Optimizing Count Responses in Surveys: A Machine-Learning Approach

Dr. Qiang Fu[*]

Department of Sociology, The University of British Columbia,

Dr. Xin Guo

Department of Applied Mathematics, The Hong Kong Polytechnic University,

Dr. Kenneth C. Land

Department of Sociology, Duke University

August 7, 2017

Abstract

Count responses with grouping and right censoring have long been used in surveys to study a variety of behaviors, status and attitudes. Yet, grouping or right-censoring decisions of count responses still rely on arbitrary choices made by researchers. We develop a new method for evaluating grouping and right-censoring decisions of count responses from a (semi-supervised) machine-learning perspective. This paper uses Poisson-multinomial mixture models to conceptualize the data generating process of count responses with grouping and right censoring, and demonstrates the link between grouping-scheme choices and asymptotic distributions of the Poisson mixture. To search for the optimal grouping scheme maximizing objective functions of the Fisher information (matrix), an innovative three-step M algorithm is then proposed to process infinitely many grouping schemes based on Bayesian A-, D- and E-optimality. A new R package is developed to implement this algorithm and evaluate grouping schemes of count responses. Results show that an optimal grouping scheme not only leads to a more efficient sampling design but also outperforms a nonoptimal one even if the latter has more groups.

Keywords: survey methodology, optimality, experimental design, search algorithm, machine learning, Fisher information, zero inflation, right censoring, Poisson distribution

1 Introduction

The design of count responses in surveys is a common yet understudied topic in social sciences. Although the collection of exact counts of frequencies or incidence in social, epidemiological and demographic surveys (e.g., number of births, frequencies of delinquent behaviors, incidents of diseases and counts of social contacts) is analytically appealing, actual count responses in survey

questions often consist of grouped counts (e.g., one response category "3–4 times" instead of two separate "3 times" and "4 times" categories) or are right-censored (e.g., the upper end response category as "6 or more times"). In fact, such grouped and right-censoring (GRC) count responses have long been adopted by social scientists to study a range of behaviors, events and attitudes (Akers et al. 1989; Bachman et al. 1990; Bailey et al. 1992; Barnes et al. 2006; Basu and Famoye 2004; Fu et al. 2013, 2016; Hagan et al. 2005; Marsden 2003; Reardon and Raudenbush 2006; Schaeffer and Dykema 2011; Straus et al. 1990; Thoits and Hewitt 2001). Scholars often find that GRC responses are useful to study sensitive research topics (e.g., juvenile delinquency, domestic violence and drug use) or to solicit information from respondents with less cognitive capacity (e.g., young adolescents or the oldest old). For example, one nationally representative survey project in the United States, the Monitoring the Future study (MTF or the National High School Senior Survey), has used GRC responses to track annual trends of delinquency and substance use among U.S. high school seniors since the 1975. Such GRC count responses have also been used by the National Longitudinal Study of Adolescent Health (Add Health) to study adolescent behaviors at home, school or neighborhood.

As documented in existing literature (Bradburn et al. 2004; Schwarz et al. 1985), the design of GRC count responses has a direct impact on the estimation of behavioral or cognitive frequencies. For example, an experimental study shows that the choice of right-censored count categories influences the estimation of TV-watching time Schwarz et al. (1985). Yet, the design of GRC count responses is still arbitrarily determined by survey investigators. This practice is surprising given the abundant presence of count responses with either right censoring or grouping or both in surveys. Under certain scenarios, determining the optimal grouping scheme of GRC responses has been implemented by sophisticated statistical procedures and context-specific research designs, de-

pending on the other variables of interest. For example, the intrinsic or contingent ordering of log-multiplicative association models may provide the optimal grouping scheme if conditional or joint distributions of variables in contingency tables are provided (Goodman 1987; Wong 2010; Smith and Garnier 1986). Likewise, given the extensive debates over the conceptualizations of gradations of democracy (Bollen 1990; Cheibub et al. 1996), it is found that the validity of dichotomous and graded measures of democracy can be evaluated by projecting their qualitative difference into two essential indicators related to democracy, international conflict and regime stability (Elkins 2000). Although these innovative studies provide useful tools for scholars to assess grouping decisions for counts that are intrinsic to specific research questions at stake, their statistical procedures or research designs require additional information on the distribution of the ungrouped outcome variable and its association with other variables. Nevertheless, the use of GRC count responses often means that investigators have yet to understand the distribution of counts that are extrinsic to a specific research question, let alone its association with other variables. A search algorithm for the optimal grouping scheme focusing exclusively on the outcome variable per se rather than its research context is therefore useful and readily facilitates the evaluation of alternative grouping schemes with different a priori assumptions.

Applying the theory of optimal experimental designs (Atkinson et al. 2007; De Leon and Atkinson 1991; Dette et al. 2004; Minkin 1987), we propose an innovative three-step algorithm for searching the parameter space and generating optimal grouping decisions for GRC count responses. In the machine-learning literature, optimal experimental design is also referred as a special case of semi-supervised machine learning (or active learning) because a learning/search algorithm interacts with users (survey investigators for the current research) to obtain optimal outputs from the parameter space (Settles 2010; Cohn et al. 1996). Based on a Poisson-multinomial

4

mixture distribution, this paper begins with configuring the data generating process of GRC count responses and develops related maximum likelihood estimators (MLE). Two members of the Poisson family of frequency distributions, the Poisson and Zero-Inflated Poisson (ZIP), are studied in detail. Combined with prior Poisson distribution parameters, the Fisher Information (matrix) of the maximum likelihood estimator is then employed to implement a new M search algorithm using Bayesian A-, D- and E-optimality. An R package `GRCdata` currently consisting of two functions `find.scheme` and `grcmle` has been written to assess the grouping decisions of count responses.

2   Grouped and Right-Censored Count Data

Before discussing optimal designs for GRC count data, the question arises as to why such response categories have been adopted by social scientists. As examples, in various surveys respondents are asked to list their numbers of close friends, weekly frequencies of alcohol intake, incidents of criminal victimization in the recent six months, times of illness in the last year and life-time history of residential moves. Admittedly, a precise enumeration of exact counts is methodologically appealing for two reasons. First, exact counts can be readily analyzed by existing statistical tools (e.g., Poisson regression models) and software packages. Second, survey investigators do not have to deal with arbitrary grouping or right-censoring decisions. Yet, one major problem encountered by survey investigators is that the precise enumeration of counts can impose a cognitive burden on interviewees and sometimes leads to excessive missing data. In other words, the GRC data structure is a compromise between what survey investigators want and what respondents are willing or able to offer (Groves et al. 2011; Schaeffer and Presser 2003). For example, although medical sociologists and psychiatrists would like to know exactly how many days in the past week respon-

dents experienced a variety of depressive symptoms, respondents, especially these with depressive symptoms, often get frustrated when required to distinguish between, for example, 2 days and 3 days. Thus, the CES-D scale, an established self-report depression measure, offers four grouped response categories: less than 1 day, 1-2 days, 3-4 days and 5-7 days (Radloff 1977). For a study on elder adults aged 65 and above, a pre-test showed that respondents were unwilling to answer even the four grouped response categories of the CES-D scale so researchers had to further collapse the four grouped categories and used a dichotomous measure instead (Blazer et al. 1991).

Likewise, for research topics that are perceived as sensitive or less socially desirable, such as personal income, number of sex partners, incidents of delinquent behaviors and history of drug use, respondents feel more comfortable in reporting grouped or right-censored categories instead of exact numbers (Sudman et al. 1996). It is not surprising that most, if not all, questions related to the frequency of juvenile delinquency and drug use in both the Monitoring the Future (MTF) study and The National Longitudinal Study of Adolescent to Adult Health (Add Health) adopted GRC responses.

Even if respondents are willing to collaborate, the difficulty in recalling the exact number of events that happened some time (e.g., several months) ago makes the exact number of events unreliable and introduces additional measurement errors (Groves et al. 2011; Schaeffer and Presser 2003). Similarly, if listing the total number of events requires extra efforts during field interview, fatigue of interviewers can result in underreported numbers of events. For example, as interviewers were instructed to probe for more discussion partners, it has been shown that interviewer effects (e.g., the failure to elicit more private network data) contributed to the extensive debates concerning increasing social isolation in the United States (Paik and Sanchagrin 2013).

3   Generating GRC Count Data

In order to define the optimality for objective functions of GRC grouping schemes, we first con-

figure a data generating process for GRC count responses. The Poisson distribution is often used

to model count data with probability mass:

$$f(y|\lambda) = e^{-y}\frac{\lambda^y}{y!}, \quad y = 0, 1, 2, \ldots, \tag{1}$$

where y is a random count variable and λ is both the mean and the variance of the Poisson dis-

tribution. To define a Poisson-based likelihood function for GRC count data, we propose a data-

generating scheme in the form of a Poisson-multinomial process. Similar Poisson-multinomial

models were previously used to study contingency tables and traffic accidents (Lang 2004; Lord

et al. 2005).

We let $\mathcal{G} = \{I_j\}_{j=1}^N$ denote a GRC grouping scheme with N groups (i.e., the total number of

response categories) and consecutive subsets $I_1, \ldots, I_N$ of nonnegative integers $\{0, 1, 2, \ldots\}$. For

identically and independently distributed (iid) observations x_i's from a Poisson(λ) distribution, we

have,

$$\alpha_j(x_i) = \begin{cases} 1, & \text{when } x_i \in I_j, \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

In other words, we have a N-dimensional random vector $(\alpha_1, \ldots, \alpha_N)$ denoting the GRC re-

sponses. For example, $(\alpha_{\{0\}}, \alpha_{\{1\}}, \alpha_{\{2\}}, \alpha_{\{3\text{-}5\}}, \alpha_{\{6\text{-}9\}}, \alpha_{\{10+\}})$ denotes the GRC response cate-

gories *never*, *once*, *twice*, *3-5 times*, *6-9 times*, *10 and more times*. Note that for any given obser-

vation in a survey sample, there is one and only one component α_j for $(\alpha_1, \ldots, \alpha_N)$ that equals

1. This N-dimensional vector then has a multinomial distribution $\mathcal{M}(1, \theta_1, \ldots, \theta_N)$, where the

parameters θ_j depend on the parameter λ of the underlying Poisson(λ) distribution:

$$\theta_j(\lambda) = \sum_{y \in I_j} e^{-\lambda} \frac{\lambda^y}{y!}, \quad j = 1, \ldots, N. \tag{3}$$

For example, the multinomial distribution corresponding to $(\alpha_{\{0\}}, \alpha_{\{1\}}, \alpha_{\{2\}}, \alpha_{\{3\text{-}5\}}, \alpha_{\{6\text{-}9\}}, \alpha_{\{10+\}})$

is

$$\mathcal{M}\left(1, e^{-\lambda}, e^{-\lambda}\lambda, e^{-\lambda}\frac{\lambda^2}{2}, \sum_{y=3}^{5} e^{-\lambda}\frac{\lambda^y}{y!}, \sum_{y=6}^{9} e^{-\lambda}\frac{\lambda^y}{y!}, \sum_{y=10}^{\infty} e^{-\lambda}\frac{\lambda^y}{y!}\right).$$

The probability mass function of $\boldsymbol{\alpha}(X) = (\alpha(X_1), \ldots, \alpha(X_N))$ is also given:

$$f(\boldsymbol{\alpha}|\lambda) = \theta_1^{\alpha_1}\theta_2^{\alpha_2}\cdots\theta_N^{\alpha_N}.$$

If there are n independent observations $\{x_i\}_{i=1}^{n}$ drawn from the Poisson(λ) distribution, the likelihood function is defined using the probability mass function of the Poisson-multinomial distribution:

$$L(\lambda) = \prod_{i=1}^{n} f(\boldsymbol{\alpha}(x_i)|\lambda). \tag{4}$$

Because this likelihood function derives from a Poisson-multinomial distribution, it is easy to show that the corresponding maximum-likelihood estimator is consistent and asymptotically normal. More importantly, the variance of its asymptotic distribution is given by the inverse of the Fisher information[1]. In other words, any consistent sequence $\hat{\lambda}_n$ of roots of the likelihood in Equation (4) satisfies $\sqrt{n}(\hat{\lambda}_n - \lambda_0) \to \mathcal{N}(0, 1/I(\lambda_0))$ in distribution, where $I(\lambda_0)$ is the Fisher information corresponding to a specific grouping scheme, and λ_0 is the underlying true parameter.

