Investigating why academically successful community college students leave college without a degree

Justin Ortagus* University of Florida

Benjamin T. Skinner† University of Florida

Melvin Tanner University of Florida

15 December 2020

Abstract

Despite the fact that earning a postsecondary degree can offer economic, social, and civic benefits, many students who begin at a community college leave without earning a credential—including some who have performed well academically and made substantial progress toward graduation. To better understand the factors which might contribute to early exit, we surveyed a number of former students in a large community college system. We improve the generalizability of the survey responses through multilevel regression with poststratification (MRP), which we use to reweight the responses to better represent the population in our original survey frame. We find that tuition and fees, living expenses, and no longer being eligible for financial aid are the factors that explain why the largest share of students leave college without a degree. We also find some variation in both financial and non-financial factors across subgroups, suggesting that targeted supports focused on student subpopulations may be useful in helping students persist or return to college and complete their degree.

Keywords community college • drop out • stop out • survey • multilevel regression with poststratification

* *Email:* jortagus@coe.ufl.edu

† *Email:* btskinner@coe.ufl.edu

Introduction

Community colleges in the United States enroll millions of underrepresented students, including a disproportionate share of low-income, racially minoritized, and adult students (Bailey et al., 2015). For community college students who earn a degree, the decision to go to college is associated with higher earnings in the labor market (Belfield & Bailey, 2011; Doyle & Skinner, 2016; Jepsen et al., 2014) and a host of non-financial benefits, such as increased civic participation, improved health, and a longer life expectancy (Doyle & Skinner, 2017; Trostel, 2015). Yet the majority of students who begin at a community college drop out of college and fail to accrue the benefits related to completing their degree (Snyder et al., 2018).

Due to the open-access enrollment policies and numerous curricular missions of community colleges, students enroll in community college courses with varying degrees of academic preparation and multiple objectives pertaining to their postsecondary coursework (Cohen & Brawer, 2008). Community colleges provide critical developmental education for academically underprepared students (Bailey et al., 2015; Bemmel et al., 2008; Cohen & Brawer, 2008) and mobility pathways for working or low-income students (Armstrong & Hamilton, 2013). Although most community college students who leave college without a degree struggle with their entry-level coursework, a smaller share of non-completers drop out after performing well in their introductory courses and making considerable progress toward degree completion (Ortagus, Tanner, et al., 2020; Shapiro et al., 2019).

At the national level, roughly 10 percent of students who leave college have already made significant progress toward completing their degree, and these previously successful non-completers have been found to be the most likely to graduate upon re-enrolling (Shapiro et al., 2019). In an effort to encourage previously successful students to re-enroll and complete their degree, many community colleges launch re-enrollment campaigns targeted at former students who have made substantial progress toward graduation (Schwartz, 2019). While re-enrollment campaigns can increase the probability of former students returning to college (Ortagus, Tanner, et al., 2020), the likelihood of any student completing college after stopping out for a period of time is substantially lower when compared to students who remain enrolled (Crosta, 2014; DesJardins et al., 2006).[1] For community colleges facing limited resources and low completion rates, the most effective mechanism to optimize completion numbers would be to prevent students from dropping out in the first place. Unfortunately, students who leave college are often difficult to contact, leaving institutions with little understanding of why these former students left or how to get them to re-enroll.

In this study, we partnered with five high-enrollment community colleges in the state of Florida to email and text 27,028 former students, requesting that they complete a short web-based survey to offer critical information regarding the specific factors that contributed to their premature departure. We focus on former students who were previously successful academically given that they are the most likely to complete college upon their return, and the sample criteria for this study include former students who left college within the past three years, earned a 2.0 GPA or higher, made progress toward degree completion (median of 42 credit hours accrued in the resulting sample), and had no behavioral or financial holds that would prevent their re-enrollment. By contacting these former students, this study offers insight into the rationale of the departure decision for students who were performing well in the classroom but dropped out of college anyway. Specifically, this study will address the following research questions:

Research Question 1: What are the specific factors that contribute to previously successful former students' decision to drop out of college?

Research Question 2: To what extent do results vary according to former students' academic and demographic characteristics?

Among the former students we contacted, 2,418 responded, representing a response rate of 8.9%. To improve the generalizability of our results, we analyze the responses using multilevel regression with poststratification (MRP), a methodological tool from political science often used to produce representative estimates of public opinion from non-representative survey data (Gelman

[1] Students who return to college after exiting early are described as stopping out, whereas former students who do not return to college are described as dropping out. Our sample includes both students who stop out and eventually return as well as those who drop out and have not returned to college.

& Little, 1997; Park et al., 2004). In a two-step procedure that we describe in more detail later in the paper, we estimate the likelihood that respondents select a given option for leaving college without a degree (e.g., textbooks were too expensive) and then reweight the estimates so that they are representative of the population of early-exiting students at our partner community colleges. Given that each participating community college enrolls a high number of students from the state of Florida–a large state with a diverse postsecondary student population–the MRP method allows this study to better estimate the generalizability of our results to the population of community college students who left college without a degree in the state of Florida and at institutions with similar demographic profiles throughout the United States.

Taken together, we find that a host of financial constraints—such as costs related to tuition and fees, living expenses, and a loss of financial aid—represent the most prevalent factors associated with why former community college students leave without a degree. Additional factors include a lack of time to study or prepare for class, increased work or family responsibilities, and substantive challenges associated with online learning. Importantly, we report variation in our findings across subgroups, as Black and Hispanic students were substantially more likely to encounter information and financial barriers that led to their early exit relative to their White peers.

Literature review

In an effort to increase the number of individuals who obtain the well-documented benefits associated with earning a college degree, the federal government provides over $120 billion on an annual basis to foster enrollment and persistence in higher education (Scott-Clayton, 2017). Public colleges and universities also invest resources in order to improve retention and completion rates, particularly those institutions with state funding linked to institutional performance measures (Ortagus, Kelchen, et al., 2020). However, community colleges face considerable financial challenges given their lack of public funding relative to flagship research universities (Hendrick et al., 2006) and low completion rates (Snyder et al., 2018).

Community colleges can play a democratizing role for individuals from disadvantaged or underserved backgrounds (Belfield & Bailey, 2011), but many community college students are unable to accrue the benefits associated with higher education because they do not complete their degree (Snyder et al., 2018). Specifically, only 38% of students who began at a community college completed their associate or bachelor's degree within six years of initial enrollment, and roughly half of students who began at a community college left college without earning their degree at any college or university (Juszkiewicz, 2017; Shapiro et al., 2017).

Factors associated with early departure from college

Previous research has reported that students' low GPA is correlated with their likelihood of dropping out of college (Hoyt & Winn, 2004; Stratton et al., 2008). Community college student attrition, in particular, has been linked to former students being academically underprepared for college-level coursework (*e.g.*, Holzer & Baum, 2017). In addition, college students who leave their institution without a degree often cite personal reasons for their early departure. For example, Johnson (2018) reported that former students often referenced a major change in family responsibilities, the need to accept a full-time job, or health concerns when identifying why they dropped out of college. Additional research points to students' family dynamics as a critical determinant of attrition while highlighting alternative personal reasons related to their decision to drop out of college, such as students' stress, anxiety, depression, burnout, and a lack of a sense of belonging on campus (Hunt et al., 2012).

Although prior work focuses disproportionately on the academic or personal challenges encountered by college dropouts, previous literature has also shown that many students who drop out of college are unable to complete their degree due to a variety of informational and financial barriers that are unrelated to their academic performance or personal circumstances (Long, 2007). Regarding informational barriers, many students who leave college without a degree cite bureaucratic or confusing administrative processes, poor academic advising, and a general lack of clarity pertaining to graduation requirements as primary reasons related to their early departure

(Bers & Schuetz, 2014; Johnson, 2018). Despite the relatively low price of community college enrollment, students who leave before earning their degree often do so in response to financial challenges unrelated to their academic performance (*e.g.*, Goldrick-Rab et al., 2016; Stinebrickner & Stinebrickner, 2008)

Previous literature has revealed that former students may leave college due to a need to work additional hours, a financial disruption related to their family obligations, and an inability to pay the required tuition and fees (Cox et al., 2016; Johnson, 2018). Financial aid programs have been found to be effective mechanisms to mitigate these financial barriers and decrease the likelihood of student attrition, particularly among low-income, racially minoritized, and academically underprepared students (*e.g.*, Goldrick-Rab et al., 2016).

Student characteristics associated with early departure from college

Prior work also suggests that students with certain types of demographic or academic characteristics are more likely to drop out of college. Several studies have indicated that Black and Hispanic students are significantly more likely to leave college without earning their degree (Juszkiewicz, 2017; Shapiro et al., 2017). Crosta (2013) highlighted the student characteristics that were associated with the decision to drop out of college before completing a degree. The author found that community college dropouts were more likely to be older and less likely to receive financial aid when compared to students who persisted. Financial barriers are prevalent for all types of college students but exacerbated among adult students with external responsibilities related to family or work (Bergman et al., 2014). A disproportionate number of community college students are working adults with family obligations, which creates time and location constraints that may force academically successful students to drop out of college (O'Toole et al., 2003; Schatzel et al., 2011; Stratton et al., 2008).

Additional research revealed that students who did not file the Free Application for Federal Student Aid (FAFSA) were more likely to drop out, especially among students who were enrolled part time (McKinney & Novak, 2013). Low-income students also have a greater likelihood of dropping out of college when compared to their more affluent peers. Even when low-income students receive need-based financial aid to pay for college, these financial aid allocations typically do not cover costs beyond tuition and fees (*i.e.*, rent and transportation) or account for familial pressures to send aid money to their low-income family members to cover rent and food expenses (Joo et al., 2008). Students who enroll in college on a part-time basis—many of whom are adults, full-time employees, and parents—are significantly more likely to drop out of college than full-time students (Attewell et al., 2012; O'Toole et al., 2003).

Conceptual framework

This study is guided by the economic theory of human capital to explain why former community college students may leave college before earning their degree. In the context of higher education, the theory of investment in human capital (Mincer, 1958) suggests that students make decisions about continuing their education based on the costs and benefits associated with enrolling (or re-enrolling) in college. The decision to remain enrolled in college, for example, is subject to a variety of considerations, such as the direct costs of tuition and the opportunity costs of forgone earnings, before determining whether higher education is a worthwhile investment. Before deciding to remain enrolled at a given college, students can weigh the costs and expected benefits of remaining enrolled in college and decide to remain enrolled only if the costs of staying in college are outweighed by the expected benefits in the future (DesJardins & Toutkoushian, 2005; Paulsen & Toutkoushian, 2008; Turner, 2004).

The economic theory of human capital also suggests that any individual's ability to generate economic value is associated with the knowledge, skills, and experiences accrued by that individual over time (Becker, 1962). Previous scholars have applied this theory to explain the benefits of obtaining additional education, particularly in relation to the decision-making process prospective students undertake when debating the merits of enrolling (or remaining enrolled) in college (Levin, 1989). The general logic of human capital theory suggests that investing in additional education (*i.e.*, remaining enrolled in college) will likely result in increased employee competencies and, as a result, a higher wage after entering the job market (Thomas & Perna, 2004). Because a human capital decision in this case may be constrained by a given student's budgetary limitations (Paulsen & Toutkoushian, 2008), this study explores the financial and non-financial reasons behind the decision to stop enrolling in college coursework before earning a degree.

The logical rationale pertaining to why students may leave college without a degree can also be explained by non-financial reasons, such as informational barriers caused by students' lack of understanding pertaining to which courses to take and the pathway to graduation. Bailey et al. (2015) highlighted the critical challenges students face when seeking to navigate the "cafeteria-style" of community colleges, noting that community college students are often "overwhelmed by the many choices available, resulting in poor program or course selection decisions, which in turn cost time and money, and likely lead many students to drop out in frustration" (p. 22). In this study, we focus specifically on students who were previously successful academically and eligible to re-enroll immediately in order to better understand the ways in which colleges can remove the financial and informational barriers that lead otherwise-successful students to make the decision to drop out of college.

Method

Approximately 8.9% of the students we contacted consented and took our survey ($N = 2,418$). While this low response rate was not unexpected—particularly considering that students in our population of interest had, by definition, stopped out of their postsecondary institution and were therefore more difficult to reach—it could affect the representativeness of our results. Specifically, if student response rates differed across observable demographic dimensions like gender, race/ethnicity, and age, or across levels of educational attainment like earned credit hours and GPA, then average responses among our sample may not reflect the average response of the population of interest.

