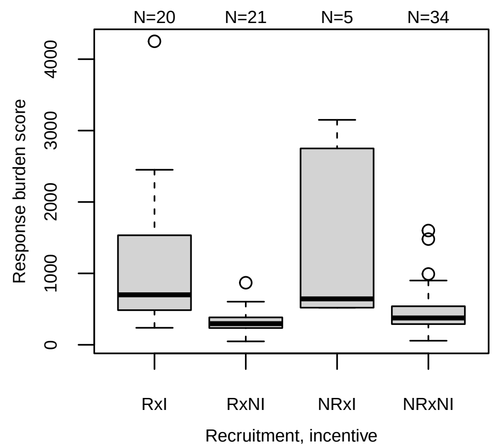

Fig. 1. Distribution of the response burden scores for *recruitment x intentive* categories. *RxI* stands for recruited, with incentives, *N* negates.

N negates and hence as an example *NRxNI* reflects the category without recruitment and incentives. The median RBS is 399 and surveys with a score higher than 1‘500 are rare (twelve points roughly correspond to a one-minute response time).

Building on Schmid and Axhausen (2019a), we estimate a logistic regression model relating response burden scores to log-transformed response rates:

$$\log\left(\frac{y_i}{100 - y_i}\right) = \beta_0 + \beta_1 \frac{x_i}{1000} + \varepsilon_i \tag{1}$$

where y_i denotes the response rate (in percent), x_i represents the ex-ante response burden score, and ε_i is a normally distributed clustered error term (similar errors for different survey waves of the same study). Observations were weighted by the square root of the sample size. The model was estimated for the entire sample (excluding surveys with a sample size less than ten) as well as separately for recruitment by incentive category. The exponential $\exp(\beta_1)$ represents a marginal change in the odds ratio (i.e., participation vs. non-participation).

If a survey instrument's response burden score increases by 100 points, the odds of participating decrease according to:

$$\left(\exp\left(\frac{\beta_1}{1000}\right) - 1\right) \cdot 100 \tag{2}$$

The basic workflow is as follows:

1. `default_data()` loads and prepares the `response_rates` data for estimation. It selects the input variables, computes the weights (`sqrt(sample_size)`), rescales the response burden score (`response_burden_score / 1000`) and log transforms the response rate (`log(response_rate / (100 - response_rate))`).

2. Fit a linear regression model with `lm()`.

3. Add clustered standard errors with `add_clustered()` which uses the **clubSandwich** (Pustejovsky, 2024) and **lmtest** package (Zeileis and Hothorn, 2002) under the hood to correct standard errors and related statistics.

4. The functions from the **texreg** (Leifeld, 2013) package (e.g., `screenreg()` or `texreg()`) work together with the class `"clustered" "lm"` (as returned by `add_clustered()`) and produce regression tables (such as the ones reported here).

3 Findings

The following code conducts the analysis for the full sample (i.e., same slope but different intercepts for the *recruitment x incentive* categories):

```
> dat <- default_data() %>%
+   filter(sample_size >= 10)  # to be consistent with previous publications
> fit <- lm(y ~ 0 + x + RxI + RxNI + NRxI + NRxNI,
+           data = dat, weights = weight)
> m1 <- add_clustered(fit, cluster = dat$survey_id, type = "CR2")
> class(m1)

[1] "clustered" "lm"
```

We repeat the above estimation for different samples, comparing the estimates of the last publication to the updated ones as well as estimate separate models for the four different recruitment and incentive categories. Table 2 summarises the results.

