Published in final edited form as:
Twin Res Hum Genet. 2006 December ; 9(6): 998–1005. doi:10.1375/183242706779462633.

PennTwins: A Population-Based Cohort for Twin Studies

Emil F. Coccaro and **Kristen C. Jacobson**
Clinical Neuroscience & Psychopharmacology Research Unit, Department of Psychiatry, The University of Chicago, United States of America

Abstract

The current article describes the creation and composition of the PennTwins Cohort and provides details on the demographic characteristics of the sample. The PennTwins Cohort was developed using a population-based method of ascertainment and currently has 9401 28- to 47-year-old twins, including 2225 confirmed twin pairs and 4951 twins whose co-twins have not yet registered. Zygosity data have been used to identify 919 monozygotic, 634 same-sex dizygotic, and 445 opposite-sex dizygotic pairs. GeoCode data on gender, age, and certain demographic characteristics were obtained for the addresses of all twins who were mailed invitations to be part of the cohort. Analysis of the available data show only very small differences between twins who are currently part of the PennTwins Cohort and potential twins who either did not respond to recruitment or who could not be located. Similarly, only very small demographic differences exist between twins from complete pairs and twins whose co-twins are not yet registered, and there are no differences across zygosity. Thus, despite a relatively low overall response rate (12% of all twins born in Pennsylvania from 1959–1978), there is no evidence that the sample differs in any meaningful respect from the larger population.

The PennTwins Cohort was begun in 1996 and consists of a population-based sample of all twins born in the state of Pennsylvania between 1959 and 1978. The primary purpose of this article is to describe the methods used to create the cohort, and to present data on the current sample composition. As part of the PennTwins project, data were obtained on gender, age, and demographic characteristics using geocoded information based on twins' addresses. Thus, the article further provides demographic information on the PennTwins Cohort, and examines potential bias due to nonparticipation.

Methods and Results

Cohort Establishment and Recruitment of Participants

The development of the PennTwins Cohort began in 1996 when one of the authors (EFC) approached the Pennsylvania Department of Health (PA-DOH) regarding the development of a cohort of adult twins born in the Commonwealth of Pennsylvania. Since the PA-DOH had computerized birth records beginning from 1959, and the aim was to recruit adult twin pairs of at least 18 years in age, we chose to survey only birth records dating from 1959 through 1978. Initial permission was received from the PA-DOH's Division of Vital

Address for correspondence: Emil F. Coccaro, Department of Psychiatry MC#3077, The University of Chicago, 5841 South Maryland Avenue, Chicago, IL 60637, USA. ecoccaro@yoda.bsd.uchicago.edu.

Statistics to conduct a search of all computerized birth records (from 1959–1978) to obtain the names of 'matched' individuals who were born to the same parents on the same day in the Commonwealth of Pennsylvania, and were therefore likely to be part of a twin pair. This led to a list of 77,012 individuals that was then cross-referenced against 4.2 million active driving license records on file at the Pennsylvania Department of Transportation (PennDOT). This resulted in an initial address list of 30,801 individuals (i.e., 39.1% of the original search list) produced by April 1997. It should be noted that addresses could not be obtained if the twin did not have an active driver's license or a state ID card for the Commonwealth of Pennsylvania in 1996 to 1997. Furthermore, the list was restricted to the twin's name as listed on the birth certificate, and therefore could not include twins whose names may have changed due to adoption or marriage. However, the list of potential twins included all twins with known addresses through PennDOT, regardless of whether the address of their co-twin was known.

In May 1997, a private mail-house (AdMail, Cherry Hill, New Jersey) was employed to mail letters of 'consent-to-contact' to each of these 30,801 individuals. By prior agreement, to maintain confidentiality the PennTwins Program was never given this list of names by PA-DOH. The mailed packets contained: (1) a cover letter signed by Secretary of Health of the Commonwealth of Pennsylvania, (2) a general information sheet for twin candidates and, (3) an institutional review board-approved consent document outlining the fact that we were developing a twin cohort for twin studies and requesting that individuals indicate whether they wished to be contacted by us. Those agreeing to further contact completed the contact information form on the back of the consent document. Those who did not wish to be contacted signed their name and gave their birth date and were assured that no further contact would be made in the future for any reason. Signed consent documents were sent directly to the PennTwins Program. Consent-to-contact letters returned due to an invalid address were subjected to a secondary search with a credit agency (i.e., TransUnion) in an attempt to determine a more current address. New contact addresses were found in about 20% of cases and such cases were then sent a new consent-to-contact mailing packet. In order to increase participation, new consent-to-contact packets were mailed out to nonresponders and to those whose addresses had changed in October 1997, and again in March, 1998.