We improve the generalizability of the survey responses through multilevel regression with poststratification (MRP), a statistical technique that allows us to reweight the responses so that they better represent the population of previously successful non-completers in our partner colleges. While MRP has been used extensively in political science to measure public opinion in the United States (Gao et al., 2019; Gelman et al., 2010; Gelman & Little, 1997; Howe et al., 2015; Kastellec et al., 2019; Kennedy & Gelman, 2019; Lax & Phillips, 2009; Lei et al., 2017; Lipps & Schraff, 2019; Little, 1993; Pacheco, 2011; Park et al., 2004; Wang et al., 2015; Warshaw & Rodden, 2012), and is increasingly used by political scientists outside of the U.S. (Lipps & Schraff, 2019; Toshkov, 2015), sociologists (Fairbrother & Martin, 2013), and epidemiologists (Downes et al., 2018; Eke et al., 2016; Zhang et al., 2014) to generate representative estimates from non-representative data, it has been seldom used in survey-based educational research.

In its most general form, MRP works using a two-step process:

1. Fit a multilevel model with K varying intercepts, α, whose categories partition J population cells. Typically, each intercept, α^k, will represent a demographic group with j unique categories: *e.g.*, $\alpha_j^{age}$where$j \in \{18-25, 26-35, 36-49, 50+\}$. With a binary outcome, the multilevel model will take the form,

$$Pr(y_i = 1) = \text{logit}^{-1}(\beta_0 + \sum_{k=1}^{K} \alpha_{j[i]}^{k}), \tag{1}$$

   in which β_0 is the grand mean. From this, we can get the average response for each cell, π_j.

2. Poststratify the average cell responses to the population average, θ, via

$$\theta = \frac{\sum_J N_j \pi_j}{\sum_J N_j}, \tag{2}$$

   the population cell sizes, N_j, as weights.

Population cell counts do not need to come from the same data set as the individual survey responses. In most MRP applications, population cell counts are provided by national census data sets such as the American Community Survey. Whatever the source, however, it must be the case that categorical variable indicators in equation (1) can be matched to population cell counts in the poststratification data set. For example, if the multilevel model contains varying intercepts for race/ethnicity and age, $\alpha^{race/ethnicity}$ and α^{age}, which have 6 and 4 categories, respectively, then the postratification matrix must have population-level counts for the $6 \times 4 = 24$ race/ethnicity by age group demographic cells possible in the multilevel model: *e.g.* number of Asian 25-39 year olds in the population.

In our study, we specifically fit,

$$Pr(y_i = 1) = \text{logit}^{-1}(\beta_0 + \alpha^{GPA}_{g[i]} + \alpha^{credit\,hours}_{h[i]} + \alpha^{gender}_{j[i]} + \alpha^{race/ethnicity}_{k[i]} + \alpha^{age}_{l[i]} + \alpha^{school}_{s[i]}), \quad (3)$$

in which $y_i \in \{0,1\}$ represents a possible survey item selection (*e.g.*, respondent selects tuition and fees as a cost-based reason contributing to early exit), β_0 is constant, and cell groups are represented by the following random intercepts: $\alpha^{GPA}_{g[i]}$ with $g \in \{[2.0-2.3),[2.3-2.7),[2.7-3.0),[3.0-3.3),[3.3-3.7),[3.7-4.0)\}$; $\alpha^{credit\,hours}_{h[i]}$ with $h \in \{<12, 12-23, 24-35, 36-48, 49-59, 60+\}$; $\alpha^{gender}_{j[i]}$ with $j \in \{$men, women, *missing*$\}$; $\alpha^{race/ethnicity}_{k[i]}$ with $k \in \{$Black, Hispanic, more than one racial/ethnic group, other racial/ethnic group, White, *missing*$\}$; $\alpha^{age}_{l[i]}$ with $l \in \{18-25, 26-35, 36-49, 50+\}$; and $\alpha^{school}_{s[i]}$ with $s \in \{1,2,3,4,5\}$. Combined, 12,960 unique cells are possible, though not every cell is represented in either the survey respondent group or the poststratified population of interest.

Using random intercepts in a multilevel model allows schools with comparatively fewer observations of a particular population cell to "borrow strength" from other schools, meaning that we can return estimates even for sparsely populated cells (Gelman et al., 2013). We assume that the random intercepts for age, gender, race/ethnicity, GPA, and earned credit hours, are normally distributed with mean of zero and group-specific variance,

$$
\begin{aligned}
&\alpha_g^{GPA} \sim N(0, \sigma_{GPA}^2) \\
&\alpha_h^{credit\,hours} \sim N(0, \sigma_{credit\,hours}^2) \\
&\alpha_j^{gender} \sim N(0, \sigma_{gender}^2) \\
&\alpha_k^{race/ethnicity} \sim N(0, \sigma_{race/ethnicity}^2) \\
&\alpha_l^{age} \sim N(0, \sigma_{age}^2),
\end{aligned}
$$

We model the school-level random intercept with

$$\alpha_s^{school} \sim N(\beta^{Comp.} X_{Comp._s} + \beta^{Cost} X_{Cost_s} + \beta^{Unem} X_{Unem_s} + \beta^{Wage} X_{Wage_s}, \sigma_{school}^2),$$

in which we use the institution's average 150% completion time and net cost as well as the institution county's unemployment rate and average weekly wage as second-level covariates to help account for differences across schools. In all cases, $\beta \sim N(0,1)$ and standard deviations are given as truncated standard normal prior, limited to positive values: $\sigma \sim N_+(0,1)$.

Due to the nature of the survey (discussed in the next section), we fit equation (3) for each possible response in the survey, r, and in turn compute cell-specific average responses, π_j^r via a modified version of equation (2),

$$\hat{\theta}_{response} = \frac{\sum_J N_j \pi_j^r}{\sum_J N_j}, \tag{4}$$

in which N_j come from the data set we used to construct and contact our initial sample of students in our population of interest. Once each $\theta_{response}$ is computed, we are able to rank them collectively by their medians to show relative prevalence of affirmative responses within our population of interest, that is, the recently stopped-out students with 2.0+ GPAs and no registration holds we were interested in reaching initially. We are also able to show differences in response across demographic subgroups (men compared to women, for example), which is useful when considering interventions targeted at the needs of specific populations.

Data

Survey

Individual-level responses come from a short web-based survey fielded in the summer of 2019 that asked former community college students about factors that might have contributed to their early departure. Partnering with five high-enrollment Florida College System (FCS) colleges, we first identified the subset of former students who, despite having made significant progress, left before earning a credential. To be included among our sampling frame, former students had to have earned at least a 2.0 GPA (cumulative) or higher, made progress toward degree completion (median of 42 credit hours accrued in the resulting sample), had no behavioral or financial holds that would prevent their re-enrollment, and to have stopped out within the prior three years. Because we were contacting former students, we were concerned about potentially low response rates. We therefore also required that students have an active cell phone number so that we could follow up via text messaging after the initial email request if necessary. We identified and contacted a total of 27,028 former students with a request to complete the survey.

The survey consisted of a total of 19 questions. Of these, the first (Q1) requested consent to participate, the last (Q19) asked if the respondent consented to be contacted for a follow-up interview, one (Q14) was open text response, and most of the remaining allowed respondents to select among a number of options. Questions in the latter group took the form of "select all that apply," with selections grouped by a general rationale for early exit. As an example, one question that asked about financial expenses took the form:

> Please indicate whether each of the following financial expenses describes a reason why you stopped attending [school name].[2]

[2]The name of the student's school was populated in the actual survey, varying across each of the five institutions.

Please check all that apply.

- ☐ Textbooks
- ☐ Tuition and fees
- ☐ Computer and internet access
- ☐ Transportation to campus
- ☐ Living expenses (rent, utilities, healthcare, childcare, food)

Students could click on the boxes to select all, none, or any combination of options before moving on to the next question group. Other question groups fell into four broad categories: *Cost*: financial expenses and financial aid issues; *Employment*: work-related issues; *Instructional*: course scheduling issues, course-related issues, issues with online courses; and *Other*: transportation/scheduling issues and personal issues. The survey was adaptive, meaning that not all students saw all questions. Students were only asked questions about employment-related issues and online coursework if they first affirmatively answered questions (Q4) “Were you employed while you were taking classes?” and (Q10) “Did you take any online courses?”, respectively. We account for this survey skip-logic in our analyses but note that most students were employed while in school (84%) and took at least one online course (58%).

In our analyses, we treat each selection as a binary outcome in a separate model: student selected (1) or did not select (0) the factor as reason for their early exit. Approximately 79.6% of students who started and consented to the survey completed it. This causes a small issue in the raw survey data since non-selection due to active non-choice and non-selection due to not having reached the question are both coded as missing. (This problem is most apparent on questions for which the student selected no factors—absent other information, it’s impossible to tell whether this is due to active non-choice or having not reached the question.) To differentiate between these two conditions, we used the survey system’s progress indicator to assess whether a student had *viewed* the question. If they had, we coded all non-selections as 0; if they had not, we left the values as missing.

We fit a total of 43 models. This includes 5-6 outcomes each for 8 question views as well as two outcomes used to account for the survey skip logic (Q4 and Q10, noted above). We use all available information for each model. Because of survey skip logic and non-survey completion, this means that survey sample sizes across models range from 1,248 to 2,418 respondents.[3] Appendix Table A.2 lists each potential factor, which is an outcome for a single model, along with the question and question category to which it belongs.

Second-level covariates

Data for the second level covariates come from two primary sources. The 150% completion rate and net cost for each school are taken from the Integrated Postsecondary Education Data System (IPEDS). The other two covariates, unemployment rate and average wage, come from the Bureau of Labor Statistics (BLS). These values are associated with the counties in which each college is located. All values are taken from data for the year just before the survey, 2018.

Poststratification weighting matrix

To construct our poststratification weighting matrix, we used administrative data from our five partner FCS schools. Unlike most other MRP studies, in which population counts come from a data source external to the original survey (*e.g.*, national census data), we are able to use the same source of data that we used to construct our original sampling frame. Table 1 shows the counts for a few poststratification cells.

One key benefit of using administrative data to construct our poststratification weighting matrix

[3]The lowest sample sizes come from questions about online courses. Of the full survey sample, 358 respondents did not reach Q10, which asked about participation in online coursework. Of those who did, only 1,269 indicated that they had taken online courses and were therefore given the opportunity to answer the series of questions (Q11.1 – Q11.5) about online coursework. Conditional on reaching Q10 and answering affirmatively, the sample of 1,248 represents a response rate 98%. Assuming all 358 non-responders to Q10 would have indicated participation in online courses, 1,248 represents a lower-bound response rate 76.7% compared to the full respondent sample.

is that we can use first-level covariates that are not typically available in external population count data. Specifically, we are able use bins of GPA and earned credit hour in addition to demographic covariates of gender, race/ethnicity, and age in our multilevel models. Because these academic characteristics are likely associated with some factors of early student exit, their inclusion provides a better fit to our data.

The sum total of all cell counts in our poststratification matrix is 58,531, which is larger than the 27,028 students we initially contacted to complete the survey. The discrepancy in the numbers reflects the fact that we required students in the initial sampling frame to have active cell phone numbers on file. Without this condition, the potential number of students to whom other conditions apply—2.0+ GPA, progress toward degree completion, no holds on the account—is higher. We poststratify to this population rather than to the initial sample frame because we believe this group better reflects the full population of interest: students who had made substantial progress to a degree with good academic standing who nonetheless exited without earning a credential.

Limitations

This study is subject to a number of limitations. First, despite the fact that we are able to use MRP to reweight our survey responses, we only poststratify to the five FCS colleges in the original sample. These are high enrollment colleges in a large state, meaning that results representative of these institutions may be representative of a larger population of early-exiting students. That said, any claims of external validity beyond the original five colleges in our study must rely on assumptions of similarity or representation beyond what we can directly model using MRP.

Second, the low numbers of some student groups in both our survey and poststratification data means that we are unable to provide useful inferences for these groups. For example, the number of students who identified as Asian, Native American, Pacific Islander, or Native Hawaiian was very low, meaning that we had to group them into a single racial/ethnic category. This lack of information is reflected in the wide poststratified posterior distributions for these subgroups. This limits our ability to speak to the experiences of these groups as they pertain to early exit, either within the five colleges we study, the state of Florida, or the nation on the whole. Finally, our results may be biased if those students who completed the survey are different from those students who did in ways we cannot observe. Though MRP has demonstrated good properties even among non-representative polls (Wang et al., 2015), we cannot test whether our analysis has corrected for any survey response bias.