Table 2. Logistic regression results: Regressing response burden score on (logit-transformed) response rates

	Updated models					Old models†			
	Pooled	RxI	RxNI	NRxI	NRxNI	Pooled	RxI	RxNI	NRxNI
Response burden	-1.01^{***}	-0.31^{**}	-2.97^{***}	-1.05	-1.63^{**}	-0.58^{**}	-0.32^{**}	-1.55	-1.25^{***}
	(0.19)	(0.11)	(0.72)	(0.40)	(0.47)	(0.17)	(0.11)	(1.16)	(0.31)
RxI	1.89^{***}					1.51^{***}			
	(0.25)					(0.20)			
RxNI	0.62					0.84^{***}			
	(0.33)					(0.15)			
NRxI	-0.73								
	(0.57)								
NRxNI	-0.64					-1.11^{***}			
	(0.34)					(0.20)			
Intercept		1.23^{***}	1.49^{***}	-0.64	-0.38		1.25^{***}	1.13^{**}	-0.81^{***}
		(0.16)	(0.26)	(0.74)	(0.41)		(0.18)	(0.32)	(0.20)
R^2	0.81	0.13	0.71	0.92	0.16	0.79	0.13	0.13	0.22
Adj. R^2	0.80	0.08	0.70	0.90	0.13	0.78	0.08	0.07	0.19
Num. obs.	79	20	20	5	34	67	19	18	30
LL	-103.96	-20.97	-15.51	-2.48	-47.70	-70.72	-20.75	-13.94	-31.10
AIC	219.93	47.94	37.02	10.97	101.39	151.45	47.49	33.89	68.20
BIC	234.15	50.92	40.00	9.80	105.97	162.47	50.33	36.56	72.40

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$. †Based on the sample of the last publication. *RxI* = Recruited, with incentives, *N* negates.

For the newly added group (no recruitment but with incentive payments, *NRxI*) the effect of the response burden is not significant, as expected because of limited sample size. Due to the logistic transformation, parameters reflect changes in log-odds when the RBS marginally increases. The effect of response burden is generally more negative than previously expected. The strongest effect can be found for the recruited subsample without incentive payment (*RxNI*), where the odds of participating decrease by -0.297 (i.e., roughly 30%) according to Eq. (2) if the RBS increases by 100 points. The other comparisons are listed in Table 3.

The back-transformed relationship between response burden and response rates (response rate curve) is visualized in Fig. 2, along their confidence intervals (i.e., the shaded area reflects the uncertainty of the curve estimates and is not a prediction interval). Recruitment shifts the curve, while incentives flatten it. Notably, the domain above a response burden score of 2‘000 is sparsely populated, and the few observations potentially strongly influence the curve's shape (however, according to *Cook's distance* no influential outliers are present in our data). The results indicate that surveys beyond 2‘000 points appear overly burdensome for respondents, sustaining high response rates only through recruitment efforts combined with incentive payments, intensive care of the respondents and general interest of the respondents in the topic of these intense studies.

Table 3. Percentage change in the odds of participating if the response burden score increases by 100 points

Category	Updated models [%]	Old models [%]
Pooled	-10.09	-5.79
RxI	-3.13	-3.23
RxNI	-29.66	-15.50
NRxI	-10.51	
NRxNI	-16.30	-12.46

RxI = Recruited, with incentives, *N* negates.

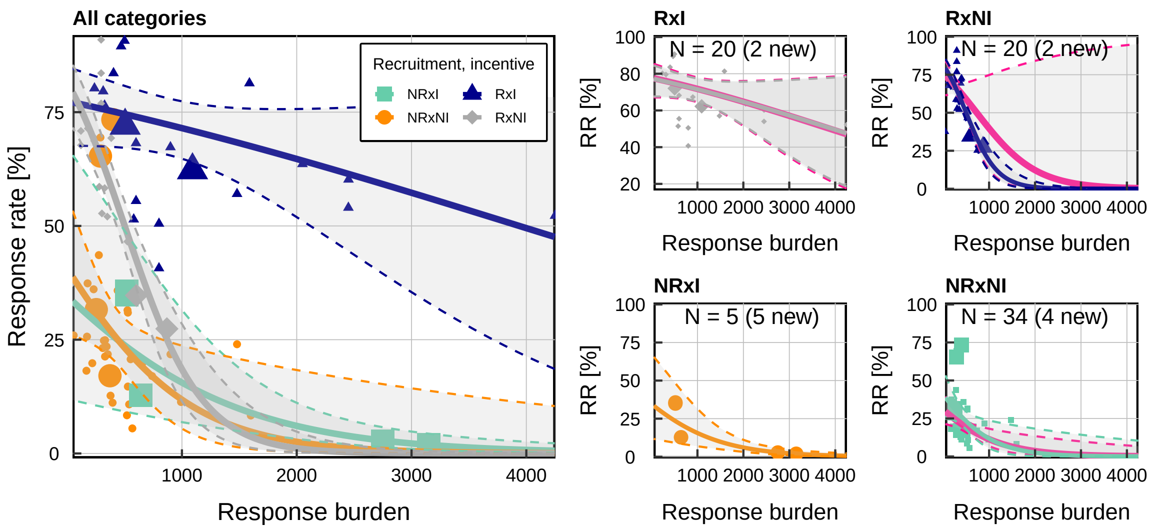

Fig. 2. Response rate curves (response rates as a function of the response burden). The left-hand panel compares the curves for each *recruitment x incentive* category based on the four separately estimated models (*RxI* stands for recruited, with incentives, *N* negates). The right-hand (smaller) panels compare the response rate curves to the ones based on the data of the previous publication (pink lines). New data points (since the last publication) are enlarged.

Generally, the 14 new data points do not dramatically change the overall shape of the curves (Fig. 2, RHS), but the curves are slightly steeper as explained based on the parameter estimates. *RxI* is almost identical (only two new data points were added). For *NRxNI* the confidence bounds increased because two of the five added surveys have unprecedented high response rates. For the category *RxNI* the function has gained support for higher response burdens which substantially steepened the curve and reduced its uncertainty. In particular, we now have higher confidence that the curve quickly joins the other response curves on the domain above 1‘500 response burden points. I.e., recruitment without incentive payments only matters for surveys with low response burden (but can make a big difference there).

Similar to Schmid and Axhausen (2019b), we can add a linear time-trend with the year 2004 (when the survey scoring effort started) serving as the base. The following code shows the trivial addition of the time-trend to the pooled model.

```
> dat$year <- dat$year - 2004
>
> fit_t <- fit_t <- lm(y ~ 0 + x + RxI + RxNI + NRxI + NRxNI + year,
+                      data = dat, weights = weight)
>
> mt <- add_clustered(fit_t, cluster = dat$survey_id, type = "CR2")
```

We repeat the steps for the individual categories and synthesise the results in a table (Table 4). In contrast to Schmid and Axhausen (2019b) we do not find a negative time-trend and therefore do not support the hypothesis of a general fatigue and less willingness to participate in our surveys.

Table 4. Logistic regression results: Adding a linear time-trend

	No time-trend					With time-trend				
	Pooled	RxI	RxNI	NRxI	NRxNI	Pooled	RxI	RxNI	NRxI	NRxNI
Response burden	-1.01^{***}	-0.31^{**}	-2.97^{***}	-1.05	-1.63^{**}	-1.02^{***}	-0.33^{*}	-3.28^{**}	-1.41^{*}	-1.50^{**}
	(0.19)	(0.11)	(0.72)	(0.40)	(0.47)	(0.20)	(0.13)	(0.85)	(0.24)	(0.44)
RxI	1.89^{***}					1.82^{***}				
	(0.25)					(0.26)				
RxNI	0.62					0.55				
	(0.33)					(0.46)				
NRxI	-0.73					-0.86				
	(0.57)					(0.58)				
NRxNI	-0.64					-0.72^{**}				
	(0.34)					(0.24)				
Intercept		1.23^{***}	1.49^{***}	-0.64	-0.38		1.06^{*}	1.49^{***}	-12.31	-0.71^{*}
		(0.16)	(0.26)	(0.74)	(0.41)		(0.43)	(0.32)	(8.03)	(0.34)
Time-trend						0.01	0.02	0.01	0.68	0.03
						(0.02)	(0.05)	(0.06)	(0.46)	(0.05)
R^2	0.81	0.13	0.71	0.92	0.16	0.81	0.14	0.72	0.97	0.20
Adj. R^2	0.80	0.08	0.70	0.90	0.13	0.80	0.04	0.68	0.95	0.14
Num. obs.	79	20	20	5	34	79	20	20	5	34
LL	-103.96	-20.97	-15.51	-2.48	-47.70	-103.83	-20.81	-15.38	0.06	-46.95
AIC	219.93	47.94	37.02	10.97	101.39	221.65	49.61	38.76	7.89	101.89
BIC	234.15	50.92	40.00	9.80	105.97	238.24	53.59	42.74	6.33	108.00

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$. *RxI* = Recruited, with incentives, *N* negates.

4 Acknowledgements

We would like to thank all the students scoring the survey instruments and contributing to the data underlying this analysis.

References

Leifeld, P. (2013). texreg: Conversion of statistical model output in R to LaTeX and HTML tables. *Journal of Statistical Software 55*(8), 1–24.

Pustejovsky, J. (2024). *clubSandwich: Cluster-robust (sandwich) variance estimators with small-sample corrections.* R package version 0.5.11.

Schmid, B. and K. W. Axhausen (2019a). Predicting response rates further updated. Technical report, IVT, ETH Zurich, Zurich.

Schmid, B. and K. W. Axhausen (2019b). Predicting response rates of all and recruited respondents: A first attempt. *Transport Findings*.

Zeileis, A. and T. Hothorn (2002). Diagnostic checking in regression relationships. *R News 2*(3), 7–10.

A Full data sample

```
> rr <-
+   response_rates %>%
+   select(year, survey_content, sample_size, response_burden_score, response_rate)
>
> names(rr) <- c("Year", "Study", "Sample size", "Burden score",
+                "Response rate [%]")
>
> tex <- kableExtra::kbl(rr, format = "latex",
+                        caption = "Updated response burden scores",
+                        label = "full_sample", digits = 0,
+                        longtable = TRUE,
+                        booktabs = TRUE) %>%
+   kableExtra::kable_styling(latex_options = c("repeat_header")) %>%
+   kableExtra::landscape()
```

Table 5. Updated response burden scores

Year	Study	Sample size	Burden score	Response rate [%]
2004	National SP survey on railway services	1561	120	68
2006	Regional mode and route choice SP	1229	120	71
2007	National SP on value of travel time savings	2317	303	53
2004	Regional SR on value of statistical life	500	440	36
2006	Regional SR on value of statistical life	1900	526	31
2005	Home ownership and use of local facilities	9330	231	36
2007	National SP on the impacts of road pricing	2227	524	47
2005	Mobility biographies and regular travel	3290	521	8
2005	Mobility biographies	1645	529	31
2005	Mobility biographies	1435	529	15
2004	Mobility biographies and home ownership	297	655	20
2006	Social network and mobility biographies	2714	992	11
2008	Mobility Plan University	363	219	20
2008	Mobility Plan USZ	1615	57	26
2009	Fuel price and rail usage	993	327	58
2009	Modelling mountaineers' travel behaviour	530	276	44
2009	Snowball sample	826	900	67
2009	Snowball sample	312	900	22
2009	Snowball sample	142	1480	57
2010	Snowball sample	50	1480	24
2010	Diary induced traffic, pen-and-paper	200	800	50
2010	Diary induced traffic, online	140	800	41
2010	2000 Watt society, pretest 1	51	326	76
2010	2000 Watt society, pretest 2	49	314	80
2010	2000 Watt society, main study	491	238	80
2010	ARE SP, pretest - mode choice only	99	235	69
2010	ARE SP, pretest - route choice only	29	280	59
2010	ARE SP, pretest - mode and route choice	484	384	72
2010	ARE SP, main study - mode choice only	893	235	69
2010	ARE SP, main study - route choice only	215	280	67

Table 5. Updated response burden scores *(continued)*

Year	Study	Sample size	Burden score	Response rate [%]
2010	ARE SP, main study - mode and route choice	3994	384	69
2011	Residential choice (Otte, no addresses)	1200	320	25
2011	Residential choice (Otte, with addresses)	1200	330	21
2011	Residential choice (Own items, no addresses)	1200	344	24
2011	Residential choice (Own items, with addresses)	1200	354	22
2011	Grimsel user SP	399	180	71
2011	Survey on bus and tram use	3177	310	23
2011	Survey on parking behaviour	1248	404	84
2012	SP survey on travel time reliability	491	400	73
2011	BABS SC (Evacuation)	4049	330	25
2012	Car sharing / pooling	1683	350	52
2012	BMVBS Zeitkosten, schriftlich	3355	600	68
2012	BMVBS Zeitkosten, online	209	600	56
2012	BMVBS Zeitkosten, gewerblich	925	500	91
2012	Climate Change Influence on Swiss Transport - Interviews	16	48	38
2013	Climate Change Influence on Swiss Transport - Written Questionnaire	5	165	80
2013	Mobility Biographies	288	1600	8
2014	Climate Change Influence on Swiss Transport - Online Questionnaire	55	168	18
2014	Mobility-Pilotprojekts zu free-floating Carsharing - Mobility-Kunden	2224	173	26
2014	Mobility-Pilotprojekts zu free-floating Carsharing - Catch a Car-Kunden	527	178	37
2014	Masterarbeit Verkehr und Soziale Netzwerke	208	580	51
2015	Masterarbeit Arbeitsplatzwahl (freie Wahl einer Arbeitsstelle)	265	290	69
2015	Masterarbeit Arbeitsplatzwahl (Wechsel der Arbeitsstelle)	11	296	91
2015	Masterarbeit Arbeitsplatzwahl (Wechsel der Arbeitsstelle)	140	296	84
2015	ARE SP 2015	6099	296	77
2015	Post-Car World (Pre-Test: Stage 1,2,3)	67	4250	52
2015	Post-Car World (Wave 1: Stage 1,2,3)	137	2450	54
2015	Post-Car World (Wave 2: Stage 1,2,3)	191	2451	60
2016	Post-Car World (Wave 3: Stage 1,2)	118	2050	64
2017	Pretest: Social Networks, Mobility Behaviour and Societal Impacts (1st survey part)	500	740	17

Table 5. Updated response burden scores *(continued)*

Year	Study	Sample size	Burden score	Response rate [%]
2017	Pretest: Social Networks, Mobility Behaviour and Societal Impacts (2nd survey part)	57	470	89
2017	Social Networks, Mobility Behaviour and Societal Impacts (1st survey part)	12000	553	21
2017	Social Networks, Mobility Behaviour and Societal Impacts (2nd survey part)	1706	1588	81
2018	SVI (pretest): Einfluss nicht verkehrlicher Variablen: Neuzuzüger	241	540	11
2018	SVI (pretest): Einfluss nicht verkehrlicher Variablen: Eingesessene	252	378	13
2018	SVI (main survey): Einfluss nicht verkehrlicher Variablen: Neuzuzüger	4825	568	6
2018	SVI (main survey): Einfluss nicht verkehrlicher Variablen: Eingesessene	4601	396	11
2017	Automated Vehicles main study (Stage 1,2,3)	482	1092	62
2021	Kontext: Yumuv. Inhalte: Haushalt, Verkehrsmittel.	10000	643	13
2022	Swiss Value of time study (pre-test & wave 1)	2545	520	36
2022	Swiss Value of time study (wave 2)	2548	520	35
2020	Swiss Mobility Panel Wave 1	27417	604	35
2021	Swiss Mobility Panel Wave 2	9442	398	73
2022	Swiss Mobility Panel Wave 3	9092	290	65
2022	Swiss Mobility Panel Wave 4 (sample refresh)	11000	869	27
2022	TimeusePlus (pre-test)	7500	2750	3
2023	TimeusePlus (main study)	69000	3150	2
2022	Multimodality in the Swiss New Normal (SNN): Pre-study	7876	373	17
2023	Multimodality in the Swiss New Normal (SNN): Stage 1 (all)	10230	254	32
2023	Multimodality in the Swiss New Normal (SNN): Stage 2-3 (telework eligible)	1280	505	72