The Poisson distribution assumes that its mean λ equals variance. However, this assumption is violated if empirical frequency distributions show excess zeros relative to a Poisson distribution

---

[1]The proof follows Theorems 6.3.7 and 6.3.10 at (Lehmann and Casella 1998), and is available upon request.

(Hall 2000; Klein et al. 2015; Lambert 1992; Puig and Valero 2006). The Zero-inflated Poisson (ZIP) distribution takes excess zeros into account and has the probability mass function:

$$f(y|\lambda, p) = \begin{cases} 1 - p + pe^{-\lambda}, & \text{when } y = 0, \\ pe^{-\lambda}\frac{\lambda^y}{y!}, & \text{when } y > 0, \end{cases} \tag{5}$$

where p is the proportion of population exposed to the Poisson(λ) distribution.

Using the same distribution of α_j in (2), the grouped and right-censored data $\alpha_j(X_i)$ defined in the last section now has a different multinomial distribution $\mathcal{M}(1, \mu_1, \ldots, \mu_N)$, where

$$\mu_1(\lambda, p) = 1 - p + p\sum_{y \in I_1} e^{-y}\frac{\lambda^y}{y!}, \quad \text{and} \quad \mu_i(\lambda, p) = p\sum_{y \in I_i} e^{-\lambda}\frac{\lambda^y}{y!}, \quad \text{for } i = 2, \ldots, N.$$

The probability mass function of $\boldsymbol{\alpha}$ then depends on two parameters p and λ of the zero-inflated Poisson distribution, $f(\boldsymbol{\alpha}|\lambda, p) = \mu_1^{\alpha_1} \cdots \mu_N^{\alpha_N}$. For independent observations $\{x_i\}_{i=1}^n$, we have the likelihood function

$$L(\lambda, p) = \prod_{i=1}^n f(\boldsymbol{\alpha}(x_i)|\lambda, p). \tag{6}$$

Again, based on Theorem 6.5.1 at Lehmann and Casella (1998), it is easy to show that the maximum-likelihood estimators remain consistent and asymptotically normal for the ZIP case. The detailed proof of their asymptotic properties is given in Fu et al. (Forthcoming). Estimators $\hat{\lambda}_n$ and $\hat{p}_n$ are asymptotically efficient in the sense that in distribution,

$$\sqrt{n}(\hat{\lambda}_n - \lambda_0) \to \mathcal{N}(0, J_{22}/|\boldsymbol{J}|),$$

$$\sqrt{n}(\hat{p}_n - p_0) \to \mathcal{N}(0, J_{11}/|\boldsymbol{J}|),$$

where J_{11} and J_{22} are the 1-1 and 2-2 entry of the Fisher information matrix $\boldsymbol{J}$ corresponding to a specific grouping scheme, and λ_0 and p_0 are the underlying true parameters. We will further illustrate and discuss the Fisher information matrix $\boldsymbol{J}$ in the following sections.

9

4 Optimal Designs for Grouped and Right Censored Count Data

The foregoing demonstration that the asymptotic distributions of the maximum-likelihood estimators of both the Poisson and the ZIP cases are characterized by the Fisher information (matrix) is important for defining Bayesian optimality of GRC grouping schemes, given the internal link between Fisher information (matrix) and grouping choices. For example, the Fisher information of the Poisson case depends on both the true unknown parameter λ_0 and the specific grouping scheme $\mathcal{G}$: if we know the true parameter λ_0, the corresponding asymptotic distribution of the estimator $\hat{\lambda}_n$ is entirely determined by grouping choices in the sense that a *better* grouping scheme is associated with a smaller variance of the asymptotic distribution, or a more efficient estimator. Although the search for an optimal grouping scheme is easier if the true parameter λ_0 is known or given, one task of optimal experimental designs is to take uncertainty of unknown parameters into account by incorporating prior knowledge (from experts, prior research or pilot studies) into a general search algorithm. Next, we further investigate the relations between the Fisher information (matrix) and grouping choices. An objective function is then proposed to synthesize these relations and facilitate our subsequent discussion of a general three-step search algorithm.

4.1 Fisher Information and Grouping Choices: the Poisson Case

For the Poisson case, the Fisher information of the previous Poisson-multinomial distribution with parameter λ and the grouping scheme $\mathcal{G} = \{I_j\}_{j=1}^N$ is

$$I(\lambda) = I_{\mathcal{G}}(\lambda) = -\mathbb{E}\left[\frac{d^2}{d\lambda^2}\log f(\boldsymbol{\alpha}|\lambda)\right] = -\sum_{j=1}^N \theta_j(\lambda)\frac{d^2}{d\lambda^2}\log\theta_j(\lambda) = \sum_{i=1}^N \frac{(\theta_j')^2}{\theta_j}. \quad (7)$$

10

Equation (7) follows the definition of probability mass function θ_j in (3) and $\sum_{j=1}^N \theta_j = 1$. We next remark on the relationship between the Fisher information and grouping choices.

Remark 4.1 *When $N = 1$, we have $\theta_1 \equiv 1$ and thus $I = 0$. This corresponds to a trivial case where data provides no information for optimal designs. When $N \geq 2$, it is easy to see $I > 0$ and the search for an optimal grouping scheme becomes possible.*

Remark 4.2 *While in empirical applications we restrict N to be finite, we can also let $N = \infty$ and make each group contain only one integer. This scenario is exactly the same as precise enumeration without any grouped counts. Under this circumstance, Equation (7) shows that the Fisher information I is $1/\lambda$, which corresponds to the asymptotic variance of the Poisson estimator.*

Remark 4.3 *If we obtain a finer grouping scheme $\mathcal{G}'$ by dividing one or more groups of $\mathcal{G}$ into subgroups, such a grouping scheme yields a larger Fisher information. To show this, let $\theta = \theta(\lambda) = \sum_{k=a+1}^c e^{-\lambda}\lambda^k/k!$ be the probability corresponding to a particular group $\{a+1,\ldots,c\}$ with $a \geq -1$ and $a+1 < c$. For a grouping scheme $\mathcal{G}$, we see from (7) that this particular group contributes $(\theta')^2/\theta$ to the overall Fisher information. Now we divide this group into two subgroups $\{a+1,\ldots,b\}$ and $\{b+1,\ldots,c\}$ with $a+1 \leq b$ and $b+1 \leq c$. For the new finer grouping scheme, the contribution of these two subgroups to the Fisher information is*

$$\frac{(\theta'_*)^2}{\theta_*} + \frac{(\theta'_{**})^2}{\theta_{**}},$$

where $\theta_* = \sum_{k=a+1}^b e^{-\lambda}\lambda^k/k!$ and $\theta_{**} = \sum_{k=b+1}^c e^{-\lambda}\lambda^k/k!$. Here we note that $\theta = \theta_* + \theta_{**}$ and $U^2 v(u+v) + V^2 u(u+v) \geq uv(U+V)^2$ for $u, v, U, V > 0$. Substituting $U = \theta'_*$, $u = \theta_*$, $V = \theta'_{**}$, and $v = \theta_{**}$, we have

$$\frac{(\theta'_*)^2}{\theta_*} + \frac{(\theta'_{**})^2}{\theta_{**}} \geq \frac{(\theta')^2}{\theta}. \tag{8}$$

In Inequality (8), we note that the equality holds if and only if $\theta'_*\theta_{**} = \theta'_{**}\theta_*$ (i.e., $Uv = uV$).

Next we further demonstrate that $\theta'_*\theta_{**} \neq \theta'_{**}\theta_*$ and the equality in (8) does not hold. If $-1 < a < b < c < \infty$, we have

$$e^{2\lambda}(\theta'_{**}\theta_* - \theta'_*\theta_{**}) = \sum_{k=a+1}^{c} \frac{\lambda^b \lambda^k}{b!k!} - \sum_{k=b+1}^{c} \frac{\lambda^a \lambda^k}{a!k!} - \sum_{k=a+1}^{b} \frac{\lambda^c \lambda^k}{c!k!}$$

$$= \sum_{k=1}^{c-b} \left(\frac{1}{b!(a+k)!} - \frac{1}{a!(b+k)!} \right) \lambda^{a+b+k} + \sum_{k=1}^{b-a-1} \left(\frac{1}{b!(c-k)!} - \frac{1}{c!(b-k)!} \right) \lambda^{c+b-k}.$$

Likewise, for special cases where $a = -1$ or $c = \infty$ we also have

$$e^{2\lambda}(\theta'_{**}\theta_* - \theta'_*\theta_{**}) = \begin{cases} \sum_{k=0}^{c} \frac{\lambda^b \lambda^k}{b!k!} - \sum_{k=0}^{b} \frac{\lambda^c \lambda^k}{c!k!}, & \text{if } -1 = a < b < c < \infty, \\ \sum_{k=a+1}^{\infty} \frac{\lambda^b \lambda^k}{b!k!} - \sum_{k=b+1}^{\infty} \frac{\lambda^a \lambda^k}{a!k!}, & \text{if } -1 < a < b < c = \infty, \\ \sum_{k=0}^{\infty} \frac{\lambda^b \lambda^k}{b!k!}, & \text{if } -1 = a < b < c = \infty. \end{cases}$$

