

Article

Avera Twin Register Growing Through Online Consenting and Survey Collection

Julie M. Kittelsrud¹, Erik A. Ehli¹, Vikki Petersen¹, Tammy Jung¹, Jeffrey J. Beck¹, Noah Kallsen¹, Patricia Huizenga¹, Brittany Holm¹ and Gareth E. Davies^{1,2}

¹Avera Institute for Human Genetics, Sioux Falls, SD, USA and ²Netherlands Twin Register, Department of Biological Psychology, Vrije Universiteit, Amsterdam, the Netherlands

Abstract

The aim of the Avera Twin Register (ATR) is to establish a prospective longitudinal repository of twins, multiples, siblings and family members' biological samples to study environmental and genetic influences on health and disease. Also, it is our intention to contribute to international genome-wide association study (GWAS) twin consortia when appropriate sample size is achieved within the ATR. The ATR is young compared with existing registers and continues to collect a longitudinal repository of biological specimens, survey data and health information. Data and biological specimens were originally collected via face-to-face appointments or the postal department and consisted of paper-informed consents and questionnaires. Enrollment of the ATR began on May 18, 2016 and is located in Sioux Falls, South Dakota, a rural and frontier area in the Central United States with a regional population of approximately 880,000. The original target area for the ATR was South Dakota and the four surrounding states: Minnesota, Iowa, North Dakota and Nebraska. The ATR has found a need to expand that area based on twin and multiple siblings who live in various areas surrounding these states. A description of the state of the ATR today and its transition to online data collection and informed consent will be presented. The ATR collects longitudinal data on lifestyle, including diet and activity levels, aging, plus complex traits and diseases. All twins and multiples participating in the ATR are genotyped on the Illumina Global Screening Array and receive zygosity results.

Keywords: Twins; multiples; twin register; lifestyle; genotyping; Global Screening Array

(Received 31 May 2019; accepted 27 June 2019; First Published online 14 October 2019)

Register Background

The Avera Twin Register (ATR) was initiated by the Avera Health system, including clinicians and researchers within the Midwestern United States (Kittelsrud et al., 2017). The ATR is relatively young in comparison with other registers worldwide. It was established in 2016 with the launch of a media campaign and announcement of the official partnership with the Netherlands Twin Register (NTR), Vrije Universiteit, Amsterdam (Willemsen et al., 2013). The Midwestern United States is heavily populated by descendants from The Netherlands and surrounding European communities. Thus, genetic and environmental comparisons of the Dutch population may provide a better understanding of health and wellness (Bryc et al., 2015). The aim of the ATR is the establishment of a longitudinal prospective repository of genetic samples and data of twins and family members to understand genetic and environmental influences of health and disease in this region. It is a goal of the ATR to contribute regional information to the wider twin research consortiums. Ultimately, Avera hopes to contribute to clinical practice by improving disease prevention and diagnostic and treatment tools.

The Avera Health system is based in Sioux Falls, South Dakota, covering a service area of 71,550 square miles (185,313 km). Sioux Falls, which is the largest city in South Dakota, has a population of about 178,000 (United States Census Bureau, 2018). The health system includes the surrounding states of Minnesota, Iowa, Nebraska and North Dakota, with clinics and hospitals servicing a population close to one million people. Avera Health is a vertically integrated health care delivery network and is sponsored by the Sisters of St. Benedict of Sacred Heart Monastery of Yankton, South Dakota, and the Sisters of the Presentation of the Blessed Virgin Mary of Aberdeen, South Dakota. The innovative health system established the Avera Institute for Human Genetics (AIHG), a molecular genetics laboratory having the capacity to conduct large high-throughput GWAS. The AIHG is certified by Clinical Laboratory Improvement Amendments and College of American Pathologists. It provides complete personalized pharmacogenomics services through the product name GeneFolio® (Avera Institute for Human Genetics). The AIHG also processes the genetic materials for all ATR participants and conducts zygosity testing.

Recruitment is through the Avera Health system and the snowball effect from twin mothers' groups and various educational opportunities. The ATR is enrolling multiples and their family members. The initial focus was to include twins and multiples focusing on South Dakota and the bordering states of North

Author for correspondence: Julie M. Kittelsrud, Email: Julie.Kittelsrud@Avera.org

Cite this article: Kittelsrud JM, Ehli EA, Petersen V, Jung T, Beck JJ, Kallsen N, Huizenga P, Holm B, and Davies GE. (2019) Avera Twin Register Growing Through Online Consenting and Survey Collection. *Twin Research and Human Genetics* 22: 686–690, <https://doi.org/10.1017/thg.2019.73>

Table 1. Benefits and challenges of online surveys

Benefits	Challenges
• Reach larger number of participants	• High nonresponse rates
• Keep costs low	• Reliability of data (i.e., height, weight)
• Data entered by participant reduce staff time	• Email addresses change frequently
• Logic within questionnaire individualizes data collection (e.g., singletons vs. twins)	• Limitation of question types (no open-ended questions, no interviewer)
• Rapid distribution, return and updates can be made	• Restricts populations (age, location, availability of online access)
• Automation (i.e., email reminders, follow-up)	• Loss of personal contact
• Convenience for respondents	

Dakota, Iowa, Nebraska and Minnesota. It has now expanded its enrollment to twins and multiples throughout the USA. Until this year, the ATR has conducted in-person visits, telephone visits, including paper-and-pencil informed consent, sent via postal services, questionnaires and DNA collection via buccal cell collection.

A growing list of advantages to collecting data via the web has contributed to the ATR's move to online consenting and questionnaires. The Research Electronic Data Capture (REDCap) software toolset was chosen for this project and is hosted within the Avera Health system. REDCap is a secure, web-based application designed to support data capture for research studies, providing (1) an intuitive interface for validated data entry, (2) audit trails for tracking data manipulation and export procedures, (3) automated export procedures for seamless data downloads to common statistical packages, and (4) procedures for importing data from external sources (Harris et al., 2009).

Conducting research via the internet has increased in popularity because of the growing numbers of online users (Dillman et al., 1998). Researchers have noted multiple benefits and challenges of conducting surveys online (Table 1). This article will showcase the ATR's transition to online questionnaires, tracking and study management.

Aims of the ATR

The ATR began with three objectives for its establishment (Kittelsrud et al., 2017). The first objective is to formulate associations with the community to enroll a prospective longitudinal repository of biological specimens, with survey data, including health information and habits from all ages of twins or multiples, and their family members. Second, the ATR seeks to study environmental and genetic influences on health, illness prevention, behavior and disease in the ATR population through collaborations with the NTR and twin consortium projects. Third, the ATR seeks to serve the needs of twins and twin families by addressing questions that are specific to twins, such as those related to their zygosity, and unique social and health goals (Segal, 2017).

Materials and Methods

This study was approved by the Avera Institutional Review Board and the Avera Department of Human Subject's protection. Initial study commencement was on May 2016 after a press release

Table 2. Enrollment data collected in 2016 to present via questionnaire (baseline)

Topic	Data included
Demographics	Participant contact information, date of birth, family relationship to multiples (parent, sibling, grandparents)
Multiple specific information	Zygoty (GSA data), similarities between multiples (spitting image, confused by parents, strangers or family)
Parent information	Location of parents' birth, birth date, health history
Personal information	Race, who gave birth to multiples, marital status, handedness, relationship status, income, education, occupation
Birth information	Was fertility assistance needed, birth weight, birth length, gestational age
Health history	System review of health history (current or past health problems), rating of health, and best life (0–10), sleep (number of hours per night).
Neuroticism-Extraversion-Openness Inventory personal analysis	60 question personal analysis
Substance use	Alcohol use (amount, frequency), smoking status (number/day)
Women's health	Fertility issues, pregnancy history, delivery complications, pregnancy complications, postpartum depression, smoking status during pregnancy
Physical activity	Activity name, frequency, intensity, duration, use of a fitness tracking device
Child development	School activities, grade level, best subject, reading difficulty/dyslexia
Adolescent female	Menstruation age
Adolescent male	Hair development (arms/beard), voice change

conference. No public advertising has been conducted to recruit participants apart from a public education event or twin event. All questionnaire materials were paper and pencil, and biological sampling was buccal cell collection via a swab passed over the cheek. One major difference between the ATR and other established registers is that all participants donate biological samples when they join the register. We have not had to go back and collect samples for specific populations in retrospect. This will potentially allow for more family genetic analysis in relationship to twin studies.