Results

In table 2 we compare survey respondents (columns 1 and 4, $N = 2,418$) with two groups: the group of students we initially contacted to complete the survey (column 2, $N = 27,028$) and the full population of interest (column 5, $N = 58,531$). In both cases, survey respondents differ across a range of demographic characteristics at conventional levels of statistical significance. Compared to both the contacted student sample and the full population of interest, survey respondents skewed older, were more likely to identify as women, and were more likely to identify as Black. Compared to the contacted sample, survey respondents were less likely to identify as Hispanic only; compared to the full population, survey respondents were less likely to identify as White or have missing information on their race/ethnicity.[4]

Academically, survey respondents tended to have fewer students with C- average GPAs and more students with B/B+ GPA averages the full sample of those contacted. Compared to the full population, respondents had fewer students with GPA averages at the low (C) and high (A-/A) ranges and more with C+ averages. Compared to both groups, survey respondents tended to have earned more credit hours prior to leaving, with a particularly pronounced difference between

[4]Respondents were able to choose any combination of racial/ethnic categories when self-identifying. The small number of those who chose multiple identities were placed in the joint "More than one racial/ethnic group" category due to sample size limitations. All other categories represent the choice of a single racial/ethnic identity. Respondents had the option not to select any racial/ethnic category, leaving a missing value; the same is true for gender (which was giving a binary male/female option set). While we cannot differentiate between missing values of gender and race/ethnicity due to refusal to answer versus simple lack of data, our random effects model is flexible enough to include these categories.

respondents and the full population of interest.

Due to these differences between survey respondents, we reweight our findings to improve their representativeness. Though we make comparisons with the students who were initially contacted, we poststratify using cell counts from the full population frame. All subsequent results, therefore, reflect the full population of early-exiting community college students in our study who had at least a 2.0 GPA, made meaningful progress toward their degree (median of 42 credit hours accrued), and had no financial holds barring their return as of their last enrolled semester.

We present our primary results in Figure 1.[5] Each row represents a reason that a former student could select as having contributed to their early exit. Because we fit each reason as an outcome in its own model and treat it as binary choice, each row can be interpreted as the percentage of students in the population of interest who cite the reason as one that contributed to their early exit. We group reasons into four broad categories: cost, employment, instructional, and other. The center shape of each line represents the median posterior value and the thick and thin horizontal lines on either side showing the 50% and 95% credible intervals, respectively.[6]

Overall, the two reasons for early exit indicated by more than half of all students involve financial costs to students: "tuition and fees" (52.2%), and "living expenses" (*e.g.*, rent, utilities, health care, child care, food) (51.3%). More than a third of students indicated another cost, "no longer being eligible for financial aid" (42%), as well as a "lack of time to study and prepare for class" (37.1%), a "switch from part-time to full-time work" (36.2%), [7] and an "inconsistent weekly schedule" (33.3%). Approximately one in four former students cite "difficulty learning in

[5]We fit our multilevel models using the Stan NUTS sampler, a variant of the Hamiltonian Monte Carlo MCMC sampler (Carpenter et al., 2017; Hoffman & Gelman, 2014). We fit 4 separate chains of 2,000 draws, throwing away the first 1,000 in each chain as warm up. The Stan script used for each model is included in the appendix.

[6]Table A.1, which compares poststratified distributions (shown in figure 1) to unadjusted survey mean values, is available in the appendix.

[7]Due to the skip logic of the survey, all poststratified responses involving work and online courses take into account the probability of affirmatively selecting the gateway question (*e.g.*, “Were you employed while you were taking classes?”) via a modified version of equation (4),

$$\hat{\theta}_{response} = \frac{\sum_J N_j \pi^{r1} \pi_j^{r2}}{\sum_J \pi^{r1} N_j},$$

where π^{r1} is the probability of answering the gateway question affirmatively and π^{r2} is the probability of selecting the subsequent work- or online course-related reason for leaving (see Park et al., 2004).

an online setting" (25.9%), "uncertainty about which classes to take next" (25.6%), "lack of desired classes at the campus location closest to them" (25.6%), "required math and science courses that were too difficult" (25.3%), "too little faculty interaction in online courses" (24.9%), the "cost of textbooks" (24.8%), and "difficulty in completing assignments" (23.7%), with one in five citing the "unavailability of a required course online" (20.3%). The remaining reasons were cited by less than 20% of students, with one, "employer stopped paying for classes," cited by less than 1 in 20 students (3.3%). That said, many were cited by approximately 1015% of students. Furthermore, smaller average percentages may cover heterogeneity in responses among different student groups. We explore some of these heterogeneous responses in the next section.

Results by subgroup

Figures 2-8 show differences in subgroup responses for a selected set of question options. We do not present results for all outcomes but rather focus on a few that show important differences among subgroups. The subgroup affirmative response rates—that is, the percentage of students within the subgroup citing the question option as a reason for their early exit—are shown within their own facet.[8] As with figure 1, the open circles within the figure represent the poststratified posterior distribution median and the thick and thin lines the 50% and 95% credible intervals, respectively.

Figure 2 presents the results for the survey response option, “Changed careers,” with 20.9% of men and 13.4% of women citing this factor as a reason for their early exit. In figure 3, we see that 18.2% of women cited a “health emergency” as a reason for exit, whereas only 12.6% of men did so. Women were also much more likely (16.3%) than men (6.8%) to say they exited because they “did not have reliable childcare” (Figure 4). For lack of childcare, we also note a u-shaped difference among age brackets, with 26-35 year-olds (15.3%) and 36-49 year-olds (17.0%) citing this reason compared to only 8.5% of 18-25 year-olds and 6.2% of those 50 years and older.

[8]Due to relatively small sample sizes and lack of clear interpretation, we omit missing categories from the gender and race/ethnicity subgroups in the figures.

In figures 5 and 6, we show results from two financial-based response options: "missed payment deadline and was dropped" and "registration hold." Per the second option, we remind that a student was only eligible for inclusion in our sample and population if they did not have any registration holds that would otherwise prevent their return. In both cases, two patterns emerge. First, students with lower GPAs are more likely to cite these financially motivated reasons than those students with the highest GPAs. For example, whereas 6.1% of students with a GPA of 3.7+ cited a missed payment and drop out, 16.5% and 12.8% of students with a C and C+/B- average, respectively, did so. Similarly, only 6.6% of students with a 3.7+ GPA cited a registration hold while median responses from students with GPAs lower than 3.0 range from 12.4% to 23.4%. We also note sharp differences between White and Black and Hispanic students for these two outcomes. While 5.0% of White students cited a missed payment that required them to drop, 11.9% of Hispanic students and 16.9% of Black students did so. Black (26.5%) and Hispanic (18.9%) students cited registration holds at much higher rates than White students (7.6%). We discuss the implications of these findings in the next section.

Figure 7 indicates the extent to which different subgroups of students selected "difficulty learning on [their] own in online settings." As an example, 33.6% of students with lower GPAs noted struggles with online learning as related to their decision to leave college without a degree, whereas only 16.6% of students with higher GPAs cited difficulties with online learning. We also show that a larger share of Black (28.4%) and Hispanic students (32.1%) noted challenges with online learning as a factor behind their decision to exit early relative to the proportion of White students (19.6%) who struggled with online learning. Similarly, we find that difficulties associated with "unreliable internet access" (Figure 8) were more prominent among Black students (12.5%) when compared to White students (3.7%) in our sample.

Discussion and conclusions

To better understand the factors associated with previously successful students' decision to drop out of college, we surveyed over 27,000 former students in a large community college system. Through this work, we address a critical problem in higher education research by improving the generalizability of the survey responses through the use of multilevel regression with poststratification in order to reweight responses to better represent the population in our original survey frame. Although prior work has focused on the reasons behind academic challenges of students who depart college without a degree, our study provides representative survey data and specifically includes former students who were performing well academically before leaving college without their degree.

Our findings show that tuition and fees, living expenses, and the loss of financial aid eligibility represent primary factors shared by the majority of former students in our sample. We also find considerable variation across subgroups of former students. For example, older students, particularly women, were more likely to highlight a lack of reliable childcare as a primary reason behind their decision to drop out of college. Women were also more likely than men to cite a health emergency, which could have applied to themselves or another person for whom they were a caregiver. Importantly, Black and Hispanic students were found to be much more likely than White students to cite a missed payment or registration hold when noting why they exited early, even though none of the surveyed students had holds that would have prevented their reregistration.

Community colleges often represent an important mechanism to allow individuals from underserved populations to climb the socioeconomic ladder (Belfield & Bailey, 2011), but many community college students are unable to accrue the financial or non-financial benefits associated with going to college because they do not earn their degree (Snyder et al., 2018). Prior work has shown that family issues or financial disruptions often force students into precarious academic environments that lead to poor grades and may explain why students drop out of college (Hoyt & Winn, 2004; Johnson, 2018; Stratton et al., 2008). Although previous research has not explored the extent to which online learning influences students' decision to drop out of college, prior studies examining the efficacy of online education have shown that Black students and students with lower levels of educational attainment have a lower likelihood of success in self-directed online learning environments when compared to their peers (Xu & Xu, 2020). Our finding that Black students and those with lower GPAs were more likely to cite troubles with unreliable internet access offers one potential reason for this difference. Taken together, this study advances our understanding of why students leave their community college without a degree by focusing specifically on providing representative survey data that centers former students who were previously successful academically.

Given that community college students who stop enrolling for any period of time are substantially less likely to graduate than their peers who remain enrolled, the critical questions facing administrators and policymakers are twofold: (1) Why do these students leave without a degree in the first place? (2) What can community colleges do to prevent their early departure? Our empirical results explain the former question, and we begin to explore the latter question in the remainder of this section. Community colleges are unlikely to be able to simply lower their tuition and fees in response to the financial challenges faced by students leaving without a degree; however, institutions may be able to bolster their efforts related to ensuring that important financial aid information is easily accessible online and students are able to complete the FAFSA. The consequences associated with not taking advantage of available financial aid are dire, as prior research has shown that students who did not file for the FAFSA were significantly more likely to drop out of college (McKinney & Novak, 2013).

Aside from broad and necessary efforts to ensure financial aid information is more easily available online to all students, community colleges can also provide targeted financial aid packages to students who are close to finishing their degree but running out of financial aid. Unfortunately, community colleges receive substantially less public funding than four-year institutions (Hendrick et al., 2006; Kelchen et al., 2020) and are already asked to do more with less. For community colleges, the combination of constrained resources and relative low completion rates outlines the dire need to make targeted, data-driven financial decisions to optimize the impact of institutional practices and policies. This study offers an important look at an understudied student population: community college students who left college despite succeeding academically before their departure. Prior work reveals that students face the greatest risk of dropping out when taking entry-level courses, particularly math courses, but the former students in our sample are typically well beyond those introductory courses and, as a consequence, more likely to complete college should they return to college (Shapiro et al., 2019). Through our novel data source and empirical approach, we offer important takeaways for administrators or policymakers seeking to better understand why these former students left in order to address how to minimize the number of early departures and increase the number of students who complete their degree.

Another important contribution of our study, which outlines why previously successful former students drop out of college, is to provide a clear framework pertaining to how to optimize efforts designed to encourage or incentivize former students to return to college and complete their degree. Due to challenges associated with declining enrollments, inadequate state funding, and low completion rates, a growing number of community colleges have launched re-enrollment campaigns designed to foster re-enrollment among former students who have already made considerable progress toward completing their degree (Schwartz, 2019). By outlining the specific factors associated with why previously successful community college students dropped out of college, future re-enrollment initiatives can seek to ameliorate the specific barriers that should be addressed to encourage and incentivize former students to return to college.

Finally, as we referenced previously, many survey-based studies in education research address important questions but fail to provide representative findings or, as a consequence, generalize to their population of interest. This study analyzes responses using MRP and thereby produces more representative estimates of why academically successful community college students left college without earning their degree. As community colleges and other institution types continue to employ surveys as a mechanism to engage students and gather evidence to make institutional decisions, our findings represent an important step toward improving the generalizability of surveys used in education research while directly addressing the completion problem facing not only community colleges but also higher education at large.

References

Armstrong, E. A., & Hamilton, L. T. (2013). *Paying for the party*. Cambridge, MA, Harvard University Press.

Attewell, P., Heil, S., & Reisel, L. (2012). What is academic momentum? And does it matter? *Educational Evaluation and Policy Analysis*, *34*(1), 27–44.

Bailey, T. R., Jaggars, S. S., & Jenkins, D. (2015). *Redesigning America's community colleges*. Harvard University Press.

Becker, G. S. (1962). Investment in human capital: A theoretical analysis. *Journal of Political Economy*, *70*(5, Part 2), 9–49.