Given that $a < b < c$, the coefficients of the polynomial $e^{2\lambda}(\theta'_*\theta_{**} - \theta'_{**}\theta_*)$ across all cases discussed above (i.e., $-1 < a < b < c < \infty$, $a = -1$ or $c = \infty$) must be positive. Since λ is also positive, $\theta'_*\theta_{**} - \theta'_{**}\theta_*$ cannot be zero and we have

$$\frac{(\theta'_*)^2}{\theta_*} + \frac{(\theta'_{**})^2}{\theta_{**}} > \frac{(\theta')^2}{\theta}.$$

We previously noted that the Fisher information of the Poisson case depends on both the specific grouping scheme $\mathcal{G}$ and the true (unknown) parameter λ. Given the foregoing three remarks on the relationship between grouping choices and Fisher information, a probability function ρ can be defined to take prior knowledge of λ into account. In general, we define an *objective* function as

$$\Omega_{\mathsf{P}}(\mathcal{G}) = \int_0^\infty I_{\mathcal{G}}(\lambda) d\rho(\lambda),$$

where ρ is a continuous or discrete distribution. The introduction of ρ allows analysts to deal with the uncertainty in estimating the true parameter λ and explore optimal grouping schemes under

different prior distributions. For example, if a survey investigator assumes that the true value of λ is known, ρ becomes a degenerate distribution with a point mass of 1 at $\lambda_0 \in (0, \infty)$. For the continuous distribution case, we could specify a uniform distribution on $[a, b]$ for ρ and obtain

$$\Omega_{\mathsf{P}}(\mathcal{G}) = \frac{1}{b-a} \int_a^b I_{\mathcal{G}}(\lambda) d\lambda.$$

ρ can also be specified as a discrete distribution supported on positive numbers $\lambda_1, \ldots, \lambda_n$ with probability masses $q_1, \ldots, q_n$, respectively, and we have

$$\Omega_{\mathsf{P}}(\mathcal{G}) = \sum_{j=1}^n I_{\mathcal{G}}(\lambda_j) q_j, \quad \sum_{j=1}^n q_j = 1.$$

4.2   Fisher Information and Grouping Choices: The Zero-Inflated Poisson Case

Given that the ZIP distribution has two parameters p and λ, its corresponding Fisher information is denoted by a symmetric and positive semi-definite matrix:

$$\boldsymbol{J}(\lambda, p) = \boldsymbol{J}_{\mathcal{G}}(\lambda, p) = \begin{bmatrix} J_{11} & J_{12} \\ J_{21} & J_{22} \end{bmatrix} = \sum_j \frac{1}{\mu_j} \begin{bmatrix} \left(\frac{\partial \mu_j}{\partial \lambda}\right)^2 & \frac{\partial \mu_j}{\partial \lambda} \frac{\partial \mu_j}{\partial p} \\ \frac{\partial \mu_j}{\partial \lambda} \frac{\partial \mu_j}{\partial p} & \left(\frac{\partial \mu_j}{\partial p}\right)^2 \end{bmatrix},$$

where $\mu_1 = 1 - p + p\theta_1$, $\mu_j = p\theta_j$ for $j \geq 2$, and $\theta_j = \sum_{y \in I_j} e^{-\lambda} \lambda^y / y!$ for $j = 1, \ldots, N$. This matrix could also be expressed in the form of Equation (7) as

$$\boldsymbol{J} = \begin{bmatrix} \frac{p(p-1)(\theta_1')^2}{\theta_1 \mu_1} + pI(\lambda) & -\frac{\theta_1'}{\mu_1} \\ -\frac{\theta_1'}{\mu_1} & \frac{1-\theta_1}{p\mu_1} \end{bmatrix}, \tag{9}$$

where $I(\lambda) = \sum_{j=1}^N (\theta_j')^2 / \theta_j$. Next, we remark on the relationship between Fisher information and grouping choices for the ZIP case.

Remark 4.4 *Optimal designs become impossible in the trivial case when $N = 1$. This scenario is similar to Remark 4.1 as the Fisher information matrix $\boldsymbol{J}$ becomes a zero matrix. Another trivial case appears when $N = 2$ and the determinant of $\boldsymbol{J}$ is zero (note that $J_{11}(\lambda, p) = p(\theta_1')^2/(\mu_1(1 - \theta_1))$). Since the asymptotic distribution of $\sqrt{n}(\hat{\lambda}_n - \lambda_0, \hat{p}_n - p_0)$ is now a degenerate distribution, optimal designs based on prior knowledge of **both** λ and p become impossible. When $N \geq 3$, Remark 4.3 implies that $I(\lambda) > (\theta_1')^2/\theta_1 + ((1 - \theta_1)')^2/(1 - \theta_1)$ and the determinant of $\boldsymbol{J}$ is calculated as:*

$$\det(\boldsymbol{J}) = \frac{1 - \theta_1}{\mu_1}I(\lambda) + \frac{(p - 1)(\theta_1')^2(1 - \theta_1)}{\theta_1\mu_1^2} - \frac{(\theta_1')^2}{\mu_1^2}$$

$$= \frac{1 - \theta_1}{\mu_1}I(\lambda) - \frac{(\theta_1')^2}{\theta_1\mu_1} = \frac{1}{\mu_1}\left((1 - \theta_1)I(\lambda) - \frac{(\theta_1')^2}{\theta_1}\right) > 0.$$

The Fisher information matrix $\boldsymbol{J}$ is therefore strictly positive definite when $N \geq 3$ and thus can be used for optimal designs.

For both the Poisson and the ZIP cases, we have demonstrated that asymptotic variances are given by the inverse of Fisher information (matrix). Given that in experimental designs an optimal design is often selected to yield the most efficient estimator (see e.g., Steinberg and Hunter 1984), an optimal grouping scheme of GRC data should, according to the same principle, *maximize* Fisher information (matrices) and produce more (asymptotically) efficient estimators. Because there are multiple ways of ordering square matrices, we introduce a local objective function S to compare Fisher information matrices: optimising S will give a *locally* optimal design (Chernoff 1953), where *local* means that the design is optimal for a specific value of an unknown parameter (or vector). To illustrate the definition of S, we follow previous research on the Loewner partial order (see e.g., Horn and Johnson 2013) and write $\boldsymbol{J}_* \succeq \boldsymbol{J}$ if $\boldsymbol{J}_* - \boldsymbol{J}$ is positively semi-definite, where $\boldsymbol{J}_*$ and $\boldsymbol{J}$ are both strictly positive definite matrices.

Definition 4.5 (Objective function) *We define a **local objective function** of positive definite matrices (e.g., the Fisher information matrices) as any function S satisfying $S(\boldsymbol{J}_*) \geq S(\boldsymbol{J})$ if $\boldsymbol{J}_* \succeq \boldsymbol{J}$.*

To *maximize* the Fisher information matrix and achieve more efficient estimation, we apply local objective functions based on three common optimality criteria (Horn and Johnson 2013; Steinberg and Hunter 1984): **A-optimality**: maximizing $S_\mathsf{A} = 1/\mathrm{tr}(\boldsymbol{J}^{-1})$, where $\mathrm{tr}(\boldsymbol{J}^{-1})$ is the trace of $\boldsymbol{J}^{-1}$; **D-optimality**: maximizing $S_\mathsf{D} = \det(\boldsymbol{J})$; and **E-optimality**: maximizing S_E, where S_E is the minimum eigenvalue of $\boldsymbol{J}$. If we assume that the two eigenvalues of $\boldsymbol{J}$ are e_1 and e_2, the A-optimality, D-optimality and E-optimality designs maximize $e_1 e_2/(e_1 + e_2)$, $e_1 e_2$ and $\min(e_1, e_2)$, respectively (Nguyen and Miller 1992). Note that S_A, S_D, and S_E satisfy the definition of local objective functions above (see e.g., Horn and Johnson 2013, page 495). Among the three optimality criteria, A-optimality minimizes average asymptotic variances of all parameter estimates; D-optimality minimizes the generalized asymptotic variance (or the volume of the confidence ellipsoid under normality) of parameter estimates; and E-optimality minimizes the maximum asymptotic variance of the estimates of (components of) parameters. Because all three optimality criteria as information functions are isotonic with respect to the Loewner ordering (Pukelsheim 1993), results from simulations (not shown) suggest that optimal grouping schemes generated by the three methods are virtually the same. Nevertheless, we recommend the use of D-optimality due to its calculation simplicity. A- and E-optimality may lead to more accurate results if there is a strong correlation between p and λ (Steinberg and Hunter 1984).

Remark 4.6 *For the Poisson case, Remark 4.3 shows that a finer grouping scheme gives a larger Fisher information. This conclusion does not always hold for the ZIP case. For example, the local objective function $S_2(\boldsymbol{J}) := J_{22} = \frac{1-\theta_1}{p\mu_1}$ depends entirely on how the first group of a grouping*

scheme is defined. S_2 *remains unchanged if one divides a group other than the first group into more subgroups. Yet, the conclusion* $S(\boldsymbol{J}_{\mathcal{G}_*}) \geq S(\boldsymbol{J}_{\mathcal{G}})$ *still holds for the ZIP case if a grouping scheme* $\mathcal{G}_*$ *is finer than* $\mathcal{G}$.

This conclusion that $S(\boldsymbol{J}_{\mathcal{G}_*}) \geq S(\boldsymbol{J}_{\mathcal{G}})$ *becomes obvious once the difference* $\Delta \boldsymbol{J} = \boldsymbol{J}_{\mathcal{G}_*} - \boldsymbol{J}_{\mathcal{G}}$ *is shown to be positive semi-definite. To investigate whether* $\Delta \boldsymbol{J}$ *is positive semi-definite, we assume without loss of generality that* $\mathcal{G}_*$ *is obtained by dividing one group from* $\mathcal{G}$ *into two and denote*

$$\Delta \boldsymbol{J} = \begin{bmatrix} \Delta J_{11} & \Delta J_{12} \\ \Delta J_{21} & \Delta J_{22} \end{bmatrix}.$$

Because both $\boldsymbol{J}_{\mathcal{G}_*}$ *and* $\boldsymbol{J}_{\mathcal{G}}$ *are symmetric matrices, we have* $\Delta J_{12} = \Delta J_{21}$. *If* $\mathcal{G}_*$ *is obtained by dividing the* j^{th} *(*$j \geq 2$*) group of* $\mathcal{G}$ *into two subgroups, we have* $\Delta J_{12} = \Delta J_{21} = \Delta J_{22} = 0$. *This conclusion follows Equation (9) because the choice of the first group, which remains the same for both* $\mathcal{G}_*$ *and* $\mathcal{G}$, *determines* ΔJ_{12}, ΔJ_{21} *and* ΔJ_{22}. *Remark 4.3 also indicates that* $\Delta J_{11} > 0$. *Therefore* $\Delta \boldsymbol{J}$ *has two nonnegative eigenvalues,* 0 *and* ΔJ_{11}, *and is positive semi-definite.*