The previous methodology included in-person visits to the AIHG facility for face-to-face consenting, paper questionnaires and buccal cell collection for those living in or around Sioux Falls, South Dakota. Those who lived outside of the local area were invited to schedule an appointment to consent via telephone. After telephone consent was completed and signed paper consents were received at the AIHG, kits were mailed to the participants. The kits contained instructions, buccal cell collection vials, paper surveys, and copies of signed consent forms.

Transition to Online Surveys

In January 2019, the ATR transitioned to the use of online consenting, surveys and tracking developed through REDCap™.

Fig. 1. Matrix example question from the participants view. Sample question about facial shape similarity for triplets 'Jonah', 'Josh', and 'Jorah'.

REDCAP number	Event Name	child1_face	child2_face	child3_face
1	1forms (Arm 1: Arm 1)		Somewhat	Somewhat
2	1forms (Arm 1: Arm 1)	Somewhat	Somewhat	
4	1forms (Arm 1: Arm 1)	Exactly	Exactly	
5	1forms (Arm 1: Arm 1)		Somewhat	Somewhat
6	1forms (Arm 1: Arm 1)	Somewhat	Somewhat	
8	1forms (Arm 1: Arm 1)	Somewhat	Somewhat	
9	1forms (Arm 1: Arm 1)			
11	1forms (Arm 1: Arm 1)	Not at all	Somewhat	Somewhat
12	1forms (Arm 1: Arm 1)		Somewhat	Somewhat
13	1forms (Arm 1: Arm 1)	Somewhat	Somewhat	
15	1forms (Arm 1: Arm 1)		Somewhat	Somewhat
16	1forms (Arm 1: Arm 1)			

Fig. 2. Matrix example of data output. It becomes unclear which twin or multiple is child 1, child 2 and so on. Therefore, the output requires inspection of individual questionnaires.

Study data was collected and managed using REDCap electronic data capture tools hosted at Avera Institute for Human Genetics (Harris *et al.*, 2009, 2019). REDCap (Research Electronic Data Capture) was initially developed and deployed at Vanderbilt University, but now has support from a wide consortium of domestic and international partners. It is a secure, web-based software platform designed to support data capture for research studies, providing (1) an intuitive interface for validated data capture; (2) audit trails for tracking data manipulation and export procedures; (3) automated export procedures for seamless data downloads to common statistical packages; and (4) procedures for data integration and interoperability with external sources. The online consent and surveys were altered as minimally as possible to retain data integrity between paper and online forms. Enrollment via the online tool was piloted with a small sample of participants who enrolled on the ATR website. Once adult participants had consented for themselves, they were asked if they had children to enroll. Branching logic within REDCap™ allowed the ability for the participant to proceed directly to the child consent and surveys for themselves and their children under 18 years old. The United States Code of Federal Regulations indicates that consent for a child under 18 must be completed by a parent or legal guardian. The initial design allowed participants to complete surveys contiguously for multiple children, including singleton siblings and twins and multiples. Based on our experiences with this format, we decided to switch to a two-step enrollment and survey process.

The online survey was then revised, and the revision became active in March of 2019 on a limited scale. The revisions included

a two-step process. The first step involved a registration form that would be completed for each participant, adult participants and children. The registration forms allowed for the differentiation of each participant who was enrolling in the register, while maintaining the connection with the family unit. After registration, the participants are sent a specific link to each person's questionnaire. For example, a parent of twins would first complete three contiguous registration forms, and then they would be emailed a survey link for herself, one for twin A and one for twin B. Similar to the paper survey and in-person process, participants are mailed a sample kit after consent and surveys are completed online. The sample kit includes a buccal cell collection kit with specific collection instructions, a welcome letter, and return postage to send the samples back to the laboratory. Survey topics are listed in Table 2. The survey topics remained unaltered in the transition to web format.

