Published in final edited form as:

Am J Med Genet B Neuropsychiatr Genet. 2012 December ; 0(8): 987–996. doi:10.1002/ajmg.b.32107.

Effectiveness of a Web-Based Protocol for the Screening and Phenotyping of Individuals with Tourette Syndrome for Genetic Studies

Crystelle Egan, PhD^{1,a}, Susan Marakovitz, PhD², Julia O'Rourke, PhD MS^{3,4}, Lisa Osiecki, BA⁵, Cornelia Illmann, PhD^{5,6}, Lauren Barton, BS^{5,6}, Elizabeth McLaughlin, BA^{5,6}, Rachel Proujansky, BA^{5,6}, Justin Royal, BA^{5,6}, Heather Cowley⁹, Martha Rangel-Lugo, PhD¹, David Pauls, PhD^{5,6}, Jeremiah M. Scharf, MD PhD^{5,6,7,8}, Carol A. Mathews, MD¹, and the Tourette Syndrome Association International Consortium for Genetics

¹Langley Porter Psychiatric Institute, Department of Psychiatry, University of California, San Francisco

²Integrated Center for Child Development, Newton, MA

³Department of Pediatrics, Lurie Center for Autism, Massachusetts General Hospital, Boston, MA

⁴Department of Medicine, Laboratory of Computer Science, Massachusetts General Hospital, Boston, MA

⁵Psychiatric Neurodevelopmental Genetics Unit, Center for Human Genetics Research, Massachusetts General Hospital, Boston, MA

⁶Department of Psychiatry, Massachusetts General Hospital; Boston, MA

⁷Department of Neurology, Massachusetts General Hospital; Boston, MA

⁸Division of Cognitive and Behavioral Neurology, Brigham and Women's Hospital, Boston, MA

⁹Tourette Syndrome Association, Bayside, NY

Abstract

Genome-wide association studies (GWAS) and other emerging technologies offer great promise for the identification of genetic risk factors for complex psychiatric disorders, yet such studies are constrained by the need for large sample sizes. Web-based collection offers a relatively untapped resource for increasing participant recruitment. Therefore, we developed and implemented a novel web-based screening and phenotyping protocol for genetic studies of Tourette Syndrome (TS), a childhood-onset neuropsychiatric disorder characterized by motor and vocal tics. Participants were recruited over a 13 month period through the membership of the Tourette Syndrome Association (TSA) (n=28,878). Of the TSA members contacted, 4.3% (1,242) initiated the questionnaire, and 79.5% (987) of these were enrollment eligible. 63.9% (631) of enrolled participants completed the study by submitting phenotypic data and blood specimens. Age was the only variable that predicted study completion; children and young adults were significantly less likely to be study completers than adults 26 and older. Compared to a clinic-based study conducted over the same time period, the web-based method yielded a 60% larger sample. Web-based participants were older and more often female; otherwise, the sample characteristics did not differ significantly. TS

^aCorresponding author: Crystelle Egan, PhD, Department of Psychiatry UCSF, 401 Parnassus Ave, Box F-0984, San Francisco, CA 94143, caegan@my.uri.edu, 401-290-7361 (phone); 415-476-7389 (fax).

Conflict of Interest

The authors have no conflicts of interest to declare.

diagnoses based on the web-screen demonstrated 100% accuracy compared to those derived from in-depth clinical interviews. Our results suggest that a web-based approach is effective for increasing the sample size for genetic studies of a relatively rare disorder and that our web-based screen is valid for diagnosing TS. Findings from this study should aid in the development of web-based protocols for other disorders.

Introduction

Tourette syndrome (TS) is a childhood-onset neuropsychiatric disorder defined by multiple frequent motor and one or more vocal tics persisting for at least a year (American Psychiatric Association, 2000; Freeman et al., 2000). Obsessive Compulsive Disorder (OCD) and Attention Deficit and Hyperactivity Disorder (ADHD) are often comorbid with TS. Between 30–60% of TS-affected individuals have OCD, 60–90% have ADHD, and up to 50% have both co-occurring diagnoses (Spencer et al., 1999; Coffey et al., 2000). Population prevalence estimates for TS vary from 0.1 to 1%, with more recent data suggesting a prevalence rate of approximately 0.3% (Robertson, 2008; Centers for Disease Control, 2009; Scharf et al., 2012; Scharf and Pauls, 2007).

While numerous family and twin studies provide strong evidence for the heritability of TS (O'Rourke et al., 2009; TSAICG, 2007), definite susceptibility genes have not yet been identified. Multiple linkage regions and candidate genes have been reported for TS using both targeted and genome-wide approaches, but these findings have not been consistently replicated, and together only account for a small proportion of the genetic variance of TS (Ercan-Sencicek, et al., 2010; O'Rourke et al., 2009; Sundaram et al., 2010). Genome-wide association studies (GWAS), as well as other currently emerging genetic approaches such as the analysis of copy number deletions (CNVs) and exome sequencing, can complement linkage and candidate gene studies and offer great promise for the identification of genetic risk factors for complex heritable disorders such as TS (Simon-Sanchez and Singleton, 2008). Scientific advances in genetic analysis and decreasing costs have enhanced the feasibility of conducting large-scale genetic studies (PGCCC, 2009; Sullivan, 2012). Data from the first TS GWAS, a multi-center case-control study with 1,496 TS cases and 55,249 controls, have recently been completed, and although several promising variants have been identified, no definitive susceptibility genes have yet been confirmed (Scharf et al., 2012).

Currently, the recruitment and phenotyping of participants has become the largest obstacle to the advancement of genetic studies (Bilder et al., 2009). For genetic studies, particularly those focused on identifying common variants such as GWAS, to be adequately powered to detect reproducible genetic associations, thousands to tens of thousands of affected participants are required (PGSCC, 2009). Five years of GWAS studies have shown that the number of identified risk variants increases linearly with increasing sample size, with success rates increasing substantially with sample sizes above 10,000. Although GWAS has seen some success for neuropsychiatric disorders, the sample sizes available are still small compared to other etiologically complex disorders (Sullivan, 2012; Visscher et al., 2012). Clinic-based methods are often insufficient to achieve these sample sizes within a reasonable timeframe, especially for relatively rare disorders such as TS. Thus, the development of novel approaches to rapidly recruit, phenotype, and collect biological samples from affected individuals is critical (Bilder et al., 2009).