Belfield, C. R., & Bailey, T. (2011). The benefits of attending community college: A review of the evidence. *Community College Review*, *39*(1), 46–68.

Bemmel, E. P., Floyd, D. L., & Bryan, V. C. (2008). Perceptions and reflections of administrators: Community colleges transitioning to baccalaureate colleges. *Community College Journal of Research and Practice*, *33*(2), 151–176.

Bergman, M., Gross, J. P. K., Berry, M., & Shuck, B. (2014). If life happened but a degree didn't: Examining factors that impact adult student persistence. *The Journal of Continuing Higher Education*, *62*(2), 90–101.

Bers, T., & Schuetz, P. (2014). Nearbies: A missing piece of the college completion conundrum. *Community College Review*, *42*(3), 167–183.

Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software, Articles*, *76*(1), 1–32.

Cohen, A. M., & Brawer, F. B. (2008). *The american community college*. John Wiley & Sons.

Cox, B. E., Reason, R. D., Nix, S., & Gillman, M. (2016). Life happens (outside of college): Non-college life-events and students' likelihood of graduation. *Research in Higher Education*, *57*(7), 823–844.

Crosta, P. M. (2014). Intensity and attachment: How the chaotic enrollment patterns of community college students relate to educational outcomes. *Community College Review*, *42*(2), 118–142.

Crosta, P. M. (2013). *Characteristics of early community college dropouts* (CCRC Analytics). Community College Research Center.

DesJardins, S. L., Ahlburg, D. A., & McCall, B. P. (2006). The effects of interrupted enrollment on graduation from college: Racial, income, and ability differences. *Economics of Education Review*, *25*(6), 575–590.

DesJardins, S. L., & Toutkoushian, R. K. (2005). Are students really rational? The development of rational thought and its application to student choice. In *Higher education: Handbook of theory and research* (pp. 191–240). Springer.

Downes, M., Gurrin, L. C., English, D. R., Pirkis, J., Currier, D., Spittal, M. J., & Carlin, J. B. (2018). Multilevel regression and poststratification: A modelling approach to estimating population quantities from highly selected survey samples. *American Journal of Epidemiology*, *187*(8), 1780–1790.

Doyle, W. R., & Skinner, B. T. (2016). Estimating the education-earnings equation using geographic variation. *Economics of Education Review*, *53*, 254–267.

Doyle, W. R., & Skinner, B. T. (2017). Does postsecondary education result in civic benefits? *The Journal of Higher Education*, *88*(6), 863–893.

Eke, P., Zhang, X., Lu, H., Wei, L., Thornton-Evans, G., Greenlund, K., Holt, J., & Croft, J. (2016). Predicting periodontitis at state and local levels in the United States. *Journal of Dental Research*, *95*(5), 515–522.

Fairbrother, M., & Martin, I. W. (2013). Does inequality erode social trust? Results from multilevel models of U.S. states and counties. *Social Science Research*, *42*(2), 347–360.

Gao, Y., Kennedy, L., Simpson, D., & Gelman, A. (2019). Improving multilevel regression and poststratification with structured priors. *arXiv preprint arXiv:1908.06716*.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis* (3rd). Chapman; Hall/CRC.

Gelman, A., Lee, D., & Ghitza, Y. (2010). Public opinion on health care reform. *The Forum*, *8*(1).

Gelman, A., & Little, T. C. (1997). Poststratification into many categories using hierarchical logistic regression. *Survey Methodology*, *23*, 127–135.

Goldrick-Rab, S., Kelchen, R., Harris, D. N., & Benson, J. (2016). Reducing income inequality in educational attainment: Experimental evidence on the impact of financial aid on college completion. *American Journal of Sociology*, *121*(6), 1762–1817.

Hendrick, R. Z., Hightower, W. H., & Gregory, D. E. (2006). State funding limitations and community college open door policy: Conflicting priorities? *Community College Journal of Research and Practice*, *30*(8), 627–640.

Hoffman, M. D., & Gelman, A. (2014). The No-U-Turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, *15*(1), 1593–1623.

Holzer, H. J., & Baum, S. (2017). *Making college work: Pathways to success for disadvantaged students*. Brookings Institution Press.

Howe, P. D., Mildenberger, M., Marlon, J. R., & Leiserowitz, A. (2015). Geographic variation in opinions on climate change at state and local scales in the usa. *Nature Climate Change*, *5*, 596–603.

Hoyt, J. E., & Winn, B. A. (2004). Understanding retention and college student bodies: Differences between drop-outs, stop-outs, opt-outs, and transfer-outs. *NASPA journal*, *41*(3), 395–417.

Hunt, P. F., Boyd, V. S., Gast, L. K., Mitchell, A., & Wilson, W. (2012). Why some students leave college during their senior year. *Journal of College Student Development*, *53*(5), 737–742.

Jepsen, C., Troske, K., & Coomes, P. (2014). The labor-market returns to community college degrees, diplomas, and certificates. *Journal of Labor Economics*, *32*(1), 95–121.

Johnson, S. L. (2018). Why seniors leave and why they return: An exploratory case study exploratory case study. *Journal of College Student Retention: Research, Theory & Practice*, 1–18.

Joo, S.-H., Durband, D. B., & Grable, J. (2008). The academic impact of financial stress on college students. *Journal of College Student Retention: Research, Theory & Practice*, *10*(3), 287–305.

Juszkiewicz, J. (2017). Trends in community college enrollment and completion data, 2017.

Kastellec, J. P., Lax, J. R., & Phillips, J. H. (2019). *Estimating state public opinion with multi-level regression and poststratification using R* [Unpublished Manuscript]. Unpublished Manuscript.

Kelchen, R., Ortagus, J. C., Baker, D., & Rosinger, K. (2020). Trends in state funding for public higher education.

Kennedy, L., & Gelman, A. (2019). *Know your population and know your model: Using model-based regression and poststratification to generalize findings beyond the observed sample* (arXiv preprint No. 1906.11323v1).

Lax, J. R., & Phillips, J. H. (2009). How should we estimate public opinion in the states? *American Journal of Political Science*, *53*(1), 107–121.

Lei, R., Gelman, A., & Ghitza, Y. (2017). The 2008 election: A preregistered replication analysis. *Statistics and Public Policy*, *4*(1), 1–8.

Levin, H. M. (1989). Mapping the economics of education: An introductory essay. *Educational researcher*, *18*(4), 13–17.

Lipps, J., & Schraff, D. (2019). Estimating subnational preferences across the European Union. *Political Science Research and Methods*, 1–9.

Little, R. J. A. (1993). Post-stratification: A modeler's perspective. *Journal of the American Statistical Association*, *88*(423), 1001–1012.

Long, B. T. (2007). The contributions of economics to the study of college access and success. *Teachers College Record*, *109*(10), 2367–2443.

McKinney, L., & Novak, H. (2013). The relationship between FAFSA filing and persistence among first-year community college students. *Community College Review*, *41*(1), 63–85.

Mincer, J. (1958). Investment in human capital and personal income distribution. *Journal of Political Economy*, *66*(4), 281–302.

Ortagus, J. C., Kelchen, R., Rosinger, K., & Voorhees, N. (2020). Performance-based funding in American higher education: A systematic synthesis of the intended and unintended consequences. *Educational Evaluation and Policy Analysis*, *42*(4), 520–550.

Ortagus, J. C., Tanner, M., & McFarlin, I. (2020). Can re-enrollment campaigns help dropouts return to college? Evidence from Florida community colleges. *Educational Evaluation and Policy Analysis*.

O'Toole, D. M., Stratton, L. S., & Wetzel, J. N. (2003). A longitudinal analysis of the frequency of part-time enrollment and the persistence of students who enroll part time. *Research in Higher Education*, *44*(5), 519–537.

Pacheco, J. (2011). Using national surveys to measure dynamic U.S. state public opinion. *State Politics & Policy Quarterly*, *11*(4), 415–439.

Park, D. K., Gelman, A., & Bafumi, J. (2004). Bayesian multilevel estimation with poststratification: State-level estimates from national polls. *Political Analysis*, *12*(4), 375–385.

Paulsen, M. B., & Toutkoushian, R. K. (2008). Economic models and policy analysis in higher education: A diagrammatic exposition. In J. Smart (Ed.), *Higher education* (pp. 1–48). Springer.

Schatzel, K., Callahan, T., Scott, C. J., & Davis, T. (2011). Reaching the non-traditional stopout population: A segmentation approach. *Journal of Marketing for Higher Education*, *21*(1), 47–60.

Schwartz, N. (2019). How colleges are bringing back stopped-out students. *Education Dive*.

Scott-Clayton, J. (2017). Undergraduate financial aid in the united states.

Shapiro, D., Dundar, A., Huie, F., Wakhungu, P. K., Yuan, X., Nathan, A., & Bhimdiwali, A. (2017). *Completing college: A national view of student completion rates–fall 2011 cohort* (Signature Report No. No. 14). National Student Clearinghouse Research Center.

Shapiro, D., Ryu, M., Huie, F., & Liu, Q. (2019). *Some college, no degree: A 2019 snapshot for the nation and 50 states* (Signature Report No. 17). National Student Clearinghouse Research Center. Herndon, VA.

Snyder, T. D., de Brey, C., & Dillow, S. (2018). *Digest of education statistics 2016* (NCES No. 2017-094). National Center for Education Statistics.

Stinebrickner, R., & Stinebrickner, T. (2008). The effect of credit constraints on the college drop-out decision: A direct approach using a new panel study. *American Economic Review*, *98*(5), 2163–84.

Stratton, L. S., O'Toole, D. M., & Wetzel, J. N. (2008). A multinomial logit model of college stopout and dropout behavior. *Economics of Education Review*, *27*(3), 319–331.

Thomas, S. L., & Perna, L. W. (2004). The opportunity agenda: A reexamination of postsecondary reward and opportunity. In *Higher education: Handbook of theory and research* (pp. 43–84). Springer.

Toshkov, D. (2015). Exploring the performance of multilevel modeling and poststratification with Eurobarometer data. *Political Analysis*, *23*(3), 455–460.

Trostel, P. A. (2015). *It's not just the money the benefits of college education to individuals and to society* (Lumina Issues Papers). Lumina Foundation.

Turner, S. (2004). Going to college and finishing college. Explaining different educational outcomes. In *College choices: The economics of where to go, when to go, and how to pay for it* (pp. 13–62). University of Chicago Press.

Wang, W., Rothschild, D., Goel, S., & Gelman, A. (2015). Forecasting elections with non-representative polls. *International Journal of Forecasting*, *31*(3), 980–991.

Warshaw, C., & Rodden, J. (2012). How should we measure district-level public opinion on individual issues? *The Journal of Politics*, *74*(1), 203–219.

Xu, D., & Xu, Y. (2020). The ambivalence about distance learning in higher education. In L. W. Perna (Ed.), *Higher education: Handbook of theory and research* (pp. 351–401). Springer International Publishing.

Zhang, X., Holt, J. B., Lu, H., Wheaton, A. G., Ford, E. S., Greenlund, K. J., & Croft, J. B. (2014). Multilevel regression and poststratification for small-area estimation of population health outcomes: A case study of chronic obstructive pulmonary disease prevalence using the behavioral risk factor surveillance system. *American Journal of Epidemiology*, *179*(8), 1025–1033.

Table 1: Example poststratification cell counts

School	Gender	Race/ethnicity	Age	GPA	Credit hours	**Count**
1	Men	Black	18-25	[2.0-2.3)	12-23	**21**
1	Men	Black	18-25	[2.0-2.3)	23-35	**27**
1	Men	Black	18-25	[2.0-2.3)	36-47	**39**
1	Men	Black	18-25	[2.0-2.3)	48-59	**16**
1	Men	Black	18-25	[2.3-2.7)	23-35	**38**
...	...	...	...	...	...	...
5	Women	(Missing)	18-25	[3.7-4.0)	36-47	**78**
5	Women	(Missing)	18-25	[3.7-4.0)	48-59	**52**
5	Women	(Missing)	18-25	[3.7-4.0)	60+	**47**
5	Women	(Missing)	26-35	[3.7-4.0)	60+	**14**
5	Women	(Missing)	36-49	[3.7-4.0)	60+	**12**

Notes. The full poststratification table contains counts for 5 schools; 3 genders: men, women, and missing; 6 race/ethnicities: Black, Hispanic, more than one race/ethnicity, white, other racial/ethnic groups, and missing; 4 age groups: 18-25, 26-35, 36-49, and 50+; 6 GPA bins: [2.0-2.3), [2.3-2.7), [2.7-3.0), [3.0-3.3), [3.3-3.7), and [3.7-4.0); and 6 earned credit hour bins: <12, 12-23, 24-35, 36-47, 48-59, and 60+ hours.