Response Rates

Despite the use of an outside tracking firm and three separate mailings, we were unable to deliver the consent-to-contact packets to 7535 individuals (24.5% of the sample), due to unknown addresses. An additional 357 twins (1.2%) were ineligible to participate in the cohort, due to death ($N = 24$) or death of co-twin ($N = 278$), medical ineligibility (i.e., mental retardation, $N = 20$), or because twins were currently living out of state ($N = 24$), were on active duty in the military ($N = 10$), or were incarcerated ($N = 1$). An additional 12,417 twins (40.3%) never responded to our mailings, and were considered uninterested in the study. Less than 4% of the sample ($N = 1151$) actively refused to participate in the cohort. Overall, we received agreement to participate from 9341 individual twins. This number represents 30.3% of the original sample of 30,801. However, if twins who were

ineligible (N = 357) or could not be located (N = 7535) are discounted, the response rate among eligible participants increases to 40.8%.

Demographic Comparisons of Twin Cohort Participants and Nonparticipants

In order to gain insight as to whether the identified PennTwins participants differed in some systematic fashion from those twins who did not respond to our mailings or who could not be located, the PA-DOH provided exact birth dates for both participating and nonparticipating twins, and further allowed the PennTwins Program to have access to the addresses (without names) of the participating and nonparticipating twins so that we could geocode these individuals. As part of the initial agreement with PA-DOH, however, we were not allowed to have similar data for the 1151 twins who returned the consent-to-contact form but declined to take part in the PennTwins Cohort. Nevertheless, we were able to obtain relevant demographic data from 96.3% of the original pool of 30,801 individuals.

A geocode is a very specific 2-dimensional location (e.g., a precise latitude and longitude) for a specific address. Once the geocode was determined from the actual address of the twin, the geocode for each twin was matched with the census block group data associated with that specific geocode. This work was performed by a third party with expertise in these kinds of analyses (Claritas Inc., Ithaca, New York). Geocodes were associated with census track characteristics for the distribution of various demographic variables (e.g., racial composition, distribution of educational attainment, martial status, and so on) in the specific census block group for every subject. A census block group represents approximately 1500 geographically related individuals (e.g., one city block in urban areas). Thus, although we do not have information regarding the specific demographics of our participants versus nonparticipants, geocode data provide estimates of the specific demographic characteristics among all individuals living within a small radius of our participating and nonparticipating twins.

Table 1 presents the mean ages and demographic characteristics of those individuals in the cohort with those individuals who are not part of the cohort. Given the large sample size (i.e., N = 9341 participants and N = 20,309 nonparticipants), all comparisons across group were statistically significant (i.e., $p < .001$). Therefore, in order to interpret the *practical significance* of any group differences, the effect size as calculated by Cohen's d index is included in the table along with the means and standard deviations (Cohen, 1988). The purpose of using the Cohen's d index is to present group differences in terms of standard deviation units, thus allowing for the examination of practical effects as well as allowing for potential comparisons across studies. 'Small' effects are typically defined by a d index between 0.20 and 0.50, with d indices between 0.50 and 0.80 constituting 'medium' effects, and 'large' effects consisting of a d index of greater than 0.80 (Cohen, 1988). Because it is assumed that a nonresponse may be indicative of a passive refusal to participate, additional analyses (available upon request), further divided those individuals who are not part of the cohort into individuals who could not be located versus individuals who did not respond to the mailings. Effect sizes from these latter comparisons were all negligible (all Cohen's $d \leq 0.20$); thus, nonresponders and those who could not be located have been combined into a single group of 'nonparticipants'.