Whole Genome Single-Nucleotide Polymorphism Typing and Zygosity Assessment

All ATR participants who give consent for DNA collection and genotyping upon registration are asked to provide a DNA sample. Genotyping and subsequent zygosity assessments are completed utilizing buccal cell-derived DNA. The AIHG joined the Infinium Global Screening Array (GSA) Consortium initiated by Illumina, Inc. in 2016, to gain access to a customized array containing highly predictive markers for translational research applications. In conjunction with collaborators at the NTR, the GSA array was designed to enhance imputation with a core backbone of approximately 660,000 markers. The GSA also includes additional fine mapping content (~30,000 markers) and other custom add-on markers of interest concerning drug metabolism, fertility, twinning and markers from the GWAS catalog. Initial validation of the array content for imputation was assessed by checking concordance, coverage and imputation quality using the Genome of the Netherlands reference sequence. A database of single-nucleotide polymorphism genotypes for each ATR participant will be included in potential GWASs.

Results of the Online Pilot Initiation (January 2019)

Twelve families enrolled via the ATR website, which included a total of 40 participants, 12 adults (one identifying as a twin) and 28 children (nine pairs of twins, seven singletons and one set of triplets). One adult did not complete enrollment, indicating the wish to enroll herself and twin children younger than 18, but did not complete consent or fill out all surveys. Three follow-up emails were sent to the adult with no response in 2 months. There was additional confusion when participants were allowed

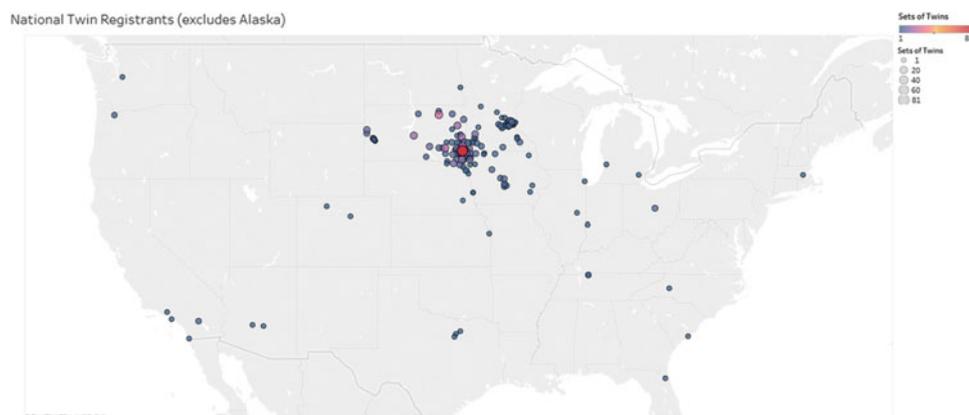


Fig. 3. A map depicting the geographic location of participants enrolled in the ATR based on zip codes reported during enrollment. The size of the circles indicates the number of twins from the respective zip code (see legend top right). One participant in Alaska was excluded from the map.

to go to a website and ‘sign up’. Many participants assumed that the ATR was not interested in nontwin siblings, parents of twins/multiples or other family members. Parents, who were singletons themselves, tended to enroll only their twin/multiple children, and withheld enrollment of themselves and any nontwin siblings. When staff identified these cases, follow-up emails or telephone calls were made to clarify. This was when the online redesign began again.

In the transition of question formats from paper to web, it was noted that all formats are not created equal. The transition of some questions could not retain the exact wording without losing meaning. Some questions that appear to be one question on paper needed to be simplified online to more than one question. This issue led our team to eliminate matrix question types in the second online version, as shown in Figure 1. This question format is one of the most difficult to answer and may result in nonresponse (Dillman & Smyth, 2007; Dillman et al., 1998). The second redesign sought to reduce manual review of data elements when the download was needed from the REDCap™ system. The misunderstanding of who was eligible to enroll in the register was clarified by the addition of a welcome page on the consent and questions that simplifies who is eligible for the study. We are yet to determine whether this has helped reduce confusion.

Current State of the Register

To date, the ATR has enrolled a total of 838 participants, including 492 individuals identifying as a twin or multiple ($n = 232$ twin-pairs, $n = 9$ triplet sets) (Figure 3). Seventeen of the twins also identify as a parent of twins. The register population includes ages 4 months to 91 years old ($M = 27.4$, $SD = 22.1$). Zygosity testing has shown that 131 twin sets are monozygotic (MZ), with 39 sets of male MZ twin-pairs and 92 sets of female MZ twin-pairs. Forty twin sets are dizygotic (DZ), with 31 same-sex DZ and 9 opposite-sex DZ twin-pairs. There are nine sets of triplets enrolled ($n = 27$). One triplet set is identical, two triplet sets have an MZ pair with a DZ child, and six are all DZ. Family members have identified themselves as parents (23.7%), siblings (9.1%), children of twins (3.4%) and grandparents and aunts (1.0%). Thus far, 99% of enrolled participants have requested to be contacted for future studies. The majority of the participants are located in the Midwestern United States, from South Dakota, Minnesota and Iowa. Many other states are represented, but with much lower numbers of participants, most of whom lived at one time or another in South Dakota, Minnesota or Iowa (Figure 2).

Collaborations

As noted in our previous publication, the AIHG and NTR have collaborated on multiple projects and continue to do so (Kittlesrud et al., 2017). Collaborations between the AIHG and NTR include studies of personality traits associated with loneliness and poly-genetic risk predictions of loneliness (Abdellaoui et al., 2018, 2019), neuroticism (Boomsma et al., 2018) and twin discordance for basal metabolic rate. More recently, consortium collaborations have focused on behavioral and genetic associations in tobacco, alcohol and cannabis use (Liu et al., 2019; Minica et al., 2018), the genes of twinning (Mbarek et al., 2016), attention-deficit/hyperactivity disorder (Demontis et al., 2019) and heritability of common brain disorders (Anttila et al., 2018). Currently, collaborative efforts are aimed at investigating the degree of genetic similarity between Midwestern American (represented by participants of the ATR), Australian and Dutch (represented by participants of the NTR) populations (Beck et al., 2019). Ongoing collaborative studies concern the identification of genetic variants contributing to birth weight variation in twins compared to singletons.

Discussion and Future Directions

Survey design has evolved, related to the mode of delivery. For example, telephone surveys required shortened responses and gave point-labeled scale answer options compared to face-to-face survey methods (Dillman & Smyth, 2007). Telemarketing surveys began bombarding unsuspecting recipients in the late 1980s to early 1990s, creating a threat to the use of telephone survey delivery methods (Couper, 2000). With the increase in web survey use, there is a concern that over surveying may occur and data integrity may suffer. Recent evidence suggests that multiple forms of communication and survey may continue to be needed to maintain data integrity (McMaster et al., 2017). The ATR will utilize multiple contact methods and continue communications through newsletters, twin events and educational seminars, and will continue to consent and enroll participants online.

The ATR continues enrollment, and ATR participants who have consented to ongoing study participation will complete periodic surveys, establishing this register as a prospective study with repeated assessments of twin families. Due to the web-savvy public and increasing interest from our twin and multiple populations, online access to enrollment and ongoing questionnaires will continue. Recommendations from previous research conducted regarding online research include having additional modes of communication if studies are longitudinal projects. Therefore, the ATR will continue newsletters sent via post

and conduct twin events. Ancestry in the Central United States has been found to have roots from Northern Europe and provides an interesting genetic and environmental comparison between our region of the USA and the Netherlands population today.

Acknowledgments. The ATR is supported by Avera Health, Avera McKennan Hospital, and AIHG. The collaboration between the Netherlands and the Avera Twin Register arose through NIHM Grant: 1RC2MH089995-01: Genomics of Developmental Trajectories in Twins.

References

- Abdellaoui, A., Chen, H. Y., Willemsen, G., Ehli, E. A., Davies, G. E., Verweij, K. J. H., . . . Cacioppo, J. T. (2019). Associations between loneliness and personality are mostly driven by a genetic association with Neuroticism. *Journal of Personality, 87*, 386–397.
- Abdellaoui, A., Nivard, M. G., Hottenga, J. J., Fedko, I., Verweij, K. J. H., Baselmans, B. M. L., . . . Cacioppo, J. T. (2018). Predicting loneliness with polygenic scores of social, psychological and psychiatric traits. *Genes, Brain and Behavior, 17*, e12472.
- Anttila, V., Bulik-Sullivan, B., Finucane, H. K., Walters, R. K., Bras, J., Duncan, L., . . . Murray, R. (2018). Analysis of shared heritability in common disorders of the brain. *Science, 360*, eaap8757.
- Beck, J. J., Hottenga, J. J., Mbarek, H., Finnicum, C. T., Ehli, E. A., . . . Davies, G. E. (2019). Genetic similarity assessment of twin-family populations by custom-designed genotyping array. *Twin Research and Human Genetics*. doi: 10.1017/thg.2019.41
- Boomsma, D. I., Helmer, Q., Nieuwboer, H. A., Hottenga, J. J., de Moor, M. H., van den Berg, S. M., . . . de Geus, E. J. (2018). An extended twin-pedigree study of neuroticism in the Netherlands twin register. *Behavior Genetics, 48*, 1–11.
- Bryc, K., Durand, E. Y., Macpherson, J. M., Reich, D., & Mountain, J. L. (2015). The genetic ancestry of African Americans, Latinos, and European Americans across the United States. *American Journal of Human Genetics, 96*, 37–53.
- Couper, M. (2000). Web surveys: A review of issues and approaches. *Public Opinion Quarterly, 64*, 464–494.
- Demontis, D., Walters, R. K., Martin, J., Mattheisen, M., Als, T. D., Agerbo, E., . . . Neale, B. M. (2019). Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. *Nature Genetics, 51*, 63–75.
- Dillman, D. A., & Smyth, J. D. (2007). Design effects in the transition to web-based surveys. *American Journal of Preventive Medicine, 32*, S90–S96.
- Dillman, D. A., Tortora, R. D., & Bowker, D. (1998). Principles for constructing web surveys. (SESRC Technical Report, pp. 98–50). Pullman, WA: Washington State University Social and Economic Sciences Research Center.
- Harris, P. A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., & Conde, J. G. (2009). Research electronic data capture (REDCap) — A metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Informatics, 42*, 377–381.
- Harris, P. A., Taylor, R., Minor, B. L., Elliott, V., Fernandez, M., O’Neal, L., . . . Duda, S. N. (2019). The REDCap consortium: Building an international community of software platform partners. *Journal of Biomedical Informatics, 95*, 103208.
- Kittelsrud, J., Ehli, E. A., Petersen, V., Jung, T., Willemsen, G., Boomsma, D., & Davies, G. E. (2017). Establishment of the Avera Twin Register in the Midwest USA. *Twin Research and Human Genetics, 20*, 414–418.
- Liu, M., Jiang, Y., Wedow, R., Li, Y., Brazel, D. M., Chen, F., . . . Vrieze, S. (2019). Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nature Genetics, 51*, 237–244.
- Mbarek, H., Steinberg, S., Nyholt, D. R., Gordon, S. D., Miller, M. B., McRae, A. F., . . . Boomsma, D. I. (2016). Identification of common genetic variants influencing spontaneous dizygotic twinning and female fertility. *American Journal of Human Genetics, 98*, 898–908.
- McMaster, H. S., LeardMann, C. A., Speigle, S., Dillman, D. A., & Millennium Cohort Family Study. (2017). An experimental comparison of web-push vs. paper-only survey procedures for conducting an in-depth health survey of military spouses. *BMC Medical Research Methodology, 17*, 73.
- Minica, C. C., Verweij, K. J. H., van der Most, P. J., Mbarek, H., Bernard, M., van Eijk, K. R., . . . Derks, E. M. (2018). Genome-wide association meta-analysis of age at first cannabis use. *Addiction, 113*, 2073–2086.
- Segal, N. (2017). *Twin mythconceptions: False beliefs, fables, and facts about twins*. San Diego, CA: Academic Press.
- United States Census Bureau. (2018). Quick facts. Retrieved from <https://www.census.gov/quickfacts/fact/table/US/PST045218>
- Willemsen, G., Vink, J. M., Abdellaoui, A., den Braber, A., van Beek, J. H., Draisma, H. H., . . . Boomsma, D. I. (2013). The Adult Netherlands Twin Register: Twenty-five years of survey and biological data collection. *Twin Research and Human Genetics, 16*, 271–281.