If $\mathcal{G}_*$ *is obtained by dividing the first group of* $\mathcal{G}$ *into two subgroups, we have* $\theta_j = \sum_{y \in I_j} e^{-y} \lambda^y / y!$ *with* $I_1, \cdots, I_N$ *denoting different groups of* $\mathcal{G}_*$. *Therefore* $I_1 \cup I_2$ *is the first group of* $\mathcal{G}$, I_3 *is the second group of* $\mathcal{G}$, *and so on. Following the definition of the Poisson-multinomial distribution, we still have* $\mu_1 = 1 - p + p\theta_1$ *and* $\mu_j = p\theta_j$ *for* $j \geq 2$. $\Delta \boldsymbol{J}$ *is then calculated as follows.*

$$\Delta J_{11} = \frac{(p\theta_1')^2}{\mu_1} + \frac{(p\theta_2')^2}{\mu_2} - \frac{(p\theta_1' + p\theta_2')^2}{\mu_1 + \mu_2} \geq 0,$$

$$\Delta J_{12} = \Delta J_{21} = -\frac{\theta_1'}{\mu_1} + \frac{\theta_1' + \theta_2'}{\mu_1 + \mu_2},$$

$$\Delta J_{22} = \frac{1 - \theta_1}{p\mu_1} - \frac{1 - \theta_1 - \theta_2}{p(\mu_1 + \mu_2)} = \frac{\theta_2}{p\mu_1(\mu_1 + \mu_2)} > 0.$$

Since $\mathrm{tr}(\Delta \boldsymbol{J}) = \Delta J_{11} + \Delta J_{22} > 0$, *the sum of the two eigenvalues of* $\Delta \boldsymbol{J}$ *is positive. Meanwhile,* $\Delta \boldsymbol{J}$ *has no negative eigenvalues because* $\det(\Delta \boldsymbol{J}) = 0$ *(proof omitted). Hence we conclude that*

$\Delta \boldsymbol{J}$ *is positive semi-definite and* $S(\boldsymbol{J}_{\mathcal{G}_*}) \geq S(\boldsymbol{J}_{\mathcal{G}})$*, if a grouping scheme* $\mathcal{G}_*$ *is finer than* $\mathcal{G}$*.*

We use a distribution $\rho(\lambda, p)$ to model prior knowledge of the parameters λ and p. Let S be a local objective function. We define another *global* objective function as

$$\Omega_{\mathsf{ZIP}}(\mathcal{G}) = \Omega_{\mathsf{ZIP},S}(\mathcal{G}) = \int_{\mathbb{R}^+ \times (0,1)} S(\boldsymbol{J}_{\mathcal{G}}(\lambda, p)) d\rho(\lambda, p).$$

Here, we choose to optimize the integral of $S(\boldsymbol{J})$ because this method has been justified by Chaloner and Verdinelli (1995) and is shown to be a preferred option for defining Bayesian D-optimality (Atkinson et al. 2007). For example, if $\rho(\lambda, p)$ is a uniform distribution on $(a, b) \times (c, d)$ we have

$$\Omega_{\mathsf{ZIP}}(\mathcal{G}) = \frac{1}{(b-a)(d-c)} \int_a^b d\lambda \int_c^d S\left(\boldsymbol{J}_{\mathcal{G}}(\lambda, p)\right) dp.$$

When $\rho(\lambda, p)$ is a discrete distribution supported on $\{(\lambda_j, p_j)\}_{j=1}^n$ with probabilities $q_1, \ldots, q_n$, respectively, we have

$$\Omega_{\mathsf{ZIP}}(\mathcal{G}) = \sum_{j=1}^n S\left(\boldsymbol{J}_{\mathcal{G}}(\lambda_j, p_j)\right) q_j, \quad \sum_{j=1}^n q_j = 1.$$

5   A Three-Step M Algorithm

Considering a global objective function $\Omega(\cdot)$ which is either Ω_{P} or Ω_{ZIP} in the preceding section, we propose a three-step M search algorithm for selecting an optimal grouping scheme that maximizes Ω. It should be noted that the application of this algorithm is not restricted to GRC data but could be extended to optimal designs for count responses in general if either grouping or censoring is present. From the perspective of semi-supervised machine learning, this M algorithm searches all possible combinations of grouping schemes and interacts with survey investigators to yield the optimal grouping scheme.

Remarks 4.3 and 4.6 show that a finer grouping scheme increases the value of Ω. Without grouping or right censoring, Ω is thus maximized by the finest scheme where each separate response group contains and only contains one integer. This finest possible grouping scheme is obviously the optimal one. In the presence of grouping and right censoring, however, the search for an optimal grouping scheme is constrained by the total number of groups N allowed. Now the search becomes challenging, if not impossible, since the search algorithm has to deal with infinitely many grouping schemes. To make sure that the infinitely many grouping schemes for the GRC responses can be processed by our search algorithm, we introduce a hypothetical integer M, which is sufficiently large, to divide the infinitely many grouping schemes into two parts: a finite set where M is contained in the last groups of schemes and an infinite set where M is not contained in the last groups. With the introduction of M, the search algorithm consists of three major steps. First, we use M to produce a finite set of possible grouping schemes. An optimal grouping scheme maximizing $\Omega(\cdot)$ is identified after a search of this finite set. The second and third steps verify whether the optimal grouping scheme returned by the first step is the global maximizer, i.e., the scheme achieving the best performance among all N-group schemes. The search algorithm stops if the optimal grouping scheme returned by the first step passes the verification. Otherwise, the iteration continues with a larger M.

Step 1: Select a sufficiently large positive integer M. Among all N-group grouping schemes where M is contained in their last groups, find the scheme $\mathcal{G}_{\mathsf{max}}$ that maximizes Ω.

The introduction of M divides the whole set of infinite grouping schemes into two parts: a finite set with M contained in the last group and an infinite set with M *not* contained in the last group. This procedure is motivated by the idea that, if M is sufficiently large, all integers larger than M from a Poisson process cannot exert much influence on the Fisher Information and thus do

not affect the choice of optimal grouping schemes. To illustrate this idea, we define the last right-censored group I_N of a N-group scheme $\mathcal{G} = \{I_i\}_{i=1}^N$ as $I_N = \{M, M+1, \ldots\}$. For a Poisson(λ) model we see that the contribution of this particular group containing M and larger integers to the Fisher information is trivial:

$$\frac{\left(e^{-\lambda} \frac{\lambda^{M-1}}{(M-1)!}\right)^2}{\sum_{k=M}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!}} \to 0, \text{ as } M \to \infty.$$

Moreover, an implication of this property is that, to increase the Fisher information, finer grouping decisions should be applied to integers with nontrivial probabilities. If the total number of groups N is fixed, a finer grouping of large integers with trivial probabilities should be avoided and a *coarse* right-censored group is preferred.

The choice of M follows a Goldilocks rule. An important assumption of the search algorithm is that M should be sufficiently large and represents the lower bound of a set of integers leading to a successful search for the global optimal scheme. Yet, researchers should not choose a too large M either: the number of all possible grouping schemes processed by the search algorithm grows quickly with larger M, and the search takes much longer time despite optimization of the algorithm (the computation time is roughly proportional to M^{N-1}). In theory M should be the lowest integer included in the last right-censored group of the global optimal scheme so that the search algorithm works without consuming too much time. As a practical guidance, researchers may start from an integer larger than the mean of the prior distribution of λ, gradually increase M if its previous value fails the verification from Step 2 and 3, and locate a sufficiently large M in a trial-and-error learning process.

One example of the search algorithm is demonstrated in Figure 1. If we set $N = 3$ and $M = 4$, Step 1 only searches six schemes as plotted in the part A of Figure 1. When M is not contained in

the last group, there are infinitely many grouping schemes to search and their overall set is denoted as $\mathcal{F}_1$. Examples of $\mathcal{F}_1$ are plotted in the part B of Figure 1.

Step 2: Compute the objective function Ω^* of $\mathcal{F}_3$ (defined below), where $\Omega^* = \max_{\mathcal{G} \in \mathcal{F}_3} \Omega(\mathcal{G})$.

The foregoing discussion in Step 1 shows that our algorithm divides the set of all possible grouping schemes into two parts and Step 1 deals with the finite set with M contained in the last group. Step 2 then deals with the other infinite set $\mathcal{F}_1$ with M *not* contained in the last group. In Step 2 the algorithm will search a finite set $\mathcal{F}_3$ of grouping schemes and calculate the objective function based on an optimal scheme from $\mathcal{F}_3$. To understand the second step, we first illustrate what $\mathcal{F}_2$ and $\mathcal{F}_3$ are and then discuss the relation between $\mathcal{F}_1$ and $\mathcal{F}_3$. First, let $\mathcal{F}_2$ be the overall set of $(N-1)$-group schemes such that M is contained in the last group. When $N = 3$ and $M = 4$, $\mathcal{F}_2$ only consists of 4 schemes and is illustrated in the part C in Figure 1. Second, for each grouping scheme $\mathcal{G}$ in $\mathcal{F}_2$, we divide its *tail* after M to make a new scheme $\mathcal{G}'$. Now the first $N-2$ groups in $\mathcal{G}'$ are exactly the same as these in the corresponding $\mathcal{G}$, but each integer greater than M is now contained and only contained in a separate group in $\mathcal{G}'$. We denote $\mathcal{F}_3$ as the total set of all grouping schemes $\mathcal{G}'$ obtained in this way from $\mathcal{F}_2$. The case with $N = 3$ and $M = 4$ is shown in parts C (for $\mathcal{F}_2$) and D (for $\mathcal{F}_3$) in Figure 1. Due to the one-to-one match between grouping schemes from $\mathcal{F}_2$ and $\mathcal{F}_3$, they have the same number of schemes.

For any N-group scheme $\mathcal{G}$ from $\mathcal{F}_1$ where M is not contained in the last group, $\mathcal{F}_3$ contains at least one scheme $\mathcal{G}'$ finer than $\mathcal{G}$. More specifically, if M is contained in the $(N-1)^{th}$ group for a grouping scheme $\mathcal{G}$, there exists some $\mathcal{G}'$ in $\mathcal{F}_3$ that has identical first $N-2$ groups as $\mathcal{G}$. $\mathcal{G}'$ must be finer than $\mathcal{G}$ given that every integer beyond M is also contained in a separate group in $\mathcal{G}'$. The case where M is contained in the kth group with $k \leq N-2$ can be deduced by analogy, as now the first $N-2$ groups of $\mathcal{G}'$ are finer than the first $k-1$ groups of $\mathcal{G}$.

Figure 1: An illustration of the search algorithm.
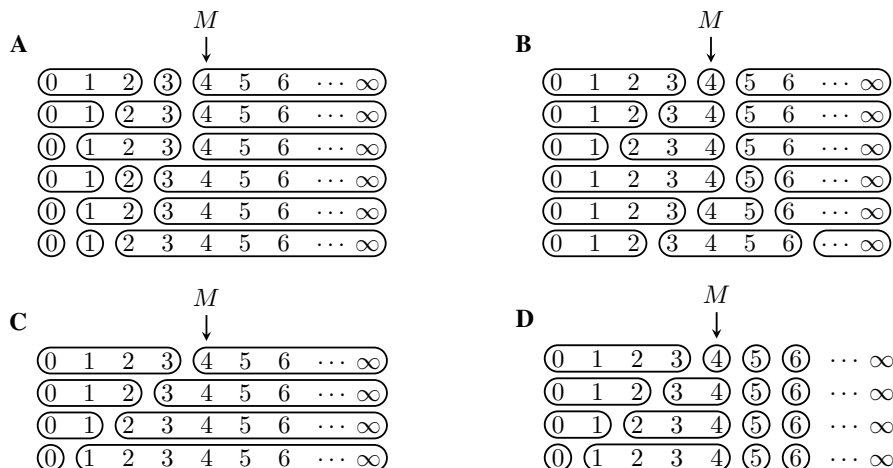
Part A: all possible 3-group schemes with $M = 4$, where M is contained in their last groups; Part B: examples of infinitely many 3-group schemes from the set $\mathcal{F}_1$, where M is not contained in their last groups; Part C: the set $\mathcal{F}_2$ of 2-group schemes obtained from a merging process of schemes in Part B. M is still contained in their last groups; Part D: the set $\mathcal{F}_3$ obtained from $\mathcal{F}_2$ by including each integer greater than M in one and only in one group.

Step 3: If $\Omega^* \leq \Omega(\mathcal{G}_{\mathsf{max}})$, $\mathcal{G}_{\mathsf{max}}$ is the global maximizer of the objective function. A larger M should be chosen otherwise.

The relationship among $\mathcal{F}_1$, $\mathcal{F}_2$ and $\mathcal{F}_3$ can be further conceptualized as follows. When M is not contained in the last group (as shown in $\mathcal{F}_1$), the search algorithm actually merges the group containing M with all its right-side groups (including the right-censored group) to form a new and bigger right-censored group. The first three grouping schemes from part B to part C in Figure 1 illustrate this merging process. Subsequently, the new grouping scheme with a bigger right-censored group (e.g., the first scheme in part C) has fewer total number of groups and is thus coarser than its original form in $\mathcal{F}_1$ (e.g., the first scheme in part B). To make a fair comparison between Fisher information of grouping schemes with M contained in the last group and with M

21

not contained in the last group, after the merging we must compensate for the loss in the latter's Fisher information due to this reduction in the total number of groups. To compensate for the loss of the Fisher information after merging, each integer greater than M in the new last group is subsequently contained and only contained in a separate group. This procedure thus forms a new (much) finer grouping scheme (i.e., from grouping schemes in part C to corresponding grouping schemes in part D). Step 3 then compares values of the objective functions between the optimal grouping scheme with M contained in the last group and the (much finer) optimal grouping scheme with M contained in other groups. Pseudo-code describing the search algorithm is listed as below to facilitate readers' understanding:

1. Input the (maximum) number of groups N, a sufficiently large integer M and the objective function of a grouping scheme Ω;

2. Among all N-group grouping schemes where M is contained in their last groups, find the scheme $\mathcal{G}_{\mathsf{max}}$ that maximizes Ω and denote this maximum value as $\Omega(\mathcal{G}_{\mathsf{max}})$;

3. Set $\mathcal{F}_2$ as the overall set of $(N-1)$-group schemes where M is contained in the last group for every grouping scheme $\mathcal{G}$ in $\mathcal{F}_2$;

4. For every $\mathcal{G}$ in $\mathcal{F}_2$, there exists one corresponding grouping scheme $\mathcal{G}'$ where every integer greater than M is contained and only contained in one separate group. The total set of all such grouping schemes $\mathcal{G}'$ is defined as $\mathcal{F}_3$. Find the grouping scheme in $\mathcal{F}_3$ that maximizes Ω and denote this maximum value as Ω^*;

5. Return $\mathcal{G}_{\mathsf{max}}$ if $\Omega^* \leq \Omega(\mathcal{G}_{\mathsf{max}})$; else choose a larger M and proceed to the first step.

To summarize, because there are infinitely many grouping schemes with M *not* contained in

the last group (e.g., grouping schemes from the set $\mathcal{F}_1$ as shown in part B), we first transform them into finite schemes with M contained in a new (big) last group ($\mathcal{F}_2$ as shown in part C) and then much finer schemes ($\mathcal{F}_3$ as shown in part D) to do a fair comparison. It is clear that these much finer grouping schemes may sometimes overcompensate for loss in Fisher information in the merging process. For example, Step 3 could falsely reject the true optimal grouping scheme if the M chosen is at or slightly higher than the lowest integer included in the last right-censored group of the global optimal scheme. Yet, the false rejection can be easily solved by increasing the value of M, as each separate group containing one integer larger than M plays less role in estimating the objective function. The global optimal grouping scheme successfully accepted by the algorithm remains the same as that falsely rejected. Actually, the search algorithm is intentionally developed in a way that it prevents any false acceptance of a wrong optimal scheme at the cost of tolerating false rejection of the true optimal grouping scheme, while a larger M further solves the false-rejection issue.

6   Data Simulation and Empirical Analysis

To illustrate the optimal designs for count data, we employ data from a nationally representative survey of youth in America, the Monitoring the Future (MTF) study. Since 1975, each year about 250,000 high-school students from approximately 130 U.S. high schools nationwide participate in this survey. In the current study, we focus on four questions from the MTF study related to 12th graders' frequencies of alcohol drinking from 1996 to 2012. The first three questions on alcohol drinking are virtually the same except for the reference period (in your lifetime, during the last 12 months and during the last 30 days): "On how many occasions have you had alcoholic beverages to drink–more than just a few sips?". The GRC response categories for the three questions are: 0

Occasions, 1-2 Occasions, 3-5 Occasions 6-9 Occasions, 10-19 Occasions, 20-39 Occasions and 40 or More. The fourth question is related to binge drinking: "Think back over the LAST TWO WEEKS. How many times have you had five or more drinks in a row? (A 'drink' is a glass of wine, a bottle of beer, a wine cooler, a shot glass of liquor, a mixed drink, etc.)". GRC response categories for this question are none, once, twice, 3 to 5 times, 6 to 9 times, 10 or more times. Table 1 shows the original counts of drinking data from 1996 to 2012. Drinking behaviors tend to be less often with shorter reference periods. Binge drinking is most rare among the 12$^{\text{th}}$ graders.

To identify appropriate prior distributions for Bayesian optimal designs, we wrote an R function `grcmle` to infer Poisson parameters based on the likelihood functions given in (4) and (6). This R function adopts maximum likelihood estimation and reports the mean, standard error and confidence interval estimated from data. As multiple waves of data are available from the MTF study, we use the mean $\pm$ 3 standard deviations as the range of the prior distribution. In the absence of multiple datasets, researchers could also determine the range of prior distributions based on the mean and standard error (e.g., the mean $\pm$ 3 standard errors) reported by this R function. Table 2 summarizes means and standard deviations of Poisson and zero-inflated Poisson parameters across the 17 years investigated. As expected, the means of λ across different survey years tend to be larger with a longer reference period. Also noteworthy is that all standard deviations calculated are much smaller than their corresponding means, suggesting that the year-to-year (zero-inflated) Poisson estimates are relatively stable.

Next, based on the aforementioned search algorithm, we developed another R function `find.scheme` in the R package `GRCdata` to search for the optimal GRC grouping scheme, which has the following parameters:

Table 1 Frequency distributions of adolescent alcoholic drinking, MTF, 1996-2012

	Lifetime drinking							Drinking in last 12 months						
Year	0	1-2	3-5	6-9	10-19	20-39	40+	0	1-2	3-5	6-9	10-19	20-39	40+
2012	674	199	235	208	281	224	416	795	335	322	225	231	160	163
2011	675	191	271	211	294	211	422	825	384	297	237	237	136	151
2010	671	193	265	216	281	244	460	784	397	321	231	250	137	199
2009	582	191	239	223	270	245	446	701	385	292	222	262	158	171
2008	634	160	251	210	292	222	470	759	380	288	218	241	165	187
2007	676	195	227	204	306	229	531	808	354	277	243	271	157	239
2006	626	169	233	226	304	242	497	756	349	302	249	262	178	196
2005	610	197	272	243	244	246	544	745	390	308	237	265	197	212
2004	571	187	245	210	307	254	583	690	363	314	246	295	214	218
2003	535	187	232	260	295	257	595	687	373	322	282	256	171	261
2002	465	151	228	214	267	211	564	592	318	311	214	239	171	240
2001	447	145	188	185	298	287	514	574	298	269	263	272	179	217
2000	412	157	246	196	283	248	538	533	303	318	236	277	190	209
1999	420	178	215	192	311	261	632	558	364	265	247	256	238	274
1998	442	168	295	233	295	307	724	588	396	327	256	331	213	350
1997	492	171	247	233	340	310	694	638	367	338	261	325	262	285
1996	515	155	226	206	282	314	625	642	350	298	254	297	219	250

	Drinking in last 30 days							Binge drinking in last two weeks					
Year	0	1-2	3-5	6-9	10-19	20-39	40+	0	1	2	3-5	6-9	10+
2012	1264	462	254	123	76	28	24	1665	197	150	142	29	19
2011	1346	432	234	122	74	22	34	1730	190	134	124	32	28
2010	1323	457	254	132	86	25	38	1716	220	158	129	35	27
2009	1224	438	257	138	85	24	24	1586	222	145	147	34	32
2008	1262	454	240	137	78	26	41	1630	215	160	135	44	24
2007	1293	451	275	144	107	37	45	1688	217	164	149	55	39
2006	1244	469	262	154	100	35	27	1661	211	160	145	45	32
2005	1257	459	304	157	104	33	32	1674	243	178	165	37	33
2004	1201	466	296	193	103	31	40	1635	253	171	170	61	26
2003	1193	504	247	196	123	41	37	1636	234	180	202	42	22
2002	1062	446	223	151	136	30	39	1456	200	161	166	47	25
2001	1036	430	239	173	114	40	27	1438	209	158	144	62	37
2000	1022	437	288	169	94	31	30	1425	202	186	150	40	36
1999	1041	425	285	214	147	40	53	1460	213	176	217	63	54
1998	1162	453	347	220	155	58	63	1634	228	224	228	61	57
1997	1153	527	324	207	170	46	47	1664	253	198	238	62	53
1996	1134	436	321	193	142	34	53	1572	236	191	202	55	47

```
find.scheme(N, densityFUN, lambda.lwr, lambda.upr, p.lwr, p.upr,

 probs, lambdas, ps, is.0.isolated = TRUE, model = c("Poisson",

        "ZIP"), matSc = c("A", "D", "E"), M = "auto").
```

N defines the (maximum) number of groups, which should be greater than 1 for the Poisson case
and greater than 2 for the ZIP case. `densityFUN` gives the probability density function of a prior
distribution, if needed. $[\texttt{lambda.lwr}, \texttt{lambda.upr}]$ and $[\texttt{p.lwr}, \texttt{p.upr}]$ define the range of λ and p,
respectively, and $[\texttt{p.lwr}, \texttt{p.upr}]$ is not needed if a Poisson model is chosen. `probs`, `lambdas`, and
`ps` define discrete prior distributions. `is.0.isolated` indicates whether zero should be contained in
a separate group. This parameter is included given that researchers are often interested in estimat-
ing prevalence or incidence rates, which requires zero to be contained in a separated group. `model`
specifies Poisson or zero-inflated Poisson cases to be used in the search algorithm. `matSc` gives
the type of local objective functions of the Fisher Information matrix for the ZIP case. Users can
choose from A-, D- and E-optimality. M is a sufficiently large integer required to implement the
search, as discussed above. If the lowest Ms needed to find the global optimal grouping schemes
are specified, most examples listed in Table 3.1 to 3.3 take several seconds to converge. Depend-
ing on the computer configuration, the program may take several minutes to converge if a large
M (e.g., 33) is chosen. If M is set as `auto`, the search algorithm will automatically determine
an adequate M needed to produce the global optimal grouping scheme and subsequently return
such optimal grouping scheme. As expected, it can take longer time for the program to converge
when the `auto` option is chosen. If M is not set as `auto`, the output of `find.scheme` includes
an indicator *succeed* indicating whether M chosen is sufficiently large for the search algorithm to
identify the global optimal grouping scheme. Users need to choose a slightly larger value for M if

succeed is false.

Based on results calculated by `find.scheme`, Table 3 shows optimal grouping schemes given combinations of the (maximum) number of groups N and prior distributions. The last entry under the *Lifetime drinking (ZIP)* scenario is estimated by the following command:

```
find.scheme(M=35, N=9, density=function(...)1, lambda.lwr=17.08,
        lambda.upr=25.17, p.lwr=0.63, p.upr=0.88, model="ZIP")
```

which yields the same optimal grouping scheme as

```
find.scheme(M=35, N=9, density=function(x)1, lambda.lwr=17.08,
                    lambda.upr=25.17)
```

We use a uniform distribution as the prior distribution in Table 3.1. It should be noted, however, that other continuous or discrete distributions can also be processed by the R function as the prior distribution. The third cell [0, 1, 2, 3-4, 5+] under the *Drinking in 30 days (Poisson)* scenario means that the optimal grouping scheme is zero, once, twice, three and four times, and five times and more given that the total number of groups is five and the range for λ's prior distribution is [1.34, 4.02]. Across different scenarios, the lowest M required to identify the global optimal scheme is also provided. As the search algorithm tolerates false rejection, the lowest M is often slightly larger than the lowest integer of the last right-censored group of the optimal scheme and this difference becomes larger as λ increases. Across the eight scenarios in Table 3.1, the cut-off integers between two adjacent groups tend to concentrate on smaller integers if the parameter space of λ is close to zero (the binge drinking scenario). If the parameter space of λ stays close to zero (e.g., rare events), any grouping decision of small integers is not supported by the search algorithm as the maximum number of groups N increases (see $N = 7$ or 8 in the binge drinking

27

scenario). This finding suggests that the GRC response is inappropriate for collecting very rare count events. The cut-off integers tend to appear first around the mean of the parameter space of λ, and then appear at other locations as N increases. The prior distributions in Table 3.2 are truncated (mean $\pm$ 3 standard deviations) Gaussian distributions, whose means and standard deviations are provided in Table 2. Table 3.2 also demonstrates that the cut-off integers of grouping decisions often exist at integers around which λ has higher probability density. For the same combination of N and range of prior distributions, the optimal schemes listed in Table 3.1 and Table 3.2 are virtually the same, suggesting that the search for optimal schemes is not sensitive to the choice of prior distributions. Table 3.3 lists optimal grouping schemes when λ is low, moderate, high and unspecified for readers' reference. Because zero is contained in a separate group for the ZIP case from Table 3.1 to 3.3, the optimal grouping scheme remains the same when λ is fixed but p varies.

To illustrate how an optimal grouping scheme is preferred to other grouping schemes, we use the grouping scheme adopted by the MTF binge drinking question as a reference grouping scheme, and compare standard errors estimated under different grouping schemes. In Table 4, the first column is the true parameters we used to simulate the Poisson distributions. Both parameters inferred from the data (see Table 2) and hypothetical parameters are used. The reference schemes are those adopted by the MTF study to measure alcohol drinking. The optimal schemes are generated by find.scheme with the same number of groups as the reference group. For each scenario, the simulation is repeated 1,000 times to calculate estimates.

When λ is small, the reference schemes appear to be acceptable as their corresponding standard errors are only slightly larger than these calculated based on optimal grouping schemes. However, the differences between standard errors estimated from the reference schemes and those of the optimal schemes grow larger as λ increases. As expected, the standard errors decrease by $\sqrt{10}$

Table 2 Maximum likelihood estimates of means and standard deviations (S.D.) of (zero-inflated) Poisson parameters

	Drinking in last 30 days			Drinking in last 12 months			Lifetime Drinking			Binge drinking in last two weeks		
	λ (Poisson)	P (ZIP)	λ (ZIP)	λ (Poisson)	P (ZIP)	λ (ZIP)	λ (Poisson)	P (ZIP)	λ (ZIP)	λ (Poisson)	P (ZIP)	λ (ZIP)
Mean	2.723	0.477	6.062	8.750	0.698	13.026	15.425	0.756	21.127	0.774	0.307	2.662
S.D.	0.461	0.039	0.561	1.209	0.039	1.089	1.834	0.040	1.348	0.129	0.029	0.226

Table 3.1 Estimated optimal grouping schemes with different numbers of groups: uniform prior distributions

N # of groups	Drinking in 30 days (Poisson) $\lambda=[1.34, 4.02]$	M*	Drinking in 12 months (Poisson) $\lambda=[5.12, 12.38]$	M*	Lifetime drinking (Poisson) $\lambda=[9.92, 20.93]$	M*	Binge drinking (Poisson) $\lambda=[0.39, 1.16]$	M*
3	0 1-2 3+	4	0 1-8 9+	11	0 1-14 15+	18	0 1 2+	3
4	0 1-2 3-4 5+	5	0 1-6 7-10 11+	13	0 1-11 12-17 18+	21	0 1 2 3+	3
5	0 1 2 3-4 5+	6	0 1-5 6-8 9-12 13+	14	0 1-10 11-14 15-19 20+	23	0 1 2 3 4+	4
6	0 1 2 3 4-5 6+	6	0 1-4 5-7 8-10 11-13 14+	15	0 1-9 10-13 14-17 18-22 23+	24	0 1 2 3 4 5+	5
7	0 1 2 3 4 5 6-7+	7	0 1-4 5-6 7-8 9-10 11-13 14+	16	0 1-8 9-11 12-14 15-18 19-22 23+	25	0 1 2 3 4 5 6+	6
8	0 1 2 3 4 5 6-7 8+	8	0 1-3 4-5 6-7 8-9 10-11 12-14 15+	17	0 1-7 8-10 11-13 14-16 17-19 20-23 24+	26	0 1 2 3 4 5 6 7+	7
9	0 1 2 3 4 5 6 7 8-9+	9	0 1-3 4-5 6-7 8-9 10-11 12-13 14-16 17+	18	0 1-7 8-10 11-12 13-14 15-17 18-20 21-24 25+	27	0 1 2 3 4 5 6 7 8+	8

N # of groups	Drinking in 30 days (ZIP) $\lambda=[4.38, 7.75]$ P=[0.36, 0.59]	M*	Drinking in 12 months (ZIP) $\lambda=[9.76, 16.29]$ P=[0.58, 0.81]	M*	Lifetime drinking (ZIP) $\lambda=[17.08, 25.17]$ P=[0.63, 0.88]	M*	Binge drinking (ZIP) $\lambda=[1.98, 3.34]$ P=[0.22, 0.39]	M*
3	0 1-6 7+	8	0 1-12 13+	15	0 1-20 21+	24	0 1-3 4+	4
4	0 1-4 5-7 8+	9	0 1-10 11-15 16+	18	0 1-18 19-24 25+	27	0 1-2 3-4 5+	5
5	0 1-4 5-6 7-9 10+	11	0 1-9 10-13 14-17 18+	19	0 1-16 17-21 22-26 27+	29	0 1 2 3-4 5+	6
6	0 1-3 4-5 6-7 8-10 11+	11	0 1-8 9-11 12-14 15-18 19+	20	0 1-15 16-19 20-23 24-28 29+	30	0 1 2 3 4-5 6+	6
7	0 1-2 3-4 5-6 7-8 9-10 11+	12	0 1-7 8-10 11-13 14-16 17-19 20+	22	0 1-14 15-18 19-21 22-24 25-28 29+	32	0 1 2 3 4 5-6 7+	7
8	0 1-2 3-4 5 6-7 8-9 10-11+	13	0 1-7 8-9 10-11 12-13 14-16 17-19 20+	22	0 1-13 14-16 17-19 20-22 23-25 26-29 30+	32	0 1 2 3 4 5 6-7+	8
9	0 1-2 3-4 5 6 7 8-9 10-11 12+	13	0 1-7 8-9 10-11 12-13 14-15 16-17 18-20 21+	23	0 1-12 13-15 16-18 19-20 21-23 24-26 27-30 31+	33	0 1 2 3 4 5 6 7 8+	9

Note: * The lowest M needed to find the global optimal grouping scheme.

Table 3.2 Estimated optimal grouping schemes with different numbers of groups: normal prior distributions

N # of groups	Drinking in 30 days (Poisson) $\lambda=[1.34, 4.02]$	M*	Drinking in 12 months (Poisson) $\lambda=[5.12, 12.38]$	M*	Lifetime drinking (Poisson) $\lambda=[9.92, 20.93]$	M*	Binge drinking (Poisson) $\lambda=[0.39, 1.16]$	M*
3	0 1-3 4+	4	0 1-8 9+	11	0 1-15 16+	18	0 1 2+	3
4	0 1-2 3-4 5+	5	0 1-6 7-10 11+	13	0 1-12 13-17 18+	21	0 1 2 3+	3
5	0 1 2 3-4 5+	6	0 1-6 7-9 10-12 13+	14	0 1-11 12-15 16-19 20+	22	0 1 2 3 4+	4
6	0 1 2 3 4-5 6+	6	0 1-5 6-7 8-9 10-12 13+	15	0 1-10 11-13 14-16 17-20 21+	23	0 1 2 3 4 5+	5
7	0 1 2 3 4 5 6 7+	7	0 1-4 5-6 7-8 9-10 11-13 14+	15	0 1-9 10-12 13-15 16-18 19-22 23+	24	0 1 2 3 4 5 6+	6
8	0 1 2 3 4 5 6 7+	8	0 1-4 5-6 7-8 9-10 11-12 13-15 16+	16	0 1-9 10-12 13-14 15-16 17-19 20-23 24+	25	0 1 2 3 4 5 6 7+	7
9	0 1 2 3 4 5 6 7 8+	9	0 1-3 4-5 6-7 8-9 10-11 12-13 14-15 16+	17	0 1-8 9-11 12-13 14-15 16-17 18-20 21-23 24+	26	0 1 2 3 4 5 6 7 8+	8

N # of groups	Drinking in 30 days (ZIP) $\lambda=[4.38, 7.75]$ P=[0.36, 0.59]	M*	Drinking in 12 months (ZIP) $\lambda=[9.76, 6.29]$ P=[0.58, 0.81]	M*	Lifetime drinking (ZIP) $\lambda=[17.08, 25.17]$ P=[0.63, 0.88]	M*	Binge drinking (ZIP) $\lambda=[1.98, 3.34]$ P=[0.22, 0.39]	M*
3	0 1-6 7+	8	0 1-13 14+	15	0 1-21 22+	24	0 1-3 4+	4
4	0 1-4 5-7 8+	9	0 1-10 11-15 16+	18	0 1-18 19-24 25+	27	0 1-2 3-4 5+	5
5	0 1-4 5-6 7-9 10+	10	0 1-9 10-13 14-17 18+	19	0 1-16 17-21 22-26 27+	29	0 1-2 3-4 5-6+	6
6	0 1-3 4-5 6-7 8-9 10+	11	0 1-8 9-11 12-14 15-18 19+	20	0 1-15 16-19 20-23 24-27 28+	30	0 1 2 3 4-5 6+	6
7	0 1-3 4-5 6-7 8-9 10-11+	12	0 1-8 9-11 12-13 14-16 17-19 20+	21	0 1-14 15-18 19-21 22-24 25-28 29+	31	0 1 2 3 4 5-6 7+	7
8	0 1-2 3-4 5 6-7 8-9 10-11+	12	0 1-7 8-10 11-12 13-14 15-16 17-19 20+	22	0 1-14 15-17 18-20 21-23 24-26 27-30 31+	32	0 1 2 3 4 5 6 7+	8
9	0 1-2-3 4-5 6 7-8 9-10-11 12+	12	0 1-7 8-9 10-11 12-13 14-15 16-17 18-20 21+	22	0 1-13 14-16 17-19 20-21 22-23 24-26 27-30 31+	33	0 1 2 3 4 5 6 7 8+	8

Note: * The lowest M needed to find the global optimal grouping scheme.

Table 3.3 Estimated optimal grouping schemes with different numbers of groups: uniform prior distributions

# of groups G	Low λ $\lambda=[0.01, 3]$	M*	Moderate λ $\lambda=[3, 10]$	M*	High λ (M=40) $\lambda=[10, 30]$	M*	Unspecified λ (M=40) $\lambda=[0.01, 30]$	M*
3	0 1 2+	3	0 1-5 6+	8	0 1-16 17+	24	0 1-2 3+	15
4	0 1 2-3 4+	4	0 1-4 5-8 9+	10	0 1-13 14-22 23+	27	0 1 2-5 6+	20
5	0 1 2 3-4 5+	5	0 1-3 4-6 7-9 10+	11	0 1-11 12-17 18-25 26+	30	0 1 2-4 5-11 12+	23
6	0 1 2 3 4-5 6+	6	0 1-3 4-5 6-7 8-10 11+	12	0 1-10 11-15 16-20 21-27 28+	31	0 1-2-3 4-8 9-17 18+	25
7	0 1 2 3 4 5 6+	7	0 1-2 3-4 5-6 7-8 9-11 12+	13	0 1-10 11-14 15-18 19-23 24-29 30+	33	0 1-2-3 4-7 8-13 14-22 23+	27
8	0 1 2 3 4 5 6 7+	8	0 1-2 3-4 5-6 7-8 9-10 11-12+	14	0 1-9 10-12 13-15 16-19 20-24 25-30 31+	34	0 1-2-3 4-6 7-10 11-16 17-24 25+	29
9	0 1 2 3 4 5 6 7 8+	8	0 1-2 3 4 5 6-7 8-9 10-12 13+	14	0 1-8 9-11 12-14 15-17 18-21 22-25 26-31 32+	35	0 1-2-3 4-5-7 8-11 12-17 18-25 26+	29

# of groups G	Moderate λ & low P $\lambda=[3, 10]$ P=[0.001, 0.4]	M*	Moderate λ & moderate P $\lambda=[3, 10]$ P=[0.4, 0.6]	M*	Moderate λ & high P $\lambda=[3, 10]$ P=[0.6, 1]	M*	Moderate λ & unspecified P $\lambda=[3, 10]$ P=[0.001, 1]	M*
3	0 1-5 6+	8	0 1-5 6+	8	0 1-5 6+	8	0 1-5 6+	8
4	0 1-4 5-8 9+	10	0 1-4 5-8 9+	10	0 1-4 5-8 9+	10	0 1-4 5-8 9+	10
5	0 1-3 4-6 7-9 10+	11	0 1-3 4-6 7-9 10+	11	0 1-3 4-6 7-9 10+	11	0 1-3 4-6 7-9 10+	11
6	0 1-3 4-5 6-7 8-10 11+	12	0 1-3 4-5 6-7 8-10 11+	12	0 1-3 4-5 6-7 8-10 11+	12	0 1-3 4-5 6-7 8-10 11+	12
7	0 1-2-3 4-5 6-7 8-9 11 12+	13	0 1-2-3 4-5 6-7 8-9 11 12+	13	0 1-2-3 4-5 6-7 8-9 11 12+	13	0 1-2-3 4-5 6-7 8-9 11 12+	13
8	0 1-2-3-4 5-6-7 8-9 10-12 13+	14	0 1-2-3-4 5-6-7 8-9 10-12 13+	14	0 1-2-3-4 5-6-7 8-9 10-12 13+	14	0 1-2-3-4 5-6-7 8-9 10-12 13+	14
9	0 1-2-3 4 5 6-7 8-9 10-12 13+	14	0 1-2-3 4 5 6-7 8-9 10-12 13+	14	0 1-2-3 4 5 6-7 8-9 10-12 13+	14	0 1-2-3 4 5 6-7 8-9 10-12 13+	14

Note: * The lowest M needed to find the global optimal grouping scheme.

when sample sizes increase by 10. Moreover, the strength of this algorithm can be illustrated by optimal schemes with fewer groups than corresponding reference schemes. Compared with estimation based on the reference schemes, researchers could achieve almost the same, sometimes better, efficiency of estimation by adopting optimal schemes with even smaller numbers of groups N. In other words, an optimal grouping scheme can outperform a nonoptimal one even if the latter has more groups.

7   Discussion and Conclusion

This research applies optimal experimental design, a branch of semi-supervised machine learning, to social science research and provides a novel algorithm to find the optimal grouping scheme of GRC count responses. One of the most striking features of social science research on survey methodology is the degree to which the design of response categories in survey questions has been neglected. Count responses with grouping and right censoring has long been collected by social scientists to study a variety of behaviors, status and attitudes. Yet, there has been little research on optimal designs for discrete response categories such that grouping or right-censoring decisions often rely on arbitrary choices of survey investigators. To search for optimal grouping schemes, this paper first uses Poisson-multinomial mixture models to conceptualize the data generating process of count data with grouping and right censoring, and then investigates the relationship between grouping-scheme choices and asymptotic distributions of the Poisson-multinomial models. Using different types of optimality in experimental designs (De Leon and Atkinson 1991), we investigate local objective functions of the Fisher information (matrix) and further demonstrate the possibility of optimal designs for GRC count responses: the optimal grouping scheme should maximize the

Table 4 Parameter estimates for different grouping schemes: results from simulation

Parameters used for simulation	Grouping schemes	$\bar{\lambda}_{MLE}$ (100)*	Standard Errors	$\bar{\lambda}_{MLE}$ (1,000)*	Standard Errors	$\bar{\lambda}_{MLE}$ (10,000)*	Standard Errors
λ=2.723	Reference	2.718	0.179	2.719	0.057	2.723	0.018
	[0, 1-2, 3-5, 6-9, 10-19, 20-39, 40+]						
	Optimal	2.718	0.166	2.719	0.054	2.723	0.017
	[0, 1, 2, 3, 4, 5-6, 7+]						
	Optimal**	2.719	0.176	2.720	0.057	2.723	0.018
	[0, 1-2, 3-4, 5+]						
λ=8.750	Reference	8.746	0.330	8.742	0.106	8.749	0.033
	[0, 1-2, 3-5, 6-9, 10-19, 20-39, 40+]						
	Optimal	8.743	0.298	8.742	0.096	8.749	0.030
	[0, 1-5, 6-7, 8-9, 10-11, 12-14, 15+]						
	Optimal**	8.746	0.322	8.742	0.105	8.749	0.033
	[0, 1-7, 8-11, 12+]						
λ=15.425	Reference	15.396	0.502	15.409	0.157	15.424	0.048
	[0, 1-2, 3-5, 6-9, 10-19, 20-39, 40+]						
	Optimal	15.395	0.403	15.409	0.129	15.423	0.039
	[0, 1-10, 11-13, 14-15, 16-18, 19-21, 22+]						
	Optimal**	15.403	0.436	15.413	0.142	15.424	0.043
	[0, 1-13, 14-18, 19+]						
λ=0.774	Reference	0.773	0.088	0.772	0.028	0.774	0.009
	[0, 1, 2, 3-5, 6-9, 10+]						
	Optimal	0.773	0.087	0.772	0.028	0.774	0.009
	[0, 1, 2, 3, 4, 5+]						
	Optimal**	0.775	0.090	0.773	0.029	0.774	0.009
	[0, 1, 2+]						
λ=1	Reference	0.999	0.102	0.998	0.033	1.000	0.010
	[0, 1, 2, 3-5, 6-9, 10+]						
	Optimal	0.999	0.101	0.998	0.032	1.000	0.010
	[0, 1, 2, 3, 4, 5+]						
	Optimal**	1.001	0.106	0.998	0.035	1.000	0.010
	[0, 1, 2+]						
λ=3	Reference	2.996	0.187	2.998	0.060	3.000	0.019
	[0, 1-2, 3-5, 6-9, 10-19, 20-39, 40+]						
	Optimal	2.996	0.173	2.996	0.055	3.000	0.018
	[0, 1, 2, 3, 4, 5-6, 7+]						
	Optimal**	3.000	0.183	2.998	0.059	3.000	0.019
	[0, 1-2, 3-4, 5+]						
λ=5	Reference	4.997	0.240	4.993	0.076	4.999	0.024
	[0, 1-2, 3-5, 6-9, 10-19, 20-39, 40+]						
	Optimal	4.994	0.225	4.994	0.073	4.999	0.023
	[0, 1-2, 3-4, 5, 6-7, 8-9, 10+]						
	Optimal**	4.991	0.240	4.994	0.079	5.000	0.025
	[0, 1-4, 5-7, 8+]						
λ=10	Reference	9.993	0.373	9.997	0.122	10.000	0.037
	[0, 1-2, 3-5, 6-9, 10-19, 20-39, 40+]						
	Optimal	9.986	0.322	9.994	0.105	10.000	0.032
	[0, 1-6, 7-8, 9-10, 11-12, 13-15, 16+]						
	Optimal**	9.990	0.349	9.997	0.114	10.001	0.035
	[0, 1-8, 9-12, 13+]						
λ=20	Reference	19.988	0.552	19.994	0.182	19.999	0.054
	[0, 1-2, 3-5, 6-9, 10-19, 20-39, 40+]						
	Optimal	19.966	0.467	19.985	0.146	20.000	0.045
	[0, 1-14, 15-17, 18-20, 21-23, 24-27, 28+]						
	Optimal**	19.971	0.502	19.987	0.159	20.000	0.049
	[0, 1-17, 18-23, 24+]						