Table 2: Comparison of respondents to those who were contacted and the population of interest across characteristics

	Respondent	Contacted	Sig.	Respondent	Population	Sig.
	(1)	(2)	(3)	(4)	(5)	(6)
Age group						
18-25	25.43	32.20	***	25.43	39.21	***
26-35	34.20	39.46	***	34.20	34.29	
36-49	27.01	21.30	***	27.01	18.99	***
50+	13.36	7.04	***	13.36	7.50	***
Gender						
Men	33.09	41.46	***	33.09	42.10	***
Women	65.92	57.61	***	65.92	56.77	***
(Missing)	0.99	0.93		0.99	1.14	
Race/ethnicity						
Black	28.21	23.96	***	28.21	21.34	***
Hispanic	13.23	16.31	***	13.23	14.01	
More than one racial/ethnic group	27.87	29.18		27.87	27.78	
Other racial/ethnic group	2.07	2.69	+	2.07	2.82	*
White	24.15	23.15		24.15	27.65	***
(Missing)	4.47	4.71		4.47	6.40	***
GPA						
[2.0-2.3)	13.65	20.36	***	13.65	15.39	*
[2.3-2.7)	23.78	24.27		23.78	22.98	
[2.7-3.0)	19.35	17.89	+	19.35	16.60	***
[3.0-3.3)	19.27	16.76	**	19.27	19.37	
[3.3-3.7)	14.60	12.31	**	14.60	14.81	
[3.7-4.0)	9.35	8.40		9.35	10.84	*
Earned credit hours						
<12	4.67	8.10	***	4.67	14.82	***
12-23	9.47	10.83	*	9.47	14.44	***
23-35	23.49	23.89		23.49	21.17	**
36-47	24.98	24.83		24.98	19.88	***
48-59	25.72	23.51	*	25.72	20.05	***
60+	11.66	8.85	***	11.66	9.64	***
N	2418	27028		2418	58531	

Notes. $+p < 0.1$, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$. All numbers are percentages. Columns (3) and (6) indicate the level of statistical significance of differences between the preceding two columns. Square brackets and parentheses around GPA intervals are inclusive and exclusive, respectively.

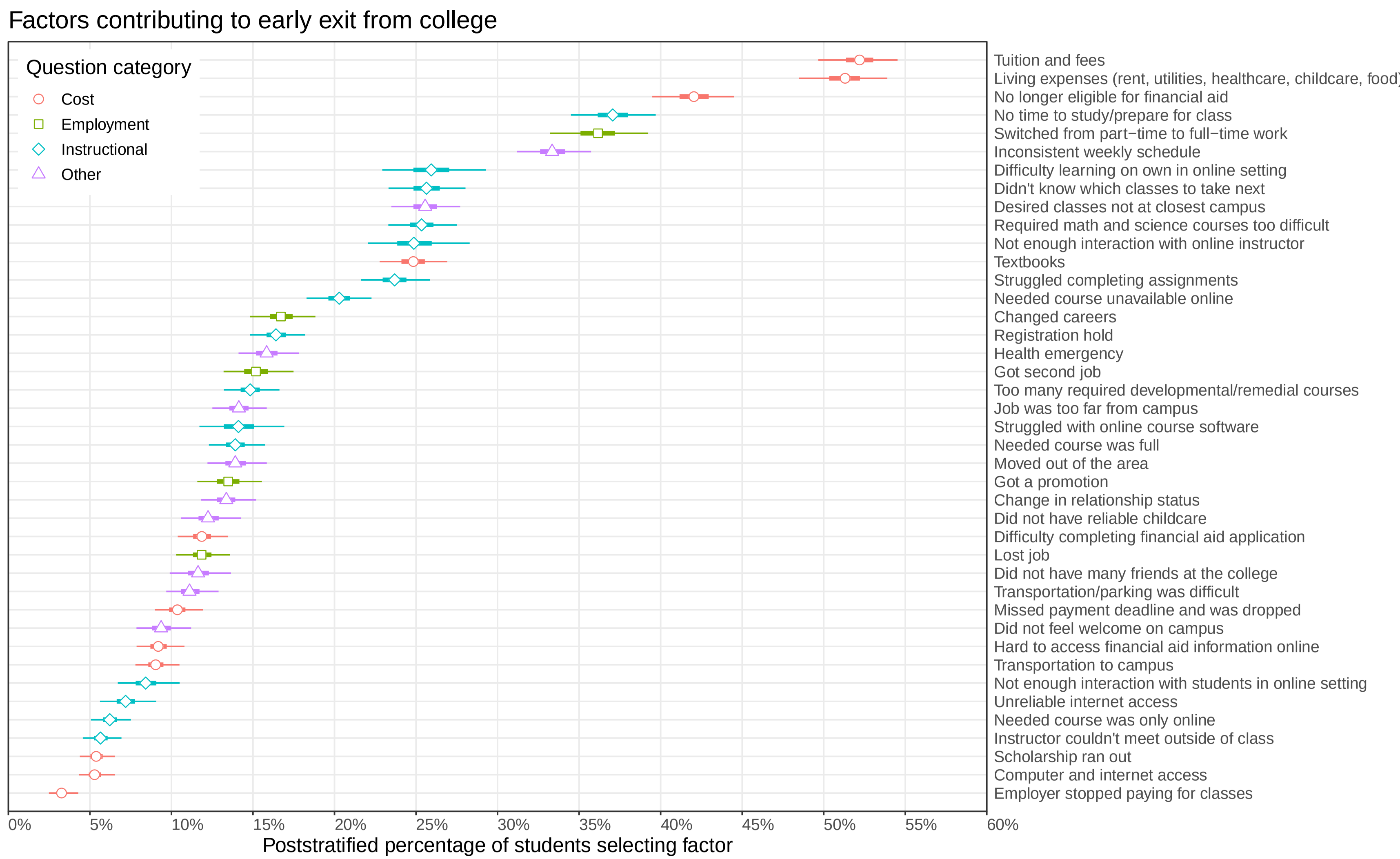

Figure 1: Each row represents a cited factor in early exit. Factors are grouped into four broad categories (indicated by shape and color). The center shape represents the median of the poststratified posterior density, which is interpretable as the percentage of students selecting the factor as having contributed to their early exit. The thick and thin lines around each median value represent the 50% and 95% credible intervals, respectively.

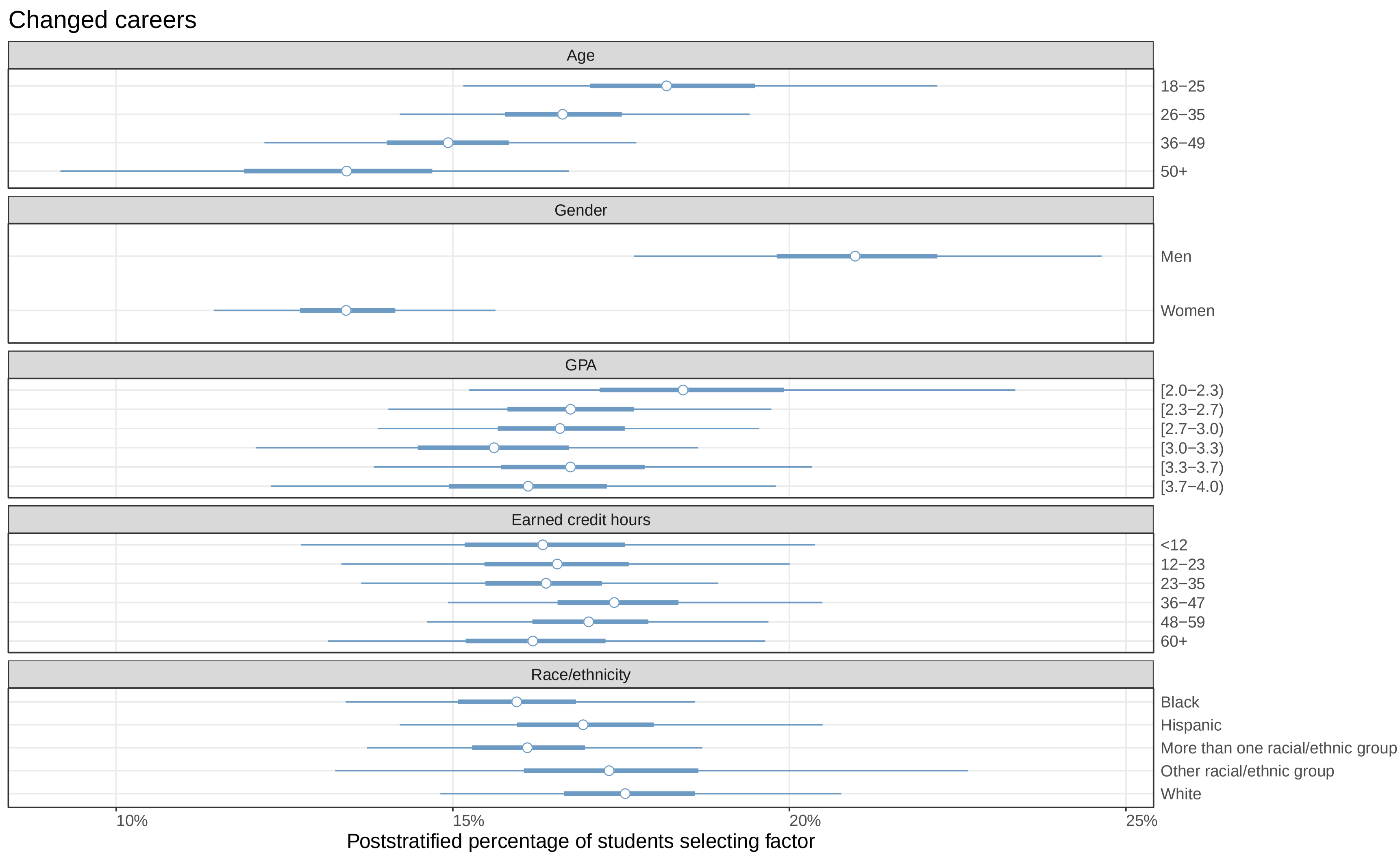

Figure 2: Each facet represents the poststratified posterior results for one question answer (outcome: was the checkbox selected), separated by within-group responses. The open circle represents the poststratified posterior median value. The thick and thin blue lines represent the 50% and 95% credible intervals, respectively.

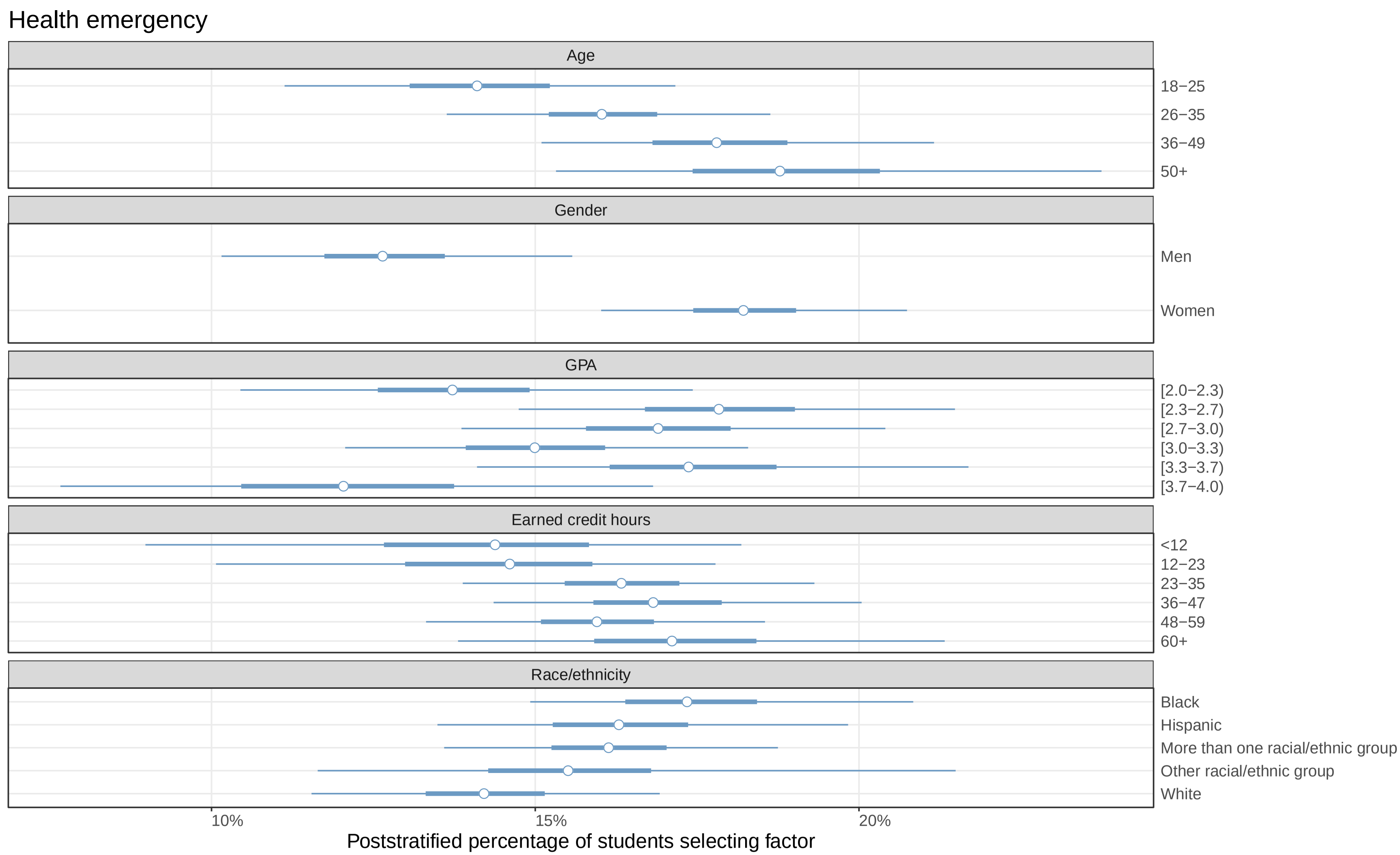

Figure 3: Each facet represents the poststratified posterior results for one question answer (outcome: was the checkbox selected), separated by within-group responses. The open circle represents the poststratified posterior median value. The thick and thin blue lines represent the 50% and 95% credible intervals, respectively.

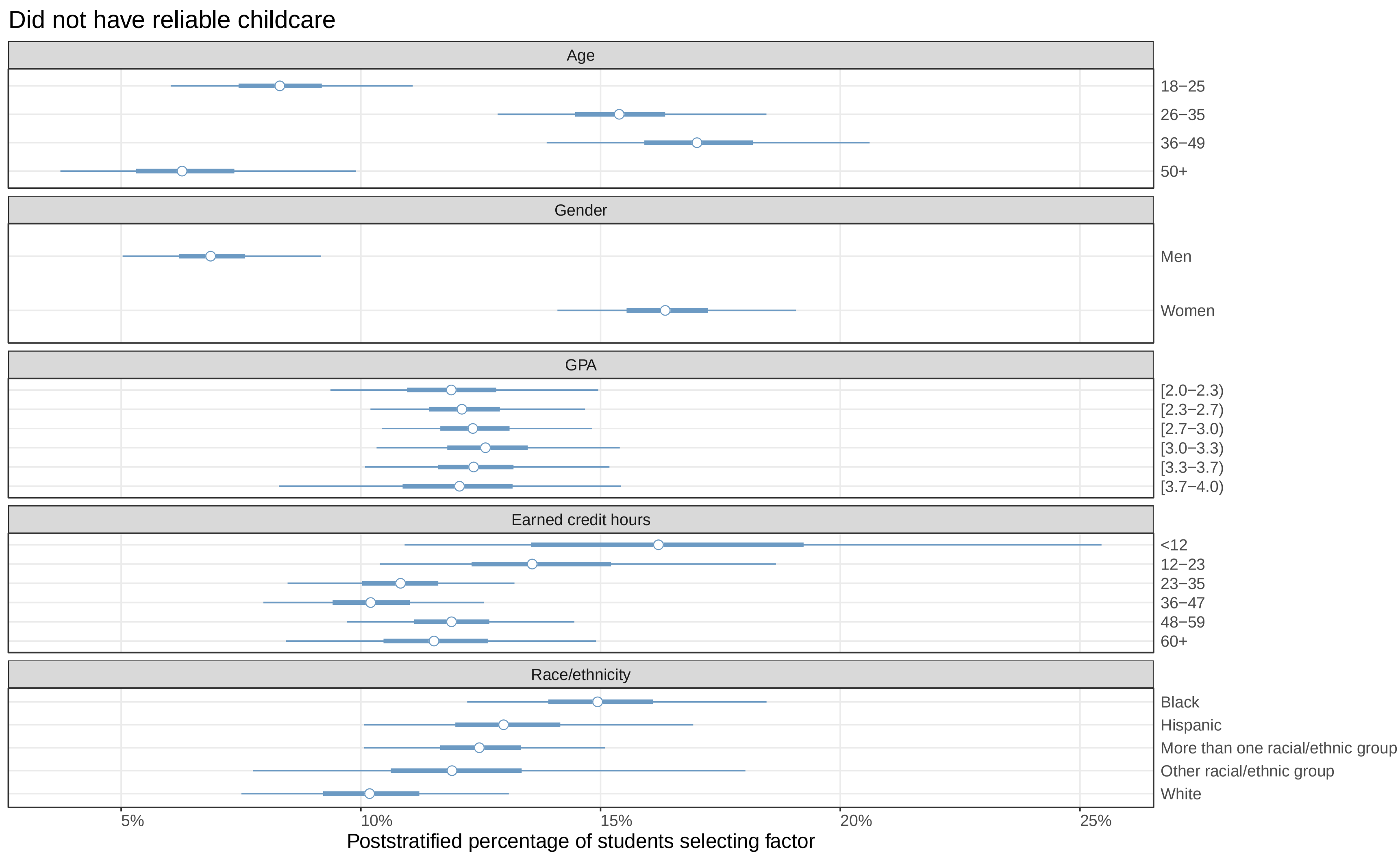

Figure 4: Each facet represents the poststratified posterior results for one question answer (outcome: was the checkbox selected), separated by within-group responses. The open circle represents the poststratified posterior median value. The thick and thin blue lines represent the 50% and 95% credible intervals, respectively.

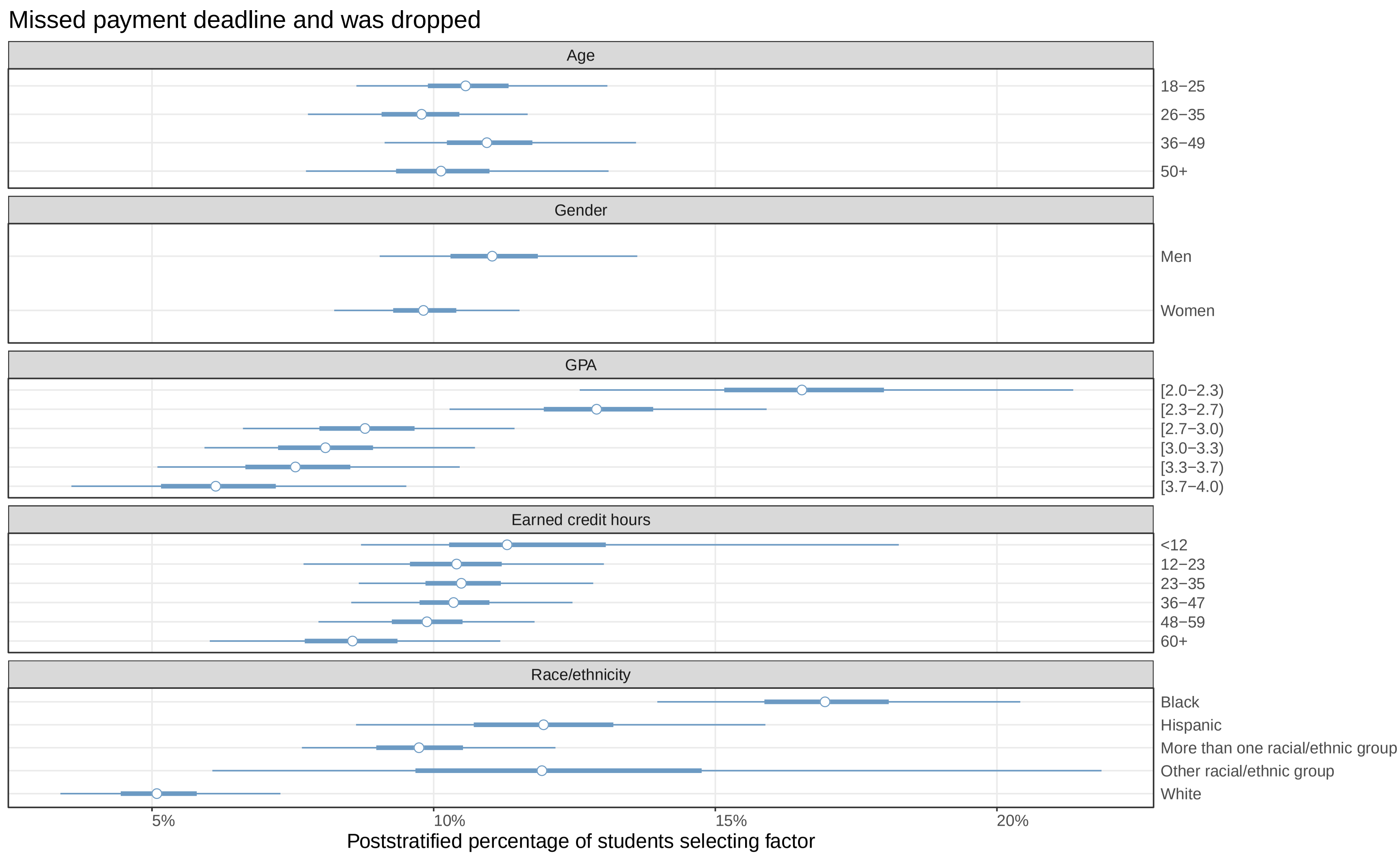

Figure 5: Each facet represents the poststratified posterior results for one question answer (outcome: was the checkbox selected), separated by within-group responses. The open circle represents the poststratified posterior median value. The thick and thin blue lines represent the 50% and 95% credible intervals, respectively.

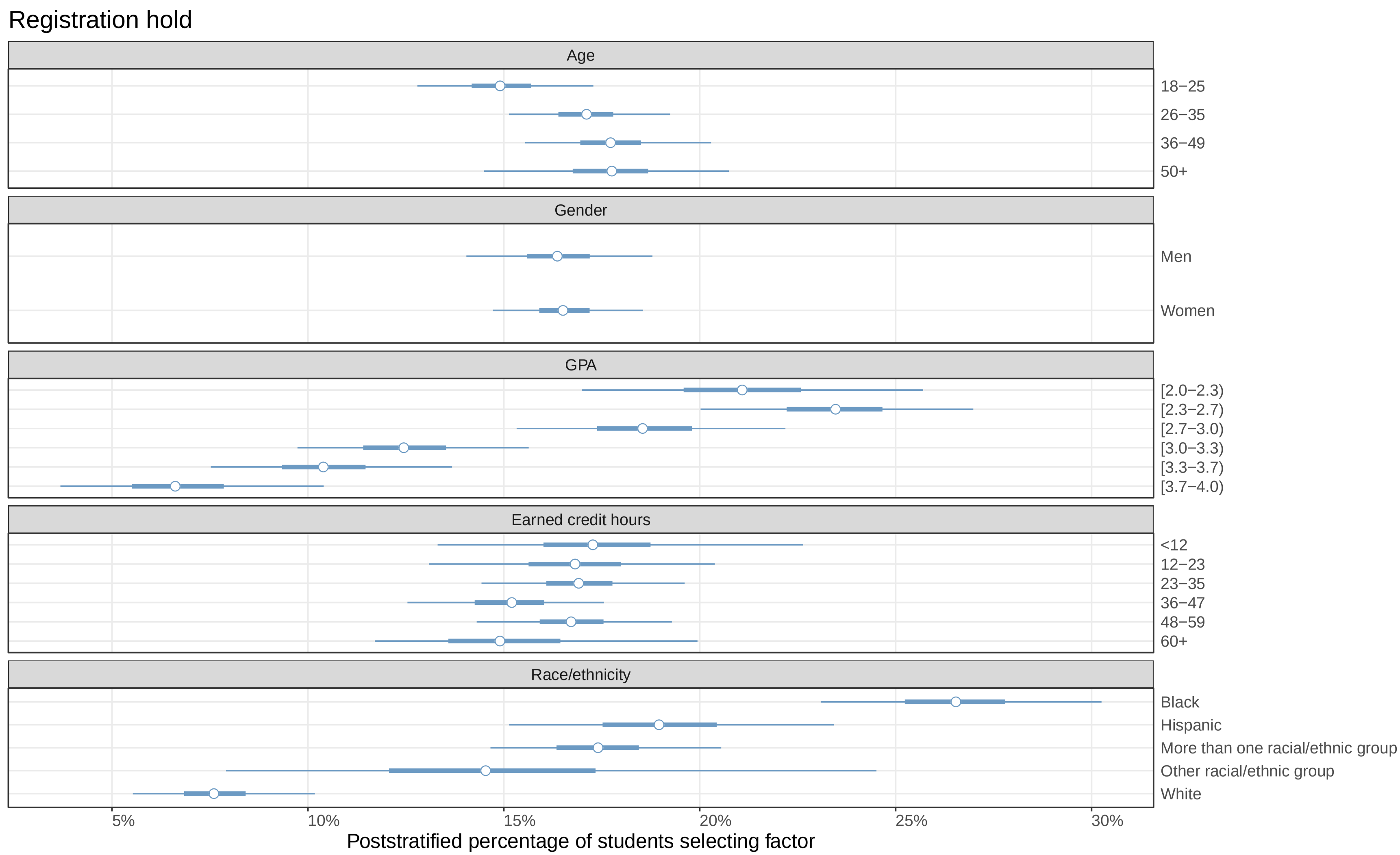

Figure 6: Each facet represents the poststratified posterior results for one question answer (outcome: was the checkbox selected), separated by within-group responses. The open circle represents the poststratified posterior median value. The thick and thin blue lines represent the 50% and 95% credible intervals, respectively.