In order to facilitate the collection of the large number of samples needed for genetic studies, researchers have developed a range of creative recruitment methods, including collaboration among multiple centers (ANZgene, 2009; Boomsma et al., 2008; PGCCC et al., 2009; Holland et al., 2009; Mick et al., 2010), utilization of professionally-moderated

registries of affected individuals and general populations (Low et al., 2010; Boomsma et al., 2008; Hiura et al., 2010), and condition-specific web-based registries and websites directly available to potential participants (Lee et al., 2010; Salkowski et al., 2001). Few reports are available on the implementation and psychometric properties of novel methods for rapid phenotyping. Studies have reported on the collection of web-facilitated phenotype data for control participants and for participants with common physical traits (Sanders et al., 2010; Eriksson et al., 2010), and one recent study established the validity of a web-based questionnaire plus documented prior clinician diagnosis for Autism Spectrum Disorder (ASD) (Lee et al., 2010).

Non-genetic clinical and survey research has demonstrated that recruiting research study participants via the internet is effective (An et al., 2007; Arab et al., 2010; Miller and Sonderland, 2010; Smith et al., 2007), especially for specialized populations (Miller and Sonderlund, 2010; Mangan and Reips, 2007; Skitka and Sargis, 2006), although it carries the attendant risk of unrepresentative samples (Klovning et al., 2009; Miller and Sonderlund, 2010; Skitka and Sargis, 2006). Web-based questionnaires have been shown to be reliable (Rankin et al., 2008; Young et al., 2009; Lieberman, 2008), and valid for various psychiatric diagnoses such as depression, ASD, and ADHD using DSM criteria (Cawthorpe, 2001; Cuijpers et al., 2008; Lin et al., 2007; Steenhuis et al., 2009). Scores from interviews and surveys adapted for Internet have been found equivalent to standard method scores (Coles et al., 2007), while some studies report more severity and pathology in Internet-assessed samples (Cuijpers et al., 2008; Sanders et al., 2010; Whitehead, 2007).

Here we report on the initial results of recruitment for a TS GWAS study using a novel web-based screening and rapid phenotyping protocol. By leveraging the resources of a national patient advocacy organization and recruiting subjects nationwide rather than solely through specialty clinics with limited catchment areas, we anticipated that we would significantly increase the number of DNA samples collected from TS-affected individuals compared to traditional recruitment methods. In addition, we expected that our procedure for web-based TS diagnosis would have equivalent accuracy to diagnoses established by clinical interview. We also present data about web-based participant characteristics, including comparison to a clinic-based sample, and data regarding study processes.

Materials and Methods

Web-based Screening and Assessment Tool

In collaboration with the Tourette Syndrome Association (TSA), investigators at the Massachusetts General Hospital (MGH) and the University of California, San Francisco (UCSF) developed a web-based rapid phenotyping protocol designed to screen and assess TS-affected individuals for eligibility in genetic studies (findtsgenes.org). Potential participants who accessed the website first reviewed a brief study description and informed consent information before answering a questionnaire that could be completed in approximately 20–30 minutes. Parents were asked to complete the questionnaire for their children under the age of 18.

Two secure databases that were linked via coded IDs, one for participant tracking and the second for phenotype data, were designed to separate patient identifiers from their phenotype data. A user-friendly web interface was created for the protocol. The website was pilot-tested with 30 participants at a TS conference in April 2010.

This research project was approved prior to implementation by the institutional review boards of MGH and UCSF.

Informed Consent

Working closely with our institutional review boards, we created a two-stage informed consent procedure that limited the initial online consent to participation in the web-based questionnaire and required separate consent documentation for the blood draw and DNA extraction and biobanking. The online procedure specifically addressed the completion of the questionnaire by a minor and comprehension of the informed consent via a series of multiple-choice questions. Participants were also asked in the web-based document if they agreed to be re-contacted for future studies of Tourette Syndrome. Separate consent was required for the blood draw, biobanking, and genetic studies, as we believe that adequate explanation of the potential risks and benefits of participation in genetic studies requires a more in-depth discussion than was possible in the brief web-based structure.

Screening

Because our clinic-based studies showed that 98.9% of self-reported clinician-diagnosed TS cases met our best estimate (BE) consensus criteria for TS (C.A. Mathews, personal communication, February 6, 2012), self-reported diagnosis of TS by a clinician was established as the first-level screening criterion. Participant endorsement of at least two motor tics and a vocal tic during childhood persisting for at least a year served as further screening criteria. This study was limited to one TS-affected individual per family who had not participated in another genetic study of TS in the past 5 years. Computer-based algorithms were developed to make TS diagnoses based on participants' responses to the screening questions, and to flag diagnostic exclusionary criteria: intellectual disability, epilepsy, and any genetic or other neurological disorder that might confound a TS diagnosis. The database manager reviewed new entries for ambiguous screening criteria and multiple entries. Questionable cases were further investigated by study staff.

Demographic and Medical History

Participants provided demographic and medical history information including age of symptom onset, age of TS diagnosis, diagnoses of OCD, ADHD, ASD, history of seizures, and other neurological and genetic conditions. Participants were also asked about their first degree and other relatives' history of tics and related disorders.

Diagnostic Measures

Brief sets of symptom-related questions were employed to confirm the diagnoses of interest (TS, OCD, and ADHD). The tic section (see Supplementary Materials) included questions about lifetime presence, duration, and type of simple and complex vocal and motor tics. The items were selected from a comprehensive tic assessment administered in our clinic-based genetic studies of TS (TSAICG, 2007) by identifying those items endorsed most frequently and most reliably associated with a BE diagnosis of TS in our existing sample of over 1,600 participants.

The OCD section assessed current and lifetime symptoms and was modeled on the Florida Obsessive Compulsive Inventory (FOCI) (Storch et al., 2009), a screen for OCD keyed to DSM-IV-TR diagnostic criteria, which includes specific examples of obsessions and compulsions, as well as measures of duration, impairment, and distress.

The ADHD section was modeled on the SNAP IV (Swanson et al., 1992), an assessment tool for ADHD corresponding to DSM-IV-TR diagnostic criteria and used in our current clinic-based studies. The SNAP IV has been used widely in intervention and genetic studies and has been found to be reliable, to have high internal consistency, and a factor structure corresponding to the ADHD subtypes (Bussing et al., 2008; Solanto and Alvir, 2009).

Recruitment

Participants were recruited primarily through the Tourette Syndrome Association (TSA), a national non-profit membership organization, from a database of 17,964 email and 28,878 mailing addresses for consumer members who agreed to contact regarding research studies (not mutually exclusive, as some TSA members provided both email and mailing addresses). From April to November 2010, we sent notices about the study via email and regular mail. An average of 2,551 email recruitment messages were sent out monthly based on the outcomes of pilot email blasts, during which 6.24% of message recipients clicked on the link to the study website and approximately 11% of the email messages were returned due to incorrect addresses. A mailing campaign was implemented concurrently. In addition, we created a Facebook page for the study and posted information on the TSA's Facebook page.

Follow-up with Enrolled Participants

After completion of the online questionnaire, eligible participants were assigned to a study site (MGH or UCSF) based on geographic location for follow-up to: 1) obtain informed consent for the blood draw, and 2) facilitate the blood draw and banking of DNA samples for the genetic study. Once the participant returned signed informed consent forms, a blood-drawing kit was shipped to the participant, who selected either to have a national mobile blood drawing service conduct the blood draw at their home or to travel to a local site of a national diagnostic laboratory. After collection, the blood specimens were shipped to the Coriell Human Genetics DNA and Cell Line Repository at the National Institute of Neurological Disorders and Stroke (NINDS) where DNA was extracted and banked indefinitely for use both in our current study and in future studies of TS and related disorders by other academic researchers. Participants were compensated with a \$5 gift card sent to their homes upon completion of the blood draw. An additional \$5 gift card was sent to those individuals who participated in the clinical confirmation interviews.

Each contact with a participant and action taken by study staff on behalf of a participant was logged in a participant tracking database. These actions included communication by phone and email, sending informed consent packets, sending blood drawing kits, booking the mobile blood drawing service, and contact with blood drawing laboratory sites. Multiple methods of contact (mail, email, home and cell telephone numbers) were used to follow up with participants. After six unsuccessful contact attempts, participants who did not respond were contacted one additional time by phone and/or email and converted to a "Lost to Follow-up" status if no response was received.

Clinical Confirmation Interview

To assess for the validity of TS, OCD, and ADHD diagnoses, ten percent of enrolled individuals were randomly selected to participate in an in-depth clinical interview that paralleled the clinic-based assessment. The interview, which utilized a validated and standardized, semi-structured instrument with high validity and reliability for TS (TSAICG, 2007), was conducted via a video web-chatting system by trained clinical psychologists so that tics, if present, could be observed by the clinician. Confirmatory clinical diagnoses of TS were then assigned based on review of the interview materials by a Ph.D.-level psychologist trained in the diagnosis of TS, OCD and ADHD.

Sample

To define the sample for the current analyses, potential participants were included if they were enrolled via the web within 6 weeks following the last set of emailed and mailed recruitment messages. We established an end date for study completion 148 days later, given that 90% of participants who completed the study did so within 148 days (data not shown).

Enrolled participants who submitted a blood specimen were classified as “completers,” and enrolled participants who did not submit a blood specimen were classified as “non-completers.”

Statistical Analyses

Descriptive statistics were tabulated for: 1) ineligible and eligible respondents, enrolled participants, and enrolled participants who provided biological specimens; 2) the demographic and clinical characteristics of enrolled participants as a group and by study completion status; 3) eligible participants’ reasons for withdrawal from the study; 4) the number of days taken to complete the study; and 5) the number of actions taken by study staff to assist each participant with study completion as a group and by study administration site.

The characteristics of study “completers” and “non-completers” were compared using chi-square analysis and independent samples t-tests. In addition, study outcomes, including completion percentages, number of days to completion, and number of actions were compared between sites using chi-square analyses with Yates’ Correction of Continuity for 2 by 2 tables and independent samples t-tests.

A direct logistic regression analysis was conducted to explore relationships among participant age, sex, racial identity, census-derived household income, clinician-diagnosed OCD, clinician-diagnosed ADHD, and the outcome variable completion status (“completer” and “non-completer”). Participant zip code was used to derive estimated household income as a proxy for SES using the most recent data available from the 2000 census. Race was dichotomized as non-white and white due to the small sample size of each of the non-white racial/ethnic groups. Age was initially examined as a continuous variable, and subsequently divided into 5 groups hypothesized to share common response patterns from anecdotal observation. A significance level of $p < .005$ accounted for Bonferroni correction for 10 examined predictors with an alpha level of $p < .05$.

Descriptive statistics were generated to determine concordance between TS diagnoses assigned by the web-based algorithm and clinician-rated diagnoses generated for the subset of participants who completed the in-depth clinical interview.

The number of collected blood samples as well as demographic and clinical characteristics of study completers were compared to a sample of TS-affected participants from a clinic-based study conducted over the same time period. The clinic-based study included five U.S. TS clinic sites (MGH, UCSF, University of Utah, Johns Hopkins, and Yale). Individual TS-affected probands and trios (TS-affected child, mother, and father) were recruited by the clinics primarily from their patient populations. The participants were phenotyped by individual interview using an in-depth questionnaire with blood sample collection arranged by the clinic study staff and then banked at the NINDS Repository. Chi-square tests were performed to compare the categorical variables sex, race, Hispanic ethnicity, ADHD, OCD, and family history of tic disorders among first-degree relatives. Yates’ Correction of Continuity was utilized for 2 by 2 tables. Age, as a continuous variable, was compared using an independent samples t-test. The alpha level of $p < 0.01$, as a relatively stringent threshold, was used for statistical significance.

Analyses were conducted using the statistical software package SPSS Statistics 18.0 (IBM, Armonk, NY).