As can be seen in Table 1, only a few of the group differences in age or demographic characteristics can be considered even 'small' effects. On average, participants in the cohort are approximately 1.1 years younger (pooled *SD* = 5.7 years) than nonparticipants (*d* = 0.19). There are no meaningful group differences in the racial composition (all effect sizes ≤ 0.11), marital status (all effect sizes ≤ 0.14), or unemployment rate (effect size = 0.13) among individuals living in the same geocode area. Importantly, there are no practical differences across group in the number of psychiatric or medical diagnoses in the geocode area (listed as number of cases per 1000 individuals; effect sizes = .02–.04). The only potentially meaningful differences among participants and nonparticipants concern variables relating socioeconomic status, that is, income and educational attainment. Participants in the PennTwins Cohort live in areas that are defined by slightly higher incomes. However, although the effect size for average individual income approaches the definition of a 'small' effect (*d* = 0.19), it is noted that the difference across groups is only ~$2,000 (*M* = $21,175 among participants and *M* = $19,180 among nonparticipants, pooled *SD* = $10,482). Similarly, PennTwins participants live in areas with higher proportions of individuals who have received some college education (effect size = 0.25), but there is less than a 5% difference across groups (*M* = 40.6% among participants and 35.8% among nonparticipants, pooled *SD* = 18.8%). In summary, while the data presented in Table 1 suggest that there may be potential ascertainment biases among twins participating in the PennTwins Cohort, these potential biases are modest at best.

The Current PennTwins Cohort

As of May, 2006, there are 9401 individuals in the PennTwins Cohort.[1] Figure 1 shows the composition of the current PennTwins Cohort. Beginning in 1997, Zygosity Questionnaires (ZQ) were sent to all twins from whom we obtained consent-to-contact. The ZQ used contained a series of standard questions concerning perceived similarity that have been commonly used in twin research, and have further been shown to discriminate same-sex monozygotic (MZ) and dizygotic (DZ) twin pairs with a high degree of accuracy in other sample (approximately 95%; Eisen et al., 1989; Magnus et al., 1983). The ZQ also contained questions concerning the name and address of the co-twin (allowing us to 'pair up' the individual twins), and questions regarding age, race/ethnicity, marital status, educational attainment of parents, and current income.

As can be seen in Figure 1, we have currently received completed ZQs from 75.9% (*N* = 7131) of the participants in the PennTwins Cohort (with a minority of twins completing the ZQ through phone interview). Of these 7131 twins, 59.0% (*N* = 4204) have participating co-twins who have also returned the ZQ, and the remaining 41.0% (*N* = 2927) are twins whose co-twins have not yet returned the ZQ. Of the 2102 'complete' twin pairs, we have used the information on gender and the responses to the similarity questions from the ZQ to assign zygosity to more than 95% of the sample. Slightly less than one half of the pairs (43.7%) are MZ twins (*N* = 919 pairs), 30.2% are same-sex DZ pairs (*N* = 634 pairs), and 21.2% are

[1]The current *N* of 9401 in 2006 represents a 'net gain' of 60 twins since the collection of the geocode data in 2000. This is largely due to the inclusion of new twins who hear about our cohort through other mechanisms (e.g., through the newsletters or word-of-mouth) and call in to volunteer.

opposite-sex DZ pairs (N = 446 pairs). An additional 4.9% (N = 103 pairs) currently have unconfirmed zygosity, due to the small degree of overlap in the distributions of similarity scores for same-sex MZ and DZ twins and/or to missing data concerning gender. We have obtained DNA data from at least 50 of these 103 pairs, and are in the process of obtaining definitive zygosity information through genotyping analysis.

Among those twins who have not yet returned the ZQ, we have been able to assign a pair id to 663 twins, as we have confirmed information on the name of their co-twin. The remaining 1607 twins either did not provide us with the name of their co-twin and/or have co-twins who have not actively consented to participate in the cohort. Thus, these twins have not been assigned a pair id. Of the 663 twins with an assigned pair id, at least 246 twins have known co-twins who have also not yet returned the ZQ (i.e., 123 pairs concordant for 'nonresponse' to the ZