Appendix A

Maximum Likelihood Estimation

The likelihood function for both the standard Poisson model and the Poisson extension to the unrelated question model (UQMP) is

$$L(\pi, \lambda) = \prod_{j=1}^{m} \prod_{i=1}^{n_j} \Big[P(\text{“yes”} \,|t_j)^{a_j} \cdot P(\text{“no”} \,|t_j)^{b_j} \Big] \tag{A1}$$

with the number of groups m, the sample size per group n_j, the time frame per group t_j that the question refers to, the observed number of yes-answers per group a_j and the observed number of no-answers per group b_j. For the two models, $P(\text{“yes”} \,|t_j)$ is computed via the Equations 5 or 10, for the standard Poisson model or the UQMP, respectively. $P(\text{“no”} \,|t_j)$ can be computed per $1 - P(\text{“yes”} \,|t_j)$. Taking the log of Equation A1 gives

$$\log L(\pi, \lambda) = \sum_{j=1}^{m} \sum_{i=1}^{n_j} \Big[a_j \cdot \log[P(\text{“yes”} \,|t_j)] + b_j \cdot \log[P(\text{“no”} \,|t_j)] \Big]. \tag{A2}$$

Maximizing equation A2 by a numerical search routine yields the maximum likelihood estimators for the parameters π and λ.

Appendix B

Statistical analysis with R

In this appendix, we give an example for the statistical analysis of our data with the proposed UQMP. The full analysis script and data set can be downloaded at https://osf.io/5pkm4/.

```
## options
# setting RNG seed to a fixed value, so the bootstrap procedure
# yields the same results in every run
set.seed(123)


## design parameters and observed variables
# probability p of getting sensitive question
p <- 245.25/365.25
# probability q of a yes-answer to the neutral question
q <- 181.25/365.25


# values for parameter estimation
lim        <- 1e-10              # lower limit for lambda
up_lim_lam <- 10                 # upper limit for lambda


# time frames t_j for all groups j = 1, 2, 3, 4
t0 <- c(.25, 1, 6, 12)


# number of bootstrap samples
nb <- 1000


# group sizes N_t
```

```
N.t.uqm <- c(672, 670, 654, 655)
# yes-answers a_t
a.uqm   <- c(283, 272, 276, 277)
# no-answers b_t
b.uqm   <- c(389, 398, 378, 378)

## functions
# function of the UQMP
pc.uqm  <- function(t,p,q,PI,lam){p*PI*(1-exp(-lam*t))+(1-p)*q}

# log-likelihood-function
MLE.uqm <- function(par, a, b){
        PI      <- par[1]
        lam     <- par[2]
        pyes    <- pc.uqm(t0,p,q,PI,lam)
        lLu     <- a*log(pyes) + b*log(1-pyes)
        MLE.uqm <- -sum(lLu)
}

# G function for testing model fit
Gf.uqm <- function(par, a, b){
        PI      <- par[1]
        lam     <- par[2]
        N.t     <- a + b
        E.t.yes <- pc.uqm(t0,p,q,PI,lam)*N.t
        E.t.no  <- N.t - E.t.yes
        G.uqm   <- 2*sum(a*log(a/E.t.yes) +
```

```
                                b*log(b/E.t.no))
        return(G.uqm)
}

## parameter estimation via bootstrap sampling
# "vectorizing" observed yes- and no-answers for every group
obs.t1.uqm <- c(rep(1, a.uqm[1]), rep(0, b.uqm[1]))
obs.t2.uqm <- c(rep(1, a.uqm[2]), rep(0, b.uqm[2]))
obs.t3.uqm <- c(rep(1, a.uqm[3]), rep(0, b.uqm[3]))
obs.t4.uqm <- c(rep(1, a.uqm[4]), rep(0, b.uqm[4]))

# bootstrap sampling
PI.b.uqm  <- numeric(nb)
lam.b.uqm <- numeric(nb)

for(i in 1:nb){
  # resampling a and b from observed data
  a.b.uqm <- c(sum(sample(x = obs.t1.uqm, size = N.t.uqm[1],
                          replace = TRUE)),
               sum(sample(x = obs.t2.uqm, size = N.t.uqm[2],
                          replace = TRUE)),
               sum(sample(x = obs.t3.uqm, size = N.t.uqm[3],
                          replace = TRUE)),
               sum(sample(x = obs.t4.uqm, size = N.t.uqm[4],
                          replace = TRUE)))
  b.b.uqm <- N.t.uqm - a.b.uqm
```

```
  # maximum likelihood estimation of the redrawn sample
  ML.uqm <- optim(par = c(0.5, 0.8), a = a.b.uqm, b = b.b.uqm,
                  fn = MLE.uqm, method = 'L-BFGS-B',
                  lower = c(lim, lim),
                  upper = c(1-lim, up_lim_lam))

  # extracting pi and lambda estimates
  PI.b.uqm[i]  <- ML.uqm$par[1]
  lam.b.uqm[i] <- ML.uqm$par[2]
}

# extracting parameter estimators, SEs and 95%-CIs
# point estimate for pi
PI.m.uqm   <- mean(PI.b.uqm)
# standard error for pi
PI.se.uqm  <- sd(PI.b.uqm)
# 95%-CI for pi
PI.ci.uqm  <- quantile(PI.b.uqm, c(.025, .975))

# point est. for lambda
lam.m.uqm  <- mean(lam.b.uqm)
# se for lambda
lam.se.uqm <- sd(lam.b.uqm)
# 95%-CI for lambda
lam.ci.uqm <- quantile(lam.b.uqm, c(.025, .975))

# G-test
```

```
Gest.uqm   <- optim(c(0.5,0.8), a = a.uqm, b = b.uqm,
                    fn = Gf.uqm,
                    method='L-BFGS-B',
                    lower = c(lim, lim),
                    upper = c(1-lim, up_lim_lam))

# extracting G-value
Gval.uqm <- Gest.uqm$value
# computing p-value for G-test decision
pval.uqm <- pchisq(Gest.uqm$value, df = 1, lower.tail = FALSE)
```

Appendix C

Follow-up study: Estimating blue eye color prevalence with a UQM curtailed sampling approach

Introduction

To investigate a possible explanation for the unforeseen results of the main study, namely the unexpected difference between prevalence estimates of blue eye color via DQ and UQM (see Hypothesis 4 of the main study), a follow-up study was preregistered (see Iberl et al., 2022c) and conducted. To be precise, we tested whether the unexpected results could have emerged due to order effects of the posed questions. The basic idea of the follow-up study was to collect further data from a sample confronted with the UQM method, but while exchanging the order of the drinking and driving and eye color questions. If an order effect was responsible for the unforeseen results, swapping the order of the questions should lead to a prevalence estimate of blue eye color via UQM that is similar to the one generated by the DQ method in the main study.

Because of this rather simple premise and to reduce sample sizes, we opted for a *sequential sampling application* for the UQM as proposed by Reiber, Schnuerch, and Ulrich (2022). Contrary to classical statistical methods, the basic idea of sequential testing is to stop the sampling process as soon as sufficient information is generated to align with a preset hypothesis. One variant of sequential sampling, *curtailed sampling*, was proposed by Wetherill (1975) (for a detailed description, see Reiber, Schnuerch, & Ulrich, 2022). In this method, certain stopping rules are set before the sampling process. Those are defined by a maximum sample size, N_{max}, defining the maximum needed sample size to align with one of the hypotheses, and c_s, a limited number of observed "successes" (here: yes- answers) needed to reject the null hypothesis. In turn, the null hypothesis is confirmed if $c_f = N_{max} - c_s + 1$ "failures" (here: no-answers) are observed. N_{max} and c_s are determined via power analysis, depending on the preset hypotheses H_0 and H_1 and on the preset error probabilities α (to falsely reject H_0) and β (to falsely reject H_1). Reiber, Schnuerch, and Ulrich (2022) created applications for curtailed sampling to RRMs, including the UQM,

which we used for the follow-up study.

In the follow-up study, we calculated a sequential binomial hypothesis test according to Reiber, Schnuerch, and Ulrich (2022). We set $\pi \leq 0.387$ as the null hypothesis, stating that the prevalence for blue eye color prevalence is lower than or equal to the upper boundary of the 95% confidence interval for the blue eye color estimated via DQ in the main study. As alternative hypothesis, we set $\pi \geq 0.489$, stating that the mentioned prevalence is at least equal to the lower boundary of the 95% confidence interval for the prevalence estimated via UQM in the main study. α and β were set to 0.05 each. The UQM variables p and q were set to the same values as in the main study, so to circa 0.67 and 0.5, respectively. With these variables, the power analysis resulted in the maximum possible sample size of $N_{max} = 579$. After $c_s = 265$ yes-answers or $c_f = 315$ no-answers would be observed, sampling would be stopped.

We predicted that there would be an order effect in accordance to our explanation for the unexpected results. So, we hypothesized that H_1 would be rejected, meaning that the estimate of blue eye color prevalence is in line with the corresponding DQ estimate of the main study.

Method

The follow-up study amounted to a simple one-group-design without experimental manipulations.

Like in the main study, we commissioned *Bilendi S.A.* to generate a sample with demographics as similar as possible to those of German citizens with a driver's license (see Kraftfahrt-Bundesamt [Federal Office for Motor Traffic], 2022).

The procedure and materials used in the follow-up study were identical to those in the main study. The only differences were that in the follow-up study, there was no DQ group, and the order of the eye colour question and the drinking and driving question was switched. Additionally, since it was not necessary to vary the time constraints in this design, the drinking and driving question always referred to the past year. The data

exclusion procedure was similar to the one used in the main study as well. But, to make sure that the sampling process was not stopped prematurely, participants who answered the survey too fast were marked as fast respondents immediately after completing the survey. To do this, we assumed the average response timings in the follow-up study to be similar to the response times of the UQM group in the main study. This seemed like a valid assumption, since the surveys were identical except for the order of two questions. Thus, we compared the RSI values (according to Leiner, 2019b) of the participants directly with the corresponding RSI values in the UQM group in the main study.

Sampling started on September 16th, 2022 and was completed on September 20th, 2022.

Results

Sampling was stopped after $N = 498$ observations. Because the stopping criterion of 265 yes-answers was reached, the null hypothesis was rejected.

The sample was similarly distributed in terms of demographic data compared with the data for Germans with driver's licenses (Kraftfahrt-Bundesamt [Federal Office for Motor Traffic], 2022), as illustrated in Table C1.

To estimate π, we used an estimator for the parameter λ that corrects for bias induced by the sequential sampling procedure (Reiber, Schnuerch, & Ulrich, 2022),

$$\hat{\gamma} = \frac{c_s - 1}{N - 1}. \tag{C1}$$

Inserting the corrected estimate $\hat{\gamma}$ into the standard formulae for calculating π in the UQM (see Equations 7 and 9) yields the point estimate $\hat{\pi} = 0.548$ (95% CI [0.483, 0.614]).

Discussion

In conclusion, the hypothesis test points toward the follow-up study's UQM estimate for the prevalence of blue eye color not being significantly lower than than the one in the main study. Thus, an order effect seems unlikely as the explanation for the unexpected differences between blue eye color prevalence estimates using DQ and UQM

Table C1

Distribution of demographics in the sample of the follow-up study compared to those of the German population owning a driver's license

Demographic		Distribution	
		sample	population
Gender	female	46.2%	43.1%
	male	53.6%	56.9%
	non-binary	0.2%	0.0%
Age	18-29 years	15.1%	16.8%
	30-39 years	19.7%	20.1%
	40-49 years	16.7%	14.2%
	50-59 years	19.1%	16.7%
	60 years and older	29.5%	31.8%

Note. The reference distribution of demographics is based on data by the Kraftfahrt-Bundesamt [Federal Office for Motor Traffic] (2022).

methods in the main study.