Results

During the 13-month recruitment period, 1,242 individuals initiated or completed the web-based questionnaire. 987 respondents were eligible and thus enrolled in the study, including participants from 48 states and the District of Columbia. The majority of enrolled participants were male (70.3%), identified as white (89.2%), and as non-Hispanic (92.6%). 63.9% of enrolled subjects completed the blood draw (“completers”) and had DNA specimens successfully biobanked (n=631). Figure 1 provides a flow chart of this process including reasons for ineligibility and non-completion. In univariable analyses, older participants and those with non-Hispanic ethnicity were significantly more likely to be completers ($p < 0.001$ and $p = 0.048$, respectively; Table 1). Of the 19.4% of enrolled participants who actively withdrew from the study, fear of phlebotomy among minor probands was the most cited reason (36.2%). Table 2 provides the complete list of reasons provided.

Predictors of Completion

934 cases were entered into the logistic regression due to missing data on race and census-derived income for 53 subjects. A test of the full predictor model against a constant-only model was significant, $\chi^2(10, N = 934) = 63.97$, $p < .001$. Participant age was the only significant predictor of study completion. For this reason, we divided age into categories in order to further examine whether particular age groups were more or less likely to complete the study. In comparison to participants 26 years of age and older, 6–8 year olds were about a fifth as likely to complete the blood draw, 9–12 year olds were half as likely, and 18–25 year olds were about a third as likely to be study completers (Table 3 and Figure 2).

Study Completion Process

Study staff generated on average 6.7 (SD=3.8) actions on behalf of “completers.” In comparison, staff generated significantly more actions, an average of 8.22 (SD=3.4), for “non-completers” ($p < .0001$). The two coordinating sites had similar completion rates (MGH: n=431, 62.1%; UCSF: n=200, 68.3%; $p = .08$). Both sites recorded an equivalent number of contacts/actions with participants (MGH: 6.7 (3.6); UCSF: 6.7 (3.0); $p = .97$).

Concordance Between Web-based and Clinical-interview TS Diagnoses

There was 100% concordance between TS diagnoses using the web-based protocol and diagnoses established using the gold standard in-depth clinical interview for the 74 participants who completed both assessments.

Comparison of the Web-based and Clinic-Based Studies

Over the same 13-month time period as the web-based study, 238 TS-affected probands were enrolled in the standard clinic-based study. Ten probands (4.2%) did not complete the clinical interview after providing blood specimens. Figure 3 compares the number of cases completed per month for the web-based and clinic-based studies. An additional 197 (primarily unaffected) first-degree family members were enrolled in the clinic-based study; 15 of these did not provide phenotype data.

In comparison to the clinic-based study sample, the web-based sample was significantly older and had a higher proportion of females. The groups did not significantly differ on other variables (Table 4).

Discussion

Results from this study indicate that in a relatively short period of time the web-based TS screening and phenotyping protocol effectively increased study enrollment and the collection of biological samples by 60% over traditional approaches while maintaining high accuracy for a diagnosis of TS. The web-based method was notable for yielding blood specimen donations from older participants and from a higher proportion of female participants. Otherwise, the samples did not differ on demographic and clinical characteristics. Unfortunately, the web-based approach did not increase the racial and ethnic diversity of the sample. As >95% of the participants agreed to future recontact, including learning about other TS-related research opportunities, this approach also provides an ongoing resource for TS investigators.

As the research literature regarding web-based phenotyping and biological sample collection is limited, it is difficult to assess the success of the approach described here relative to those from other disorders and web-based protocols. Two prior studies indicated that different disorders and approaches may yield varying sample sizes compared to the 631 in our study, from 2,750 participants a year for ASD (Lee et al., 2010), to 141 eligible participants over 18 months for a study of aneurysms (Salkowski et al., 2000). These differences may be due to variations such as disease prevalence, recognition of the disorder, and degree of impairment.

In this study, the validity of the rapid phenotyping protocol for TS diagnosis was confirmed by in-depth interviews conducted with a sub-sample of the web-based participants. We had 100% concordance for TS diagnoses in the 10% of participants interviewed, and are confident in the specificity of the screening instrument and our inclusion algorithm for identifying clear-cut cases of TS. However, as we interviewed a relatively small number of participants in this study, further validation is needed. Clinical interviews of additional participants are currently underway, including interviews of all individuals who were excluded from the study because they did not meet our inclusion criteria, in order to confirm and refine the utility of the screen for accurately identifying TS.

The validity of Internet-facilitated questionnaires for psychiatric diagnosis has been reported for other disorders (Cawthorpe, 2001; Cuijpers et al., 2008; Lin et al., 2007; Steenhuis et al., 2009). Our finding contributes to the limited published research on the psychometric properties of web-based questionnaires designed specifically for rapid phenotyping. The only published validity study to our knowledge of a web-based rapid phenotyping questionnaire also found it valid for diagnosis when accompanied by a prior clinical diagnosis, in this case of ASD, compared to a standard clinical assessment battery (Lee et al., 2010).

Age was the only variable that significantly predicted completion in the web-based study, although we hypothesized that other variables (e.g., severity of the disorder (as defined by comorbid ADHD and/or OCD), racial background, SES, and gender) could also play a role. Younger participants, especially those aged 5–12, were significantly less likely to submit a blood sample than adults 26 years of age and older. For young participants, we can assume that fear of phlebotomy was the most likely barrier to completion based on the reasons for study discontinuation given by participants who actively withdrew. Young participants may be more effectively recruited through specialty clinics, as participants willing to undergo a blood draw may be more readily identified in these settings and existing relationships with care providers may reduce phlebotomy-related fears.

The fact that non-completers received significantly more staff attention than completers further indicates that factors other than staff follow up contributed to non-completion. In

additional to fear of phlebotomy, concern about DNA privacy appears to be a considerable barrier to full participation, based on responses from those non-completers who provided reasons for withdrawal. We believe that, while it may lead to non-completion of the study in some cases, the use of a two-stage informed consent process allowed participants more opportunity to consider the potential risks of participation.

There are several caveats to consider in implementing this type of web-based protocol. First, successful enrollment in this study was founded on a large recruitment base of nearly 30,000 TSA members. Only 4.3% of the contacted TSA members initiated the online questionnaire and 2.1% submitted blood samples for the genetic study. Second, the process, while automated to a large degree, still requires staff effort and can be time consuming. On average seven actions by study staff were required to facilitate completion of one blood draw, and 25% of completers required more intensive follow up, between 9 and 26 actions. However, as a large proportion of the cost associated with web-based screening are one-time costs (e.g., software design and implementation), this approach provides a viable and cost effective alternative to the standard clinic based approaches for participant recruitment, particularly for relatively rare disorders.

Limitations

This study examined results from a preliminary sample of participants in an ongoing recruitment process for genetic studies of TS. Thus, we expect more TSA members to enroll in and complete the study as it is further publicized and members are contacted a second time. A significant limitation of this study is the lack of racial and ethnic diversity in the sample. TS diagnosis rates for Hispanic and non-Hispanic blacks are less than half the rate for whites (Centers for Disease Control, 2009); still, our sample reflects low participation, especially by African American participants who comprised .4% of enrollees in the web-based study. Different methods of recruitment are clearly required to involve a more racially and ethnically representative sample. In addition, although the TS diagnoses were clearly valid in the web-based screen compared to the clinical interview, the validity of the web-based questionnaire for OCD and ADHD diagnoses has not yet been established. As mentioned above, a validity study is underway to compare web-facilitated diagnoses with best-estimate consensus diagnoses from in-depth clinical interviews. Finally, although zip code of the participants was used as a proxy for socioeconomic status, and is a commonly utilized method in epidemiological studies, this approach is imperfect, and may not accurately represent the SES of the participants (Soobader, et al, 2001).

Conclusions

Use of web-based screening and phenotyping for genetic studies, especially for rare disorders, offers relatively untapped potential for increasing sample sizes to the numbers necessary for genetic studies of complex disorders (either GWAS or studies of rare mutations). This study offers strong support for the effectiveness of a web-based method for the recruitment, rapid phenotyping, and collection of blood specimens from TS-affected participants and highlights the potential of this approach to increase participation numbers over traditional clinic-based methods. In addition, this approach is potentially useful for a variety of study designs, as family members of identified probands could also easily be recruited and assessed via the web.

As noted previously, participants in the web-based collection tended to be older and more often female than clinic-based participants. Because of the early age of onset of Tourette Syndrome, and the 3:1 male to female gender ratio, female participants and adult participants tend to be under-represented in research studies of TS; these participants may in fact substantially enhance the power of genetic studies as they may represent an increased

severity of illness (represented by tic persistence in adults) or an increased genetic load (Pauls et al., 1981; Szatmari et al., 2012). Furthermore, this study establishes the validity of web-based diagnosis of TS. Although we cannot be certain that this approach will generalize to other disorders, we hope that our findings can inform the development of future web-based approaches. Further studies involving other affected populations are necessary to demonstrate conclusively that this approach is effective for other disorders.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This study was funded by an American Recovery and Reinvestment Act award from the National Institute for Neurological Disorders and Stroke (NINDS) U01 NS40024-S1, the primary NINDS award U01 NS40024, MH085057 and an additional award from the Tourette Syndrome Association (TSA).

We gratefully acknowledge the individuals with TS and their families who participated in this study. We would like to acknowledge the efforts of the TSA staff members involved in this project, especially Mary Carball. We also acknowledge the members of the Tourette Syndrome Association International Consortium for Genetics (TSAICG), listed alphabetically by city: D. Cath and P. Heutink, Departments of Psychiatry and Human Genetics, Free University Medical Center, Amsterdam, The Netherlands; M. Grados, H.S. Singer, and J.T. Walkup, Departments of Psychiatry and Neurology, Johns Hopkins University School of Medicine, Baltimore, MD; J.M. Scharf, C. Illmann, D. Yu, S. E. Stewart, and D.L. Pauls, Psychiatric and Neurodevelopmental Genetics Unit, Center for Human Genetic Research, Massachusetts General Hospital Harvard Medical School, Boston, MA; N.J. Cox, Departments of Medicine and Human Genetics, University of Chicago, Chicago, IL; S. Service, and N.B. Freimer, Departments of Psychiatry, Human Genetics and Statistics, U.C.L.A Medical School, Los Angeles, CA; M.M. Robertson, St George's Hospital and Medical School, University College London, London, England; G.A. Rouleau, J.-B. Riviere, S. Chouinard, F. Richer, P. Lesperance, and Y. Dion, University of Montreal, Montreal, Quebec, Canada; R.A. King, J.R. Kidd, A.J. Pakstis, J.F. Leckman, and K.K. Kidd, Department of Genetics and the Child Study Center, Yale University School of Medicine, New Haven, CT; R. Kurlan, Atlantic Neuroscience Institute, Overlook Hospital, Summit, NJ, USA; B.A. Oostra, Department of Clinical Genetics, Erasmus University, Rotterdam, The Netherlands; G. Lyon, W. McMahon, Departments of Psychiatry and Human Genetics, University of Utah School of Medicine, Salt Lake City, UT; C.A. Mathews, Department of Psychiatry, University of California, San Francisco, San Francisco, CA; P. Sandor and C.L. Barr, Department of Psychiatry, The Toronto Hospital and University of Toronto, Toronto, Ontario, Canada.

References

- American Psychiatric Association. Diagnostic and statistical manual of mental disorders, DSM-IV-TR. Revised 4th ed.. Washington, DC: Psychiatric Publishing Inc; 2000. p. 943
- An LCHD, Perry CL, Lein EB, Klatt C, Farley DM, Bliss RL, Pallonen UELH, Ehlinger EP, Ahluwalia JS. Feasibility of Internet health screening to recruit college students to an online smoking cessation intervention. *Nicotine and Tobacco Research*. 2007; 9(Suppl 1):S11–S18. [PubMed: 17365722]
- Australia and New Zealand Multiple Sclerosis Genetics Consortium [ANZgene]. Genome-wide association study identifies new multiple sclerosis susceptibility loci on chromosomes 12 and 20 *Nature Genetics*. 2009; 41:824–828.
- Arab L, Hahn H, Henry J, Chacko S, Winter A, Cambou MC. Using the web for recruitment, screen, tracking, data management, and quality control in a dietary assessment clinical validation trial. *Contemp Clin Trials*. 2010; 31:138–46. [PubMed: 19925884]
- Bilder RM, Sabb FW, Cannon TD, London ED, Jentsch JD, Stott Parker D, Poldrack D, Evans C, Freime NB. Phenomics: the systematic study of phenotypes on a genome-wide scale. *Neuroscience*. 2009; 164:30–42. [PubMed: 19344640]
- Boomsma DI, Willemsen G, Sullivan PF, Heutink P, Meijer P, Sondervan D, Kluft C, Smit G, Nolen WA, Zitman FG. others. Genome-wide association of major depression: description of samples for the GAIN Major Depressive Disorder Study: NTR and NESDA biobank projects. *Eur J Hum Genet*. 2008; 16:335–342. [PubMed: 18197199]

- Bussing R, Fernandez M, Harwood M, Hou W, Garvan CW, Eyberg SM, Swanson JM. Parent and teacher SNAP-IV ratings of attention deficit hyperactivity disorder symptoms. *Assessment*. 2008; 15:317–328. [PubMed: 18310593]
- Cawthorpe D. An evaluation of a computer-based psychiatric assessment: evidence for expanded use. *Cyberpsychol Behav*. 2001; 4:503–510. [PubMed: 11708730]
- Centers for Disease Control. Tourette syndrome in persons aged 6–17 Years — United States, 2007. *Morbidity and Mortality Weekly Report*. 2009; 58:581–585. [PubMed: 19498335]
- Coffey BJ, Biederman J, Smoller JW, Geller DA, Sarin P, Schwartz S, et al. Anxiety disorders and tic severity in juveniles with Tourette's disorder. *Journal of the American Academy of Child and Adolescent Psychiatry*. 2000; 39:562–568. [PubMed: 10802973]
- Coles ME, Cook LM, Blakeb TR. Assessing obsessive compulsive symptoms and cognitions on the Internet: evidence for the comparability of paper and Internet administration. *Behaviour Research and Therapy*. 2007; 45:2232–2240. [PubMed: 17306222]
- Cuijpers BP, Boluijt P, van Straten A. Screening of depression in adolescents through the Internet: sensitivity and specificity of two screening questionnaires. *European Child and Adolescent Psychiatry*. 2008; 17:32–38. [PubMed: 17876508]
- Eriksson N, Macpherson JM, Tung JY, Hon LS, Naughton B, et al. Web-based, participant-driven studies yield novel genetic associations for common traits. *PLoS Genet*. 2010; 6(6):e1000993. [PubMed: 20585627]
- Ercan-Sencicek AG, Stillman AA, Ghosh AK, Bilguvar K, O'Roak BJ, Mason CE, et al. L-histidine decarboxylase and Tourette's syndrome. *New England Journal of Medicine*. 2010; 362:1901–1908. [PubMed: 20445167]
- Freeman R, Fast D, Burd L, Kerbeshian J, Robertson M, Sandor P. An international perspective on Tourette syndrome: selected findings from 3500 individuals in 22 countries. *Developmental Medicine and Child Neurology*. 2000; 42:436–447. [PubMed: 10972415]
- Hiura Y, Tabara Y, Kokubo Y, Okamura T, Miki T, Tomoike H, Iwai N. A genome-wide association study of hypertension-related phenotypes in a Japanese population. *Circulation Journal*. 2010; 74(11):2353–2359. [PubMed: 20877124]
- Holland A, Whittington J, Cohen O, Curfs L, Delahaye F, Dudley O, Horsthemke B, Lindgren A-C, Nourissier C, Sharma N, et al. The European Prader-Willi Syndrome Clinical Research Database: an aid in the investigation of a rare genetically determined neurodevelopmental disorder. *Journal of Intellectual Disability Research*. 2009; 53:538–547. [PubMed: 19457156]
- Klovning A, Hogne S, Hunskaara S. Web-based survey attracted age-biased sample with more severe illness than paper-based survey. *Journal of Clinical Epidemiology*. 2009; 62:1068–1074. [PubMed: 19246177]
- Lee H, Marvin AR, Watson T, Piggot J, Law JK, Law PA, Constantino JN, Nelson SF. Accuracy of phenotyping of autistic children based on Internet implemented parent report. *Am J Med Genet B Neuropsychiatr Genet*. 2010; 153B:1119–1126. [PubMed: 20552678]
- Lieberman DZ. Evaluation of the stability and validity of participant samples recruited over the Internet. *Cyberpsychol Behav*. 2008; 11(6):743–745. [PubMed: 18954280]
- Lin CC, Bai YM, Liu CY, Hsiao MC, Chen JY, Tsai SJ, Ouyang WC, Wu CH, Li YC. Web-based tools can be used reliably to detect patients with major depressive disorder and subsyndromal depressive symptoms. *BMC Psychiatry*. 2007; 7:12. [PubMed: 17425774]
- Low SK, Kuchiba A, Zembutsu H, Saito A, Takahashi A, Kubo M, Daigo Y, Kamatani N, Chiku S, Totsuka H. others. Genome-wide association study of pancreatic cancer in Japanese population. *PLoS One*. 2010; 5:e11824. [PubMed: 20686608]
- Mangan MA, Reips UD. Sleep, sex, and the Web: surveying the difficult-to-reach clinical population suffering from sexomnia. *Behav Res Methods*. 2007; 39:233–236. [PubMed: 17695349]
- Mick E, Todorov A, Smalley S, Hu X, Loo S, Todd RD, Biederman J, Byrne D, Dechairo B, Guiney A. others. Family-based genome-wide association scan of attention-deficit/hyperactivity disorder. *J Am Acad Child Adolesc Psychiatry*. 2010; 49:898–905. e3. [PubMed: 20732626]
- Miller PG, Sonderlund AL. Using the Internet to research hidden populations of illicit drug users: a review. *Addiction*. 2010; 105:1557–1567. [PubMed: 20626378]