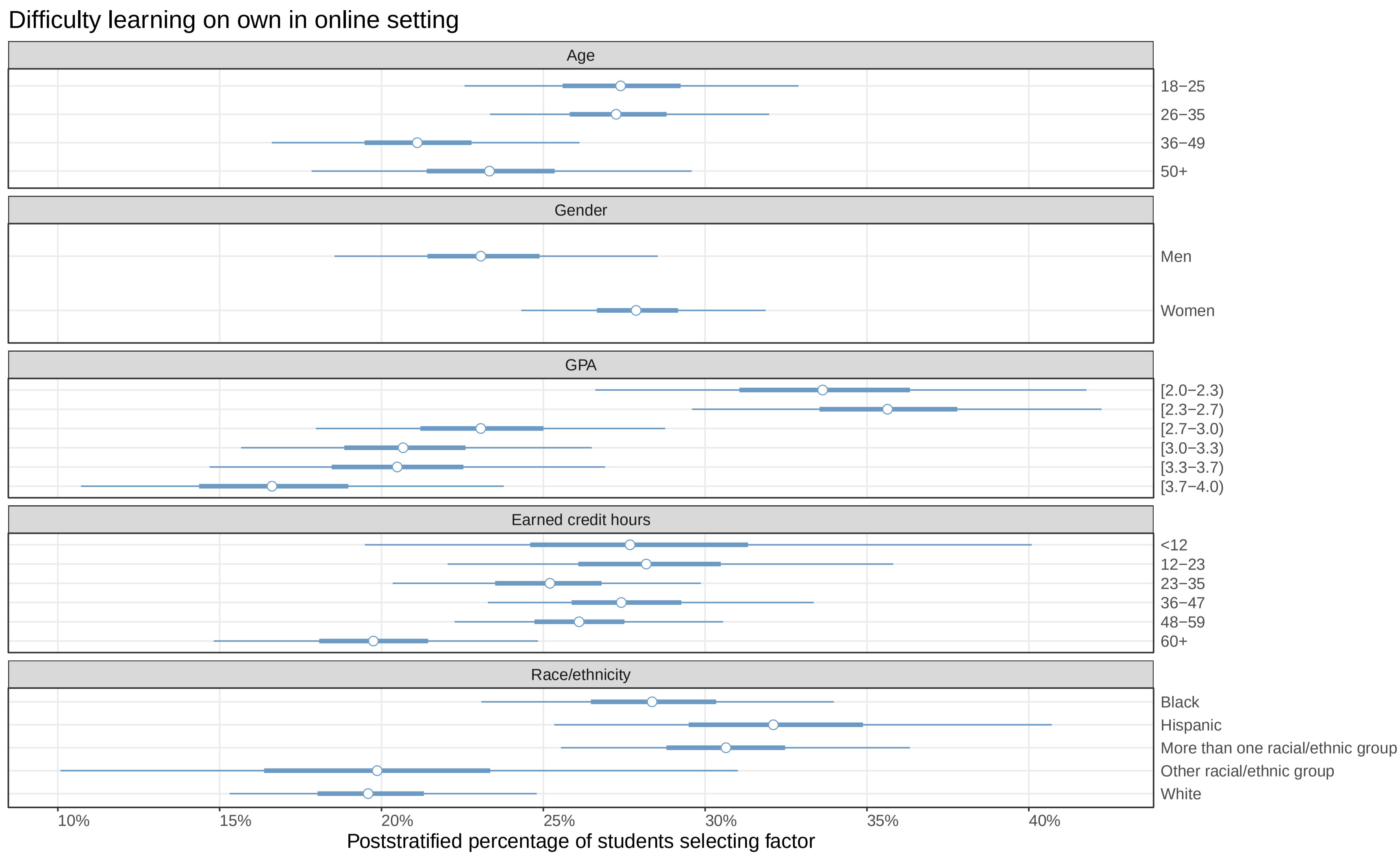

Figure 7: Each facet represents the poststratified posterior results for one question answer (outcome: was the checkbox selected), separated by within-group responses. The open circle represents the poststratified posterior median value. The thick and thin blue lines represent the 50% and 95% credible intervals, respectively.

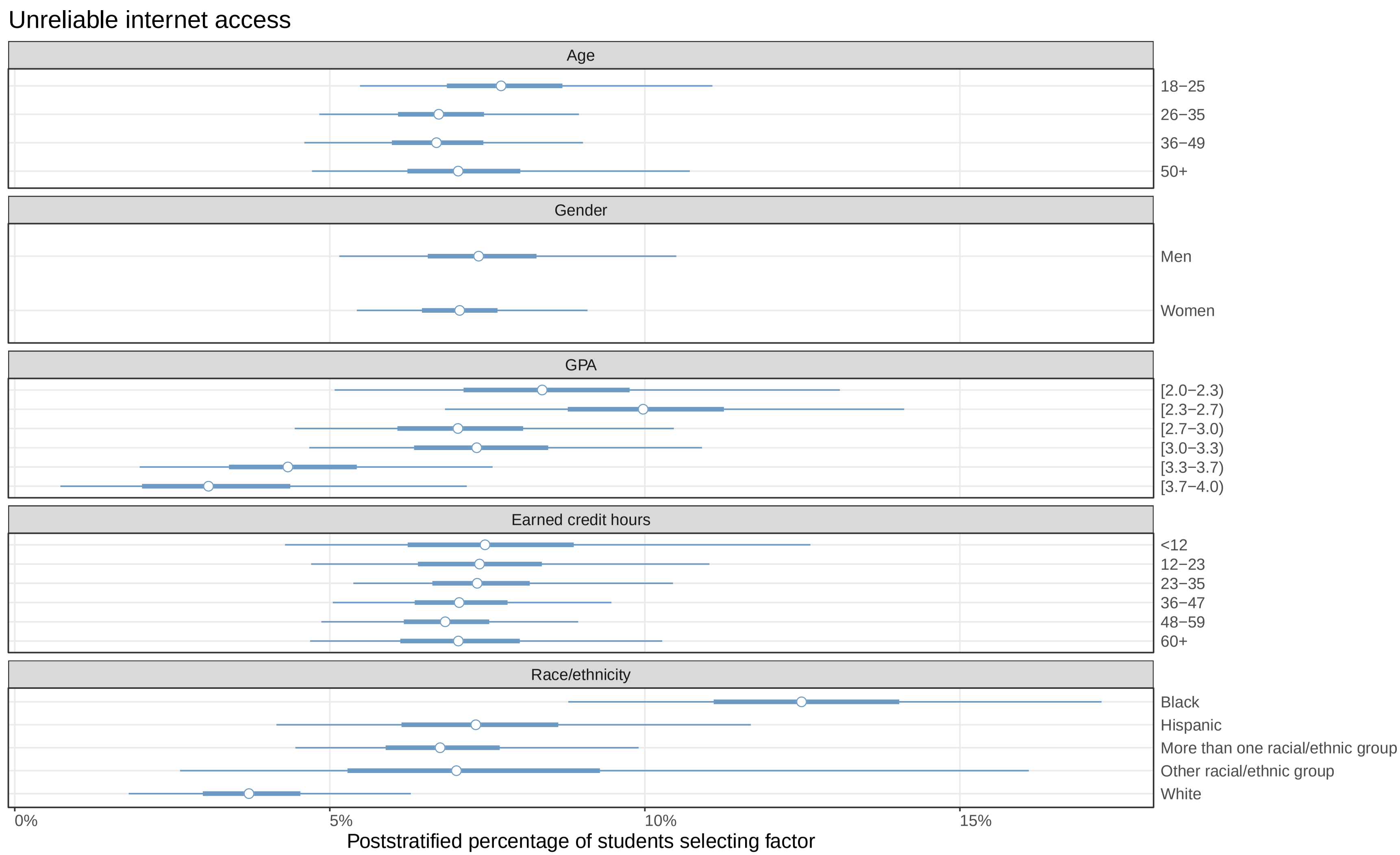

Figure 8: Each facet represents the poststratified posterior results for one question answer (outcome: was the checkbox selected), separated by within-group responses. The open circle represents the poststratified posterior median value. The thick and thin blue lines represent the 50% and 95% credible intervals, respectively.

Stan file

```
// ------------------------------------------------------------
//
// [ PROJ ] Postratify Helios survey responses
// [ FILE ] mrp.stan
// [ AUTH ] Justin Ortagus, Benjamin Skinner, Melvin Tanner
// [ INIT ] 2 March 2020
//
// ------------------------------------------------------------

data {
  int<lower=1> N;     // # of cells
  int<lower=1> M;     // # of schools
  int<lower=1> L;     // # of 2nd-level covariates
  int<lower=1> J_sch; // # of school categories
  int<lower=1> J_gen; // # of gender categories
  int<lower=1> J_rac; // # of race/ethnicity categories
  int<lower=1> J_age; // # of age categories
  int<lower=1> J_gpa; // # of gpa categories
  int<lower=1> J_hrs; // # of hrs categories
  int views[N];       // # of cell members who saw the question
  int clicks[N];      // # of cell members who clicked box
  int sch[N];         // school categories
  int gen[N];         // gender categories
  int rac[N];         // race/ethnicity categories
  int age[N];         // age cat
  int gpa[N];         // gpa cat
  int hrs[N];         // hrs cat
  matrix[M,L] z;      // 2nd-level variables
}
parameters {
  // general intercept
  real beta_0;

  // using non-centered parameterization to help with convergence
  //
  // theta ~ N(mu, sigma) ==> mu + sigma * theta_std where theta_std ~ N(0,1)

  // standardized coefficients for reparametrized parameters
  vector[J_sch] sch_alpha_std;
  vector[J_gen] gen_alpha_std;
  vector[J_rac] rac_alpha_std;
  vector[J_age] age_alpha_std;
  vector[J_gpa] gpa_alpha_std;
  vector[J_hrs] hrs_alpha_std;
  vector[L] beta_std;

  // standard deviations for reparameterized parameters (restricted positive)
  real<lower=0> sch_alpha_sd;
  real<lower=0> gen_alpha_sd;
  real<lower=0> rac_alpha_sd;
  real<lower=0> age_alpha_sd;
  real<lower=0> gpa_alpha_sd;
  real<lower=0> hrs_alpha_sd;
  real<lower=0> beta_sd;
}
transformed parameters {
  // actual model parameters
  vector[J_sch] sch_alpha;
  vector[J_gen] gen_alpha;
  vector[J_rac] rac_alpha;
  vector[J_age] age_alpha;
  vector[J_gpa] gpa_alpha;
  vector[J_hrs] hrs_alpha;
  vector[L] beta;
```

```
  // linear combination
  vector[N] p_hat;

  // if: alpha ~ N(0, alpha_sd)
  // --> 0 + alpha_sd * alpha_std --> alpha_sd * alpha_std
  sch_alpha = sch_alpha_sd * sch_alpha_std;
  gen_alpha = gen_alpha_sd * gen_alpha_std;
  rac_alpha = rac_alpha_sd * rac_alpha_std;
  age_alpha = age_alpha_sd * age_alpha_std;
  gpa_alpha = gpa_alpha_sd * gpa_alpha_std;
  hrs_alpha = hrs_alpha_sd * hrs_alpha_std;
  beta = beta_sd * beta_std;

  // linear combination (vectorized using indices)
  p_hat = beta_0
    + sch_alpha[sch]
    + gen_alpha[gen]
    + rac_alpha[rac]
    + age_alpha[age]
    + gpa_alpha[gpa]
    + hrs_alpha[hrs]
    + z[sch,] * beta;
}
model {
  // standardized coefficient priors: N(0,1)
  sch_alpha_std ~ std_normal();
  gen_alpha_std ~ std_normal();
  rac_alpha_std ~ std_normal();
  age_alpha_std ~ std_normal();
  gpa_alpha_std ~ std_normal();
  hrs_alpha_std ~ std_normal();
  beta_std ~ std_normal();
  beta_0 ~ std_normal();

  // standardized standard deviation priors: N+(0,1) due to restriction above
  sch_alpha_sd ~ std_normal();
  gen_alpha_sd ~ std_normal();
  rac_alpha_sd ~ std_normal();
  age_alpha_sd ~ std_normal();
  gpa_alpha_sd ~ std_normal();
  hrs_alpha_sd ~ std_normal();
  beta_sd ~ std_normal();

  // likelihood (binomial rather than bernoulli b/c we collapse [0/1]
  // observations to the cell level [clicks/views] as sufficient
  // statistic
  clicks ~ binomial_logit(views, p_hat);
}
```

