Characteristics of an Online Consumer Genetic Research Cohort

J.Y. Tung, N. Eriksson, A.K. Kiefer, J.M. Macpherson, B.T. Naughton, A.B. Chowdry, C.B. Do, D.A. Hinds, A. Wojcicki, J.L. Mountain

23andMe, Inc, Mountain View, CA

Abstract

The increasing availability of internet access and the decreasing costs of generating genetic information have created a new group of individuals who are active internet users with access to their own genetic information. We hypothesized that these individuals would be interested in sharing their genotypic and phenotypic information for research if given the opportunity to do so by providing data online, thus removing any geographic restriction for participating in a study. All customers of 23andMe, a direct-to-consumer genetic testing company, were recruited for an ongoing research project entitled 23andWe. We evaluated consent rate, demographics (including sex, age, education, household income, and genetic ancestry), and research questionnaire response rate in this cohort. Close to 90% of customers consented to participate in 23andWe, and of those, again close to 90% of participants answered at least one research question, with an average of 11 questionnaires completed, and 64 singular "Research Snippet" questions answered. Despite the fact that the majority of customers did not obtain the 23andMe service in order to participate in research, we have found that most are willing to contribute their data for research and that the vast majority actively contribute additional phenotypic data online, suggesting that tapping into consumer interest in genetics and research is a viable method for conducting large-scale studies.

Methods

Data on sex, age, education, household income, genetic ancestry, and number of survey or Research Snippet questions answered were pulled from the 23andMe research database on or before September 28, 2011, and only from individuals who have consented to participate in research. All data were either self-reported, except for sex and ancestry, which were inferred from genotype data.

Results

Despite the fact that less than 10% of 23andMe customers were recruited through specific research projects (such as projects that focus on Parkinson's disease or sarcoma), most customers consented to participate in our research (for the latest version of our consent form, visit https://www.23andme.com/about/consent/) (Figure 1).

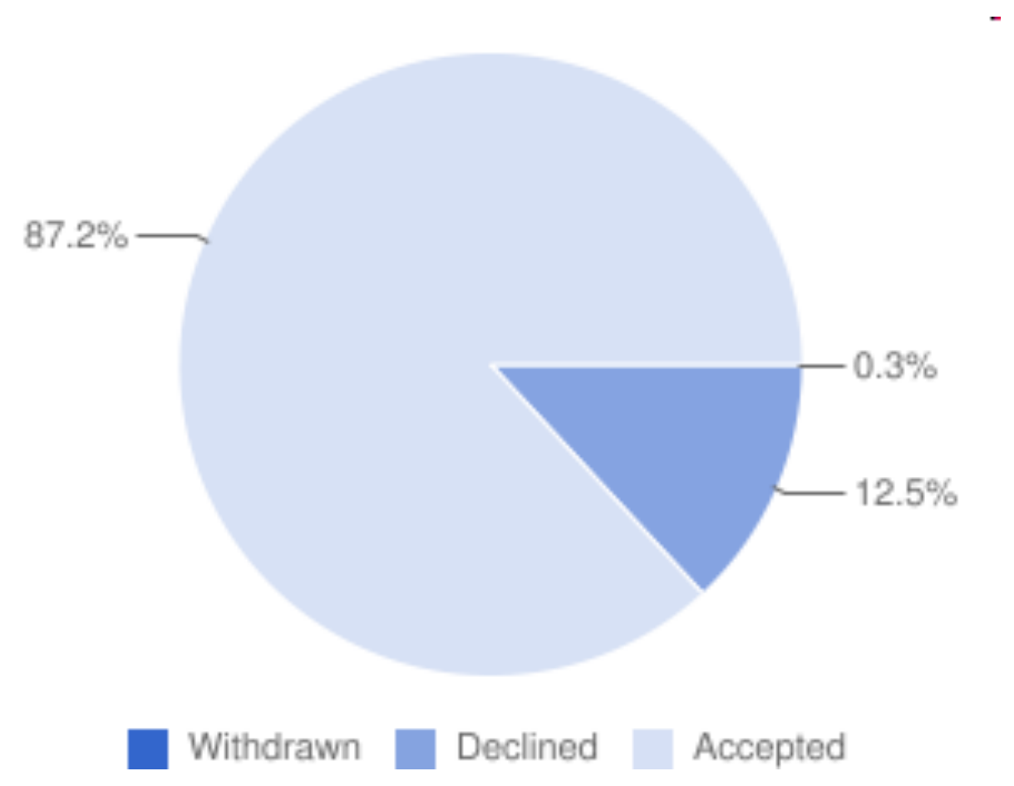

Figure 1. Percentage of 23andMe customers who have accepted, declined, or withdrawn (accepted at first, but later declined or closed their account) consent to participate in 23andWe research.

Figure 2 depicts the demographics of our research cohort. Men slightly outnumber women, and the age distribution is bimodal centering approximately around 30 and 60 years. Most of our participants have at least some college education and have a household income of at least $100,000 per year. The vast majority are of European descent, although recently the percentage of individuals of African descent in our database has risen rapidly through the efforts of our Roots into the Future initiative (https://www.23andme.com/roots/).

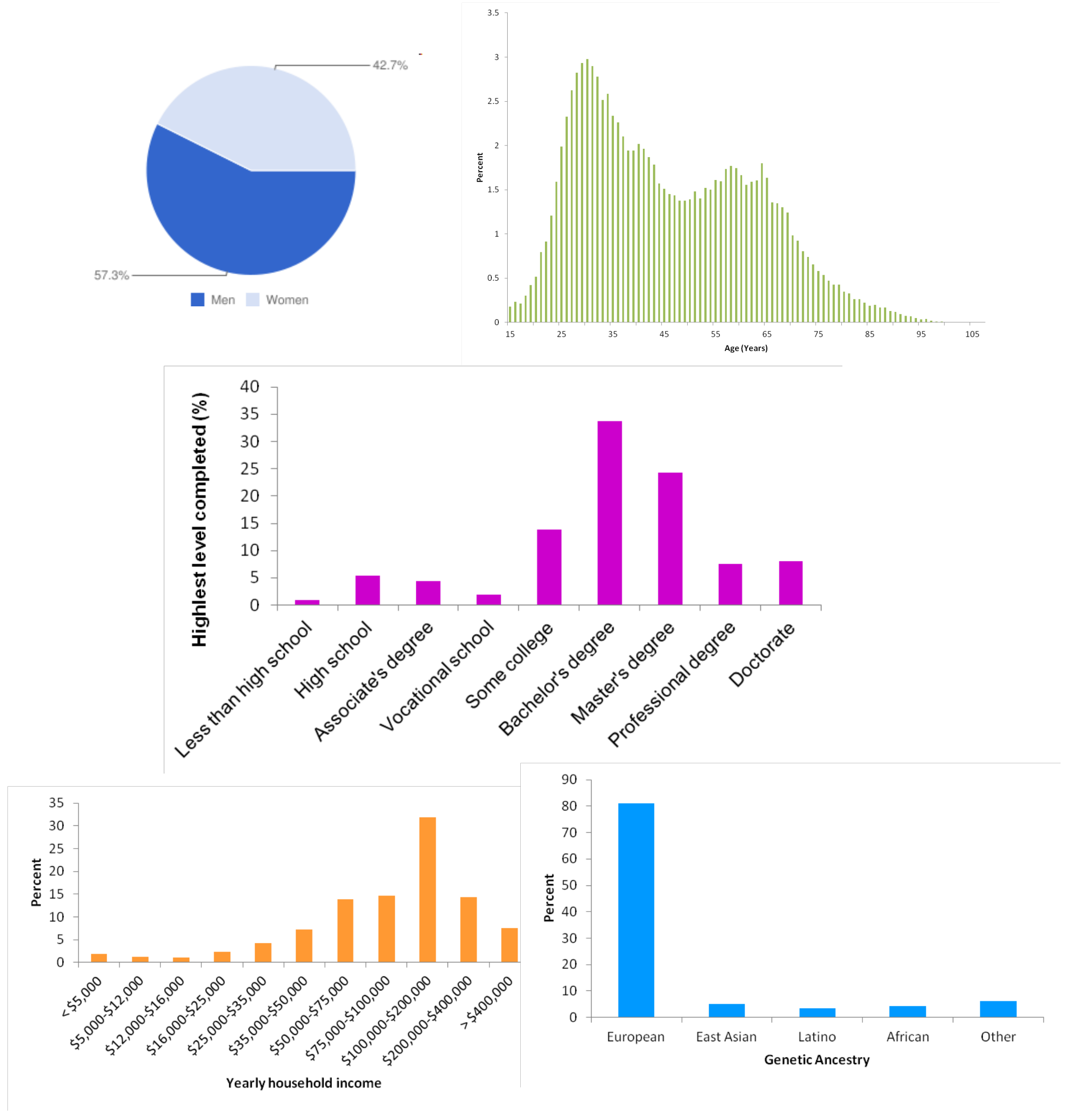

Figure 2. Breakdown of 23andMe research cohort by sex, age, highest level of education, household income, and genetic ancestry. Approximately 10% of respondents reported education and household income.

Approximately 88% of the our research participants have provided phenotype data about themselves through a research survey (generally multi-question questionnaires focusing on a particular topic) or a Research Snippet (single, one-off questions presented in random order). Figure 3 shows the distribution of the number of surveys and Research Snippets answered by our research participants. The average number of surveys taken was just over 11 (of those who took at least one survey), while the average number of Research Snippets answered was over 60 (of those who answered at least one Research Snippet). Figure 4 shows the breakdown of average number of surveys and Research Snippets answered by sex, age, and genetic ancestry (educational status and household income are not included as relatively small numbers of people answered those questions).

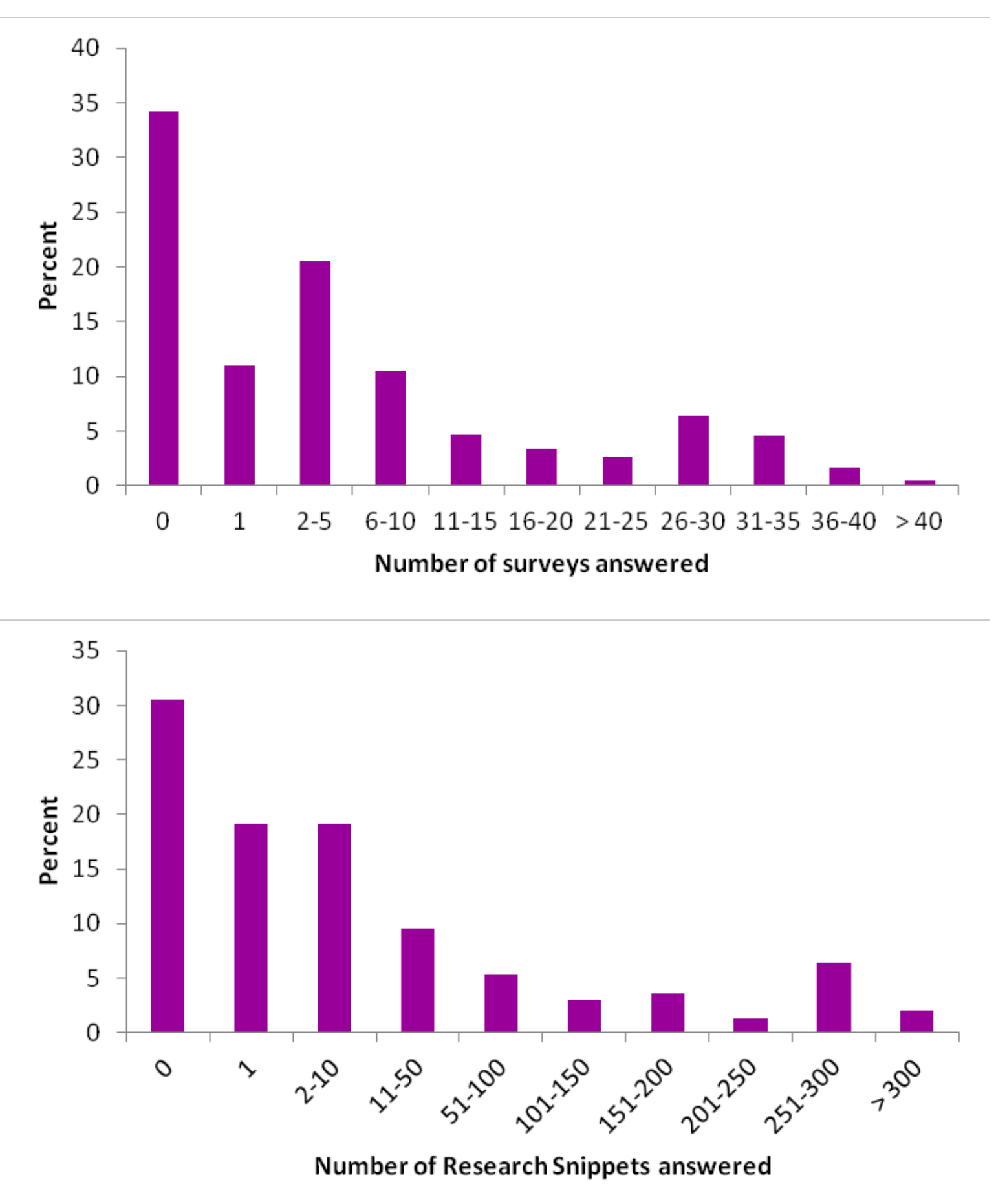

Figure 3. Number of surveys and Research Snippets answered

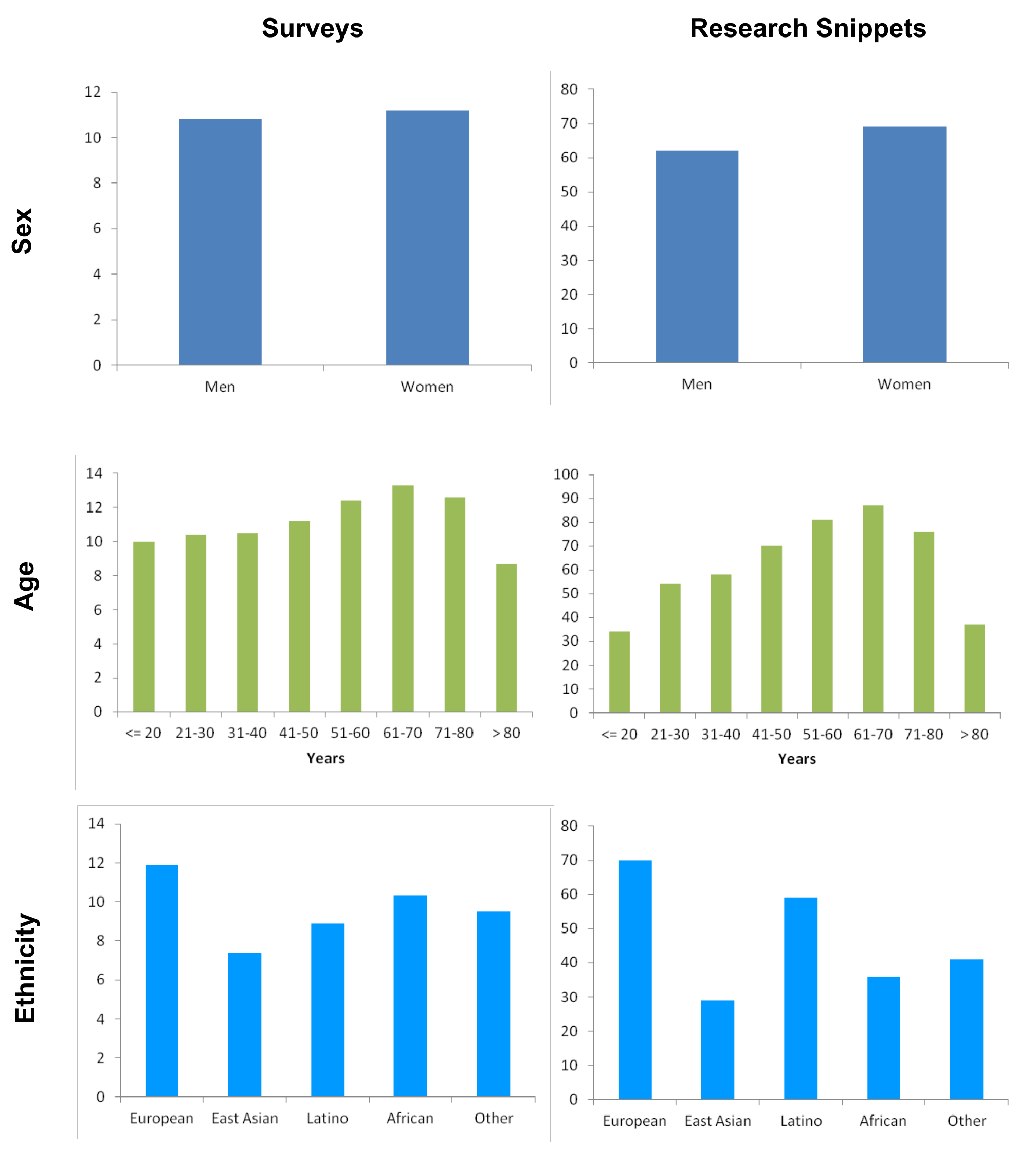

Figure 4. Average number of surveys and Research Snippets answered broken down by sex, age, and genetic ancestry (of those who took at least one survey or at least one Research Snippet, respectively).

Discussion

Of all 23andMe customers, close to 90% consent to participate in research. Of those, close to 90% choose to contribute additional phenotypic data by answering research questions. This cohort has already been employed in a number of publications[1, 2, 3]. For many reasons, this cohort may not be representative of, for example, the U.S. population; however, this is mainly an issue for genome-wide association studies only when there is genetic variation underlying the demographic differences that cannot be controlled for. Many genome-wide association studies also employ non-representative populations for reasons of practicality. The high rate of participation in this group suggests that tapping into consumer interest in research is a viable method for conducting large-scale studies.

Acknowledgments

We thank the customers of 23andMe for their support and time spent contributing to this research. Without their enthusiasm, there would be no research. We would also like to thank the employees of 23andMe for making such an ambitious research project technically possible.

References

1. Eriksson et al. (2010). "Web-based, participant-driven studies yield novel genetic associations for common traits." PLoS Genet 6(6):e1000993.

2. Do et al. (2011). "Web-based genome-wide association study identified two novel loci and a substantial genetic component for Parkinson's disease." PLoS Genet 7 (6):e1002141.

3. Tung et al. (2011). "Efficient replication of over 180 genetic associations with self-reported medical data." PLoS ONE 6 (8):e23473.