- O'Rourke JA, Scharf JM, Yu D, Pauls DL. The genetics of Tourette syndrome: a review. *Journal of Psychosomatic Research*. 2009; 67:533–545. [PubMed: 19913658]
- Pauls DL, Cohen DJ, Heimbuch R, Detlor J, Kidd KK. Familial pattern and transmission of Gilles de la Tourette syndrome and multiple tics. *Arch Gen Psychiatry*. 1981; 38:1091–1093. [PubMed: 6945827]
- Psychiatric GWAS Consortium Coordinating Committee [PGCCC]. Genomewide association studies: history, rationale, and prospects for psychiatric disorders. *Am J Psychiatry*. 2009; 166:540–556. [PubMed: 19339359]
- Rankin KM, Rauscher GH, McCarthy B, Erdal S, Lada P, Il'yasova D, Davis F. Comparing the reliability of responses to telephone-administered versus self-administered Web-based surveys in a case-control study of adult malignant brain cancer. *Cancer Epidemiol Biomarkers Prev*. 2008; 17:2639–2646. [PubMed: 18843005]
- Robertson M. The prevalence and epidemiology of Gilles de la Tourette syndrome: Part 1: the epidemiological and prevalence studies. *Journal of Psychosomatic Research*. 2008; 65:461–472. [PubMed: 18940377]
- Salkowski A, Tromp G, Greb A, Womble D, Kuivaniemi H. Web-site-based recruitment for research studies on abdominal aortic and intracranial aneurysms. *Genet Test*. 2001; 5:307–310. [PubMed: 11960575]
- Sanders ARLD, Duan J, Dennis JM, Li R, Kendler KS, Rice JP, Shi J, Mowry BJ, Amin F, Silverman JM, Buccola NG, Byerley WF, Black DW, Freedman R, Cloninger CR, Gejman PV. The Internet-based MGS2 control sample: self report of mental illness. *Am J Psychiatry*. 2010; 167:854–65. [PubMed: 20516154]
- Scharf J, Miller L, Mathews C, Ben-Shlomo Y. Prevalence of Tourette syndrome and chronic tics in the population-based Avon longitudinal study of parents and children cohort. *Journal of the American Academy of Child and Adolescent Psychiatry*. 2012; 51:192–201. [PubMed: 22265365]
- Scharf, JM.; Pauls, DL. Genetics of tic disorders. In: Rimoin, DL.; Connor, JM.; Pyeritz, RE.; Korf, BR., editors. *Emery and Rimoin's principles and practices of medical genetics*. 5th ed.. Philadelphia: Churchill Livingstone/Elsevier; 2007. p. 2737-2754.
- Scharf JM, Yu D, Mathews CA, Neale BM, Stewart SE, Fagerness JA, et al. Genome-wide association study of Tourette's syndrome. *Mol Psychiatry*. Aug 14.2012 Epub ahead of print.
- Simón-Sánchez J, Singleton A. Genome-wide association studies in neurological disorders. *The Lancet Neurology*. 2008; 7:1067–1072.
- Skitka LJ, Sargis EG. The Internet as psychological laboratory. *Annu Rev Psychol*. 2006; 57:529–55. [PubMed: 16318606]
- Smith KS, Eubanks D, Petrik A, Stevens VJ. Using web-based screening to enhance efficiency of HMO clinical trial recruitment in women aged forty and older. *Clin Trials*. 2007; 4:102–105. [PubMed: 17327250]
- Solanto MV, Alvir J. Reliability of DSM-IV Symptom Ratings of ADHD. *Journal of Attention Disorders*. 2009; 13:107. [PubMed: 19372494]
- Soobader M, Le Clere FB, Hadden W, Maury B. Using aggregate geographic data to proxy individual socioeconomic status: does size matter? *American Journal of Public Health*. 2001; 91:632–636. [PubMed: 11291379]
- Spencer T, Biederman J, Coffey B, Geller D, Wilens T, Faraones S. The 4-year course of tic disorders in boys with attention-deficit/hyperactivity disorder. *Archives of General Psychiatry*. 1999; 56:842–847. [PubMed: 12884890]
- Steenhuis M, Serra M, Minderaa R, Hartman CA. An Internet version of the Diagnostic Interview Schedule for Children (DISC-IV): correspondence of the ADHD section with the paper-and-pencil version. *Psychological Assessment*. 2009; 21:231–234. [PubMed: 19485678]
- Storch EA, Khanna M, Merlo LJ, Loew BA, Franklin M, Reid JM, Goodman WK, Murphy TK. Children's Florida Obsessive Compulsive Inventory: psychometric properties and feasibility of a self-report measure of obsessive-compulsive symptoms in youth. *Child Psychiatry & Human Development*. 2009; 40:467–483. [PubMed: 19326209]
- Sullivan P. Don't give up on GWAS. *Molecular Psychiatry*. 2012; 17:2–8. [PubMed: 21826059]

- Sundaram SK, Huq AM, Wilson BJ, Chugani HT. Tourette syndrome is associated with recurrent exonic copy number variants. *Neurology*. 2010; 74:1583–1590. [PubMed: 20427753]
- Swanson, J.; Nolan, W.; Pelham, W. The SNAP-IV rating scale. 1992. URL: <http://www.adhd.net> [14.11.2009]
- Szatmari P, Liu X, Goldberg J, Zwaigenbaum L, Paterson A, Woodbury-Smith M, Georgiades S, Duku E, Thompson A. Sex differences in repetitive stereotyped behaviors in autism: implications for genetic liability. *American Journal of Medical Genetics Part B. Neuropsychiatric Genetics*. 2012; 159B:5–12. [PubMed: 22095612]
- Tourette Syndrome Association International Consortium for Genetics [TSAIG]. Genome scan for Tourette Disorder in affected-sibling-pair and multigenerational families. *American Journal of Human Genetics*. 2007; 80:265–272. [PubMed: 17304708]
- Visscher PM, Brown MA, McCarthy MI, Yang J. Five years of GWAS discovery. *American Journal of Human Genetics*. 2012; 90:7–24. [PubMed: 22243964]
- Whitehead LC. Methodological and ethical issues in Internet-mediated research in the field of health: an integrated review of the literature. *Soc Sci Med*. 2007; 65:782–791. [PubMed: 17512105]
- Young NL, Varni JW, Snider L, McCormick A, Sawatzky B, Scott M, King G, Hetherington R, Sears E, Nicholas D. The Internet is valid and reliable for child-report: an example using the Activities Scale for Kids (ASK) and the Pediatric Quality of Life Inventory (PedsQL). *J Clin Epidemiol*. 2009; 62:314–312. [PubMed: 18834710]

Figure 1.