Table A.1: Percentage of students indicating reason for early exit

	Survey mean	Poststratified median
Cost		
Computer and internet access	5.3	5.3
	(4.4, 6.1)	[4.3, 6.5]
Difficulty completing financial aid application	11.4	11.9
	(10.2, 12.7)	[10.4, 13.5]
Employer stopped paying for classes	3.6	3.3
	(2.8, 4.3)	[2.5, 4.3]
Hard to access financial aid information online	9	9.2
	(7.8, 10.1)	[7.9, 10.8]
Living expenses (rent, utilities, healthcare, childcare, food)	49.8	51.3
	(47.8, 51.8)	[48.5, 53.9]
Missed payment deadline and was dropped	10.7	10.4
	(9.4, 11.9)	[9, 11.9]
No longer eligible for financial aid	42.9	42
	(40.9, 44.9)	[39.5, 44.5]
Scholarship ran out	5.1	5.4
	(4.2, 6)	[4.4, 6.5]
Textbooks	23.4	24.8
	(21.8, 25.1)	[22.8, 26.9]
Transportation to campus	8.4	9
	(7.3, 9.5)	[7.8, 10.5]
Tuition and fees	53.5	52.2
	(51.5, 55.5)	[49.7, 54.5]
Employment		
Changed careers	15.4	16.7
	(14, 16.9)	[14.8, 18.8]
Got a promotion	12.4	13.5
	(11.1, 13.7)	[11.6, 15.5]
Got second job	14.2	15.2
	(12.8, 15.6)	[13.2, 17.5]
Lost job	12.3	11.9
	(10.9, 13.6)	[10.3, 13.6]
Switched from part-time to full-time work	32.1	36.2
	(30.3, 34)	[33.2, 39.2]
Instructional		
Didn't know which classes to take next	24.2	25.6
	(22.5, 25.9)	[23.3, 28]
Difficulty learning on own in online setting	25.8	25.9
	(24.1, 27.5)	[22.9, 29.3]
Instructor couldn't meet outside of class	5.1	5.7
	(4.2, 5.9)	[4.6, 6.9]
Needed course unavailable online	19.6	20.3
	(18.1, 21.2)	[18.3, 22.3]
Needed course was full	13.4	13.9
	(12, 14.7)	[12.3, 15.7]
Needed course was only online	5.8	6.2
	(4.9, 6.8)	[5.1, 7.5]
No time to study/prepare for class	34.6	37.1

Continued on next page...

...table A.1 continued

	Survey mean	Poststratified median
	(32.7, 36.5)	[34.5, 39.7]
Not enough interaction with online instructor	24.2	24.9
	(22.5, 25.9)	[22, 28.3]
Not enough interaction with students in online setting	7.7	8.4
	(6.6, 8.8)	[6.7, 10.5]
Registration hold	17.5	16.4
	(16, 19)	[14.8, 18.2]
Required math and science courses too difficult	27	25.3
	(25.3, 28.8)	[23.3, 27.5]
Struggled completing assignments	22.5	23.7
	(20.8, 24.1)	[21.6, 25.9]
Struggled with online course software	13.1	14.1
	(11.7, 14.4)	[11.7, 16.9]
Too many required developmental/remedial courses	15.4	14.8
	(13.9, 16.8)	[13.2, 16.6]
Unreliable internet access	7.1	7.2
	(6, 8.1)	[5.6, 9.1]
Other		
Change in relationship status	13.5	13.4
	(12.1, 14.8)	[11.8, 15.2]
Desired classes not at closest campus	25	25.6
	(23.2, 26.7)	[23.5, 27.7]
Did not feel welcome on campus	8.4	9.4
	(7.3, 9.5)	[7.9, 11.2]
Did not have many friends at the college	9.3	11.6
	(8.2, 10.5)	[9.9, 13.7]
Did not have reliable childcare	12.8	12.2
	(11.5, 14.2)	[10.6, 14.3]
Health emergency	17.3	15.8
	(15.8, 18.8)	[14.1, 17.8]
Inconsistent weekly schedule	32.7	33.3
	(30.8, 34.5)	[31.2, 35.7]
Job was too far from campus	13.6	14.1
	(12.2, 15)	[12.5, 15.8]
Moved out of the area	12.3	13.9
	(11, 13.6)	[12.2, 15.8]
Transportation/parking was difficult	10.5	11.1
	(9.3, 11.7)	[9.7, 12.9]

Notes. Survey mean estimates are the number of clickes (positive selections) divided by total views. The 95% confidence intervals for the means are reported in parentheses. Bayesian point estimates are poststratified medians, with 95% credible intervals reported in square brackets.

Table A.2: Question number and name concordance

Question number	Question name
Question 3.1	Cost: textbooks
Question 3.2	Cost: tuition and fees
Question 3.3	Cost: computer and internet access
Question 3.4	Cost: transportation to campus
Question 3.5	Cost: living expenses (rent, utilities, healthcare, childcare, food)
Question 4	Employed while taking classes?
Question 5.1	Employment: lost job
Question 5.2	Employment: changed careers
Question 5.3	Employment: switched from part-time to full-time work
Question 5.4	Employment: got second job
Question 5.5	Employment: got a promotion
Question 6.1	Cost: difficulty completing financial aid application
Question 6.2	Cost: no longer eligible for financial aid
Question 6.3	Cost: scholarship ran out
Question 6.4	Cost: missed payment deadline and was dropped
Question 6.5	Cost: employer stopped paying for classes
Question 6.6	Cost: hard to access financial aid information online
Question 7.1	Instructional: registration hold
Question 7.2	Instructional: didn't know which classes to take next
Question 7.3	Instructional: needed course unavailable online
Question 7.4	Instructional: needed course was full
Question 7.5	Instructional: needed course was only online
Question 8.1	Other: moved out of the area
Question 8.2	Other: inconsistent weekly schedule
Question 8.3	Other: desired classes not at closest campus
Question 8.4	Other: transportation/parking was difficult
Question 8.5	Other: job was too far from campus
Question 9.1	Instructional: no time to study/prepare for class
Question 9.2	Instructional: struggled completing assignments
Question 9.3	Instructional: required math and science courses too difficult
Question 9.4	Instructional: too many required developmental/remedial courses
Question 9.5	Instructional: instructor couldn't meet outside of class
Question 10	Did you take any online courses?
Question 11.1	Instructional: not enough interaction with online instructor
Question 11.2	Instructional: unreliable internet access
Question 11.3	Instructional: not enough interaction with students in online setting
Question 11.4	Instructional: difficulty learning on own in online setting
Question 11.5	Instructional: struggled with online course software
Question 12.1	Other: change in relationship status
Question 12.2	Other: health emergency
Question 12.3	Other: did not feel welcome on campus
Question 12.4	Other: did not have reliable childcare
Question 12.5	Other: did not have many friends at the college

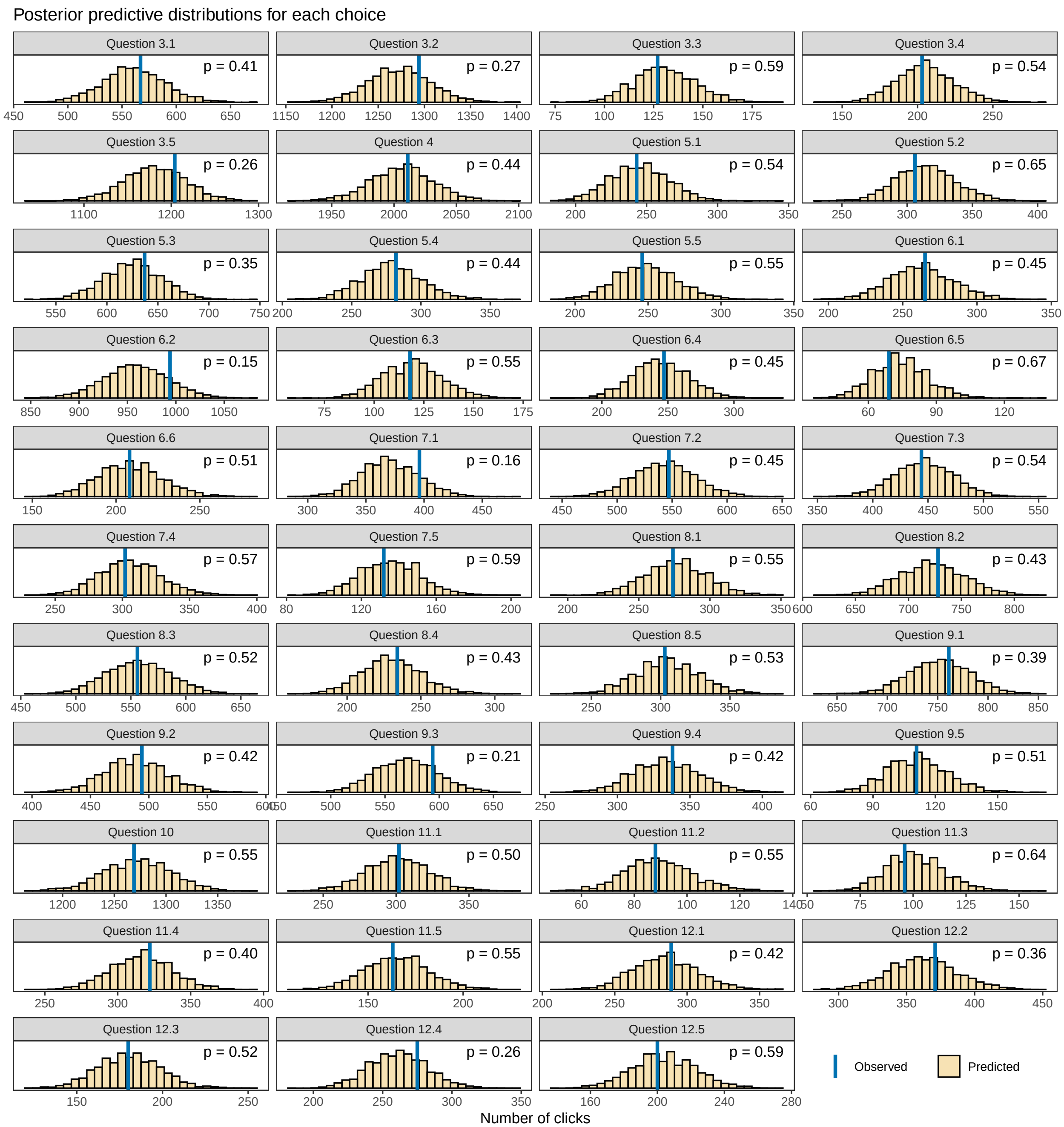

Figure A.1: Posterior predictive distributions (histogram) with observed counts of students who clicked an option (vertical line), each facet representing a unique survey choice and model. Bayesian p-values show $Pr(T(y^{rep},\theta) \geq T(y,\theta) \mid y)$, the proportion of simulated draws (4,000 for each facet) greater than or equal to the observed count. A concordance for question number and name can be found in Table A.2.