Note: *Each simulation is repeated 1,000 times to calculate estimates. Numbers in the parentheses are sample sizes used for simulation. ** Optimal grouping schemes with fewer groups

global objective function of the Fisher information (matrix). We also propose a new three-step general algorithm to process infinitely many grouping schemes and identify the global optimal grouping scheme. To process all possible grouping schemes, this algorithm introduces a sufficiently large integer M, which is in theory the lowest integer contained in the right-censored group of the global optimal scheme. The introduction of M not only makes the search feasible but also tolerates false rejection of the global optimal grouping scheme. A new R package `GRCdata` is developed to implement this algorithm and help survey investigators to assess grouping schemes of count responses. The use of two R programs `grcmle` and `find.scheme` in `GRCdata` is illustrated by empirical examples of alcohol drinking. Results from data simulation show that the optimal designs yielded by this new algorithm considerably outperform existing designs: the optimal grouping scheme, even with fewer total number of groups, can lead to more efficient estimation.

The M algorithm and software programs presented in this research readily provides survey investigators a new tool for evaluating grouping and right-censoring decisions of count responses in surveys. While survey methodologists do need to take a series of factors (e.g., the coherence of response categories over time and across questions, or whether a specific count is of research interest or has substantive meaning) into account when designing response categories (Schaeffer and Dykema 2011), the new R package developed allow scholars to incorporate their prior knowledge in optimal designs of survey questions. Although this research only addresses (zero-inflated) Poisson models of count data, it should be noted that the application of the M search algorithm is not restricted to the two statistical models investigated and can be extended to other models of count data, such as negative binomial models and hurdle models. If the assumption that the Fisher information increases with a finer grouping scheme holds for other discrete or continuous data gen-

erating processes, this M algorithm can be employed for designing survey responses in general. Such potential applications of this algorithm to broader issues in survey methodology merit further attention.

References

Akers, Ronald L, Anthony J La Greca, John Cochran, and Christine Sellers. 1989. "Social learning theory and alcohol behavior among the elderly." *The Sociological Quarterly* 30:625–638.

Atkinson, Anthony, Alexander Donev, and Randall Tobias. 2007. *Optimum experimental designs, with SAS*. Oxford, UK: Oxford University Press.

Bachman, Jerald G, Lloyd D Johnston, and Patrick M O'Malley. 1990. "Explaining the recent decline in cocaine use among young adults: further evidence that perceived risks and disapproval lead to reduced drug use." *Journal of Health and Social Behavior* 31:173–184.

Bailey, Susan L, Robert L Flewelling, and J Valley Rachal. 1992. "Predicting continued use of marijuana among adolescents: The relative influence of drug-specific and social context factors." *Journal of Health and Social Behavior* 33:51–65.

Barnes, Grace M, Joseph H Hoffman, John W Welte, Michael P Farrell, and Barbara A Dintcheff. 2006. "Effects of parental monitoring and peer deviance on substance use and delinquency." *Journal of Marriage and Family* 68:1084–1104.

Basu, Bharati and Felix Famoye. 2004. "Domestic violence against women, and their economic dependence: a count data analysis." *Review of Political Economy* 16:457–472.

Blazer, Dan, Bruce Burchett, Linda K George, et al. 1991. "The association of age and depression among the elderly: an epidemiologic exploration." *Journal of gerontology* 46:M210–M215.

Bollen, Kenneth A. 1990. "Political democracy: Conceptual and measurement traps." *Studies in Comparative International Development* 25:7–24.

Bradburn, Norman M, Seymour Sudman, and Brian Wansink. 2004. *Asking questions: the definitive guide to questionnaire design–for market research, political polls, and social and health questionnaires.* San Francisco, CA: John Wiley & Sons.

Chaloner, Kathryn and Isabella Verdinelli. 1995. "Bayesian experimental design: A review." *Statistical Science* 10:273–304.

Cheibub, Jose Antonio, Adam Przeworski, Fernando Papaterra Limongi Neto, and Michael M Alvarez. 1996. "What makes democracies endure?" *Journal of democracy* 7:39–55.

Chernoff, Herman. 1953. "Locally optimal designs for estimating parameters." *The Annals of Mathematical Statistics* 24:586–602.

Cohn, David A, Zoubin Ghahramani, and Michael I Jordan. 1996. "Active learning with statistical models." *Journal of artificial intelligence research* 4:129–145.

De Leon, AC Ponce and Anthony C Atkinson. 1991. "Optimum experimental design for discriminating between two rival models in the presence of prior information." *Biometrika* 78:601–608.

Dette, Holger, Viatcheslav B Melas, and Andrey Pepelyshev. 2004. "Optimal designs for a class of nonlinear regression models." *Annals of statistics* 32:2142–2167.

Elkins, Zachary. 2000. "Gradations of democracy? Empirical tests of alternative conceptualizations." *American Journal of Political Science* 44:293–300.

Fu, Qiang, Xin Guo, and Kenneth C Land. Forthcoming. "A Poisson-Multinomial Mixture Approach to Grouped and Right-Censored Counts." *Communications in Statistics-Theory and Methods* .

Fu, Qiang, Kenneth C Land, and Vicki L Lamb. 2013. "Bullying victimization, socioeconomic

status and behavioral characteristics of 12th graders in the United States, 1989 to 2009: repetitive trends and persistent risk differentials." *Child indicators research* 6:1–21.

Fu, Qiang, Kenneth C Land, and Vicki L Lamb. 2016. "Violent Physical Bullying Victimization at School: Has There Been a Recent Increase in Exposure or Intensity? An Age-Period-Cohort Analysis in the United States, 1991 to 2012." *Child Indicators Research* 9:485–513.

Goodman, Leo A. 1987. "New methods for analyzing the intrinsic character of qualitative variables using cross-classified data." *American Journal of Sociology* 93:529–583.

Groves, Robert M, Floyd J Fowler Jr, Mick P Couper, James M Lepkowski, Eleanor Singer, and Roger Tourangeau. 2011. *Survey methodology*. Hoboken, NJ: John Wiley & Sons.

Hagan, John, Carla Shedd, and Monique R Payne. 2005. "Race, ethnicity, and youth perceptions of criminal injustice." *American sociological review* 70:381–407.

Hall, Daniel B. 2000. "Zero-inflated Poisson and binomial regression with random effects: a case study." *Biometrics* 56:1030–1039.

Horn, Roger A. and Charles R. Johnson. 2013. *Matrix analysis*. Cambridge, UK: Cambridge University Press, second edition.

Klein, Nadja, Thomas Kneib, and Stefan Lang. 2015. "Bayesian Generalized Additive Models for Location, Scale, and Shape for Zero-Inflated and Overdispersed Count Data." *Journal of the American Statistical Association* 110:405–419.

Lambert, Diane. 1992. "Zero-inflated Poisson regression, with an application to defects in manufacturing." *Technometrics* 34:1–14.

Lang, Joseph B. 2004. "Multinomial-Poisson homogeneous models for contingency tables." *Annals of Statistics* 32:340–383.

Lehmann, Erich Leo and George Casella. 1998. *Theory of point estimation*. New York, NY:

Springer-Verlag.

Lord, Dominique, Simon P Washington, and John N Ivan. 2005. "Poisson, Poisson-gamma and zero-inflated regression models of motor vehicle crashes: balancing statistical fit and theory." *Accident Analysis & Prevention* 37:35–46.

Marsden, Peter V. 2003. "Interviewer effects in measuring network size using a single name generator." *Social Networks* 25:1–16.

Minkin, Salomon. 1987. "Optimal designs for binary data." *Journal of the American Statistical Association* 82:1098–1103.

Nguyen, Nam-Ky and Alan J Miller. 1992. "A review of some exchange algorithms for constructing discrete D-optimal designs." *Computational Statistics & Data Analysis* 14:489–498.

Paik, Anthony and Kenneth Sanchagrin. 2013. "Social isolation in America: An artifact." *American Sociological Review* 78:339–360.

Puig, Pedro and Jordi Valero. 2006. "Count data distributions: some characterizations with applications." *Journal of the American Statistical Association* 101:332–340.

Pukelsheim, Friedrich. 1993. *Optimal design of experiments*. New York, NY: John Wiley & Sons.

Radloff, Lenore Sawyer. 1977. "The CES-D scale a self-report depression scale for research in the general population." *Applied psychological measurement* 1:385–401.

Reardon, Sean F and Stephen W Raudenbush. 2006. "3. A Partial Independence Item Response Model for Surveys with Filter Questions." *Sociological Methodology* 36:257–300.

Schaeffer, Nora Cate and Jennifer Dykema. 2011. "Questions for surveys current trends and future directions." *Public opinion quarterly* 75:909–961.

Schaeffer, Nora Cate and Stanley Presser. 2003. "The science of asking questions." *Annual review of sociology* 29:65–88.

Schwarz, Norbert, Hans-Juergen Hippler, Brigitte Deutsch, and Fritz Strack. 1985. "Response Categories: Effects on behavioural reports and comparative judgements." *Public Opinion Quarterly* 49:388–395.

Settles, Burr. 2010. *Active learning literature survey*. University of Wisconsin-Madison.

Smith, Herbert L and Maurice A Garnier. 1986. "Association between background and educational attainment in France." *Sociological Methods & Research* 14:317–344.

Steinberg, David M and William G Hunter. 1984. "Experimental design: review and comment." *Technometrics* 26:71–97.

Straus, Murray Arnold, Richard J Gelles, and Christine Smith. 1990. *Physical violence in American families: Risk factors and adaptations to violence in 8,145 families*. New Brunswick, NJ: Transaction Publishers.

Sudman, Seymour, Norman M Bradburn, and Norbert Schwarz. 1996. *Thinking about answers: The application of cognitive processes to survey methodology*. San Francisco, CA: Jossey-Bass.

Thoits, Peggy A and Lyndi N Hewitt. 2001. "Volunteer work and well-being." *Journal of health and social behavior* 42:115–131.

Wong, Raymond Sin-Kwok. 2010. *Association models*. Thousand Oaks, CA: Sage Publications.