Flow chart of web-based recruitment, enrollment, and completion

¹This number reflects individuals who were contacted directly, not the unknown number of individuals recruited via Facebook or website postings.

Figure 2.
Completion rates by age group

Figure 3.
Cumulative number of completed cases by month for web-based and clinic-based studies

Table 1
Characteristics of web-based study enrolled participants as a group and by completion status

Categorical Variables	Total Enrolled		Non-Completer		Completer		Non-Completers compared to Completers	P-value
	Frequency (%)		Frequency (%)		Frequency (%)			
Total Participants	987 (100.0)		356 (36.1)		631 (63.9)		NA	
Participants by Site:								
MGH	694 (70.3)		263 (37.9)		431 (62.1)			
UCSF	293 (29.8)		93 (31.7)		200 (68.3)		.077	completion rate MGH v. UCSF)
Sex:								
Female	294 (29.7)		102 (28.7)		192 (30.4)		.608	
Male	693 (70.2)		254 (71.3)		439 (69.6)			
Race:								
Am. Indian/Alaska Native	1 (.1)		1 (.3)		0		.361 [/]	
Asian	10 (1.0)		1 (.3)		9 (1.4)		.105 [/]	
African American	4 (.4)		1 (.3)		3 (.5)		1.00 [/]	
More than one race	74 (7.5)		26 (7.3)		48 (7.6)		.962	
Unknown/Other	18 (1.8)		10 (2.8)		8 (1.3)		.136	
White	880 (89.2)		317 (89.0)		563 (89.2)		1.00	
Ethnicity:								
Hispanic	69 (7.0)		33 (9.3)		36 (5.7)		.048*	
ADHD Diagnosis								
Yes	453 (45.9)		171 (48.0)		282 (44.7)		.499	
Don't Know	58 (5.9)		18 (5.1)		40 (6.3)			
OCD Diagnosis								
Yes	507 (51.4)		188 (52.8)		319 (50.6)		.737	
Don't Know	73 (7.4)		27 (7.6)		46 (7.3)			
ADHD & OCD Diagnoses								
Yes	295 (29.9)		118 (33.1)		177 (28.1)		.244	
Don't Know	107 (10.8)		37 (10.4)		70 (11.1)			
First Degree Family Member with TS or another tic disorder (yes)	289 (29.3)		114 (32.0)		175 (27.7)		.177	
Any Family Member with TS or another tic disorder (yes)	445 (45.1)		166 (46.6)		279 (44.2)		.536	
Continuous Variables	Mean (SD)		Mean (SD)		Mean (SD)		P-value	

Categorical Variables	Total Enrolled		Non-Completer		Completer		Non-Completers compared to Completers	
	Frequency (%)		Frequency (%)		Frequency (%)		P-value	
Age	23.44 (SD=16.10)		19.33 (SD=13.38)		25.77 (SD=17.03)		.000**	
Income by Participant's Zip Code	63,626.96 (SD=22,668.93)		62,685.79 (SD=23,080.13)		64,160.34 (SD=22,434.19)		.335	

J Fisher's Exact Test used due to small cell size

* Significant at the $p \leq .05$ level

** Significant at the $p \leq .01$ level

Table II

Reasons for study withdrawal by enrolled web-based participants (n=69)

Reason	Percentage (Number)
Fear of phlebotomy by minor proband	36.2 (25)
Changed mind	14.5 (10)
Concern about DNA privacy	11.6 (8)
Too busy	10.1 (7)
Minor declined participation	8.7 (6)
Other	15.9 (11)
Declined to provide reason	2.9 (2)

Table III

Predictors of web-based study completion

Variable	Wald Statistic	df	Significance level	Odds Ratio
Male	.232	1	.630	1.08
Higher census-derived income	1.060	1	.303	1.00
Non-white race	1.003	1	.317	1.29
ADHD diagnosis	.685	1	.408	1.20
OCD diagnosis	.330	1	.566	1.12
Both ADHD & OCD diagnoses	2.507	1	.113	.62
Ages 6–8 compared to 26+	45.730	1	<.001*	.17
Ages 9–12 compared to 26+	13.849	1	<.001*	.47
Ages 13–17 compared to 26+	3.832	1	.050	.67
Ages 18–25 compared to 26+	21.671	1	<.001*	.35

**
Significant at the $p < .005$ level

Table IV

Comparison of web-based and clinic-based study participant characteristics

Variable	Web-based Participants (n=631)	Clinic-based participants (n=238)	P-value
Female	30.4% (n=192)	19.3% (n=46)	.001*
Male	69.6% (n=439)	80.7% (n=192)	
Non-white	10.8% (n=68)	7.2% (n=17)	.328
White	89.2% (n=563)	92.8% (n=219)	
Hispanic	5.7% (n=36)	12.4% (n=9)	.325
Non-Hispanic	94.3% (n=592)	96.2% (n=229)	
ADHD diagnosis	44.7% (n=282)	48.1% (n=114)	.069 ^a
No ADHD diagnosis	55.3% (n=349) ^a	51.9% (n=123)	
OCD diagnosis	50.6% (n=319)	47.7% (n=113)	.497 ^a
No OCD diagnosis	49.4% (n=312) ^a	52.3% (124)	
History of TS or another tic disorder in 1 st degree relative	27.7% (n=175)	34.6% (n=82)	.059 ^a
No History of TS or another tic disorder in 1 st degree relative	72.3% (n=456)	65.4% (n=155)	
Mean Age	25.77 (17.03)	20.11 (13.85)	<.001*

* Significant at the p=<.01 level

^a "Don't know" responses included with "no" responses due to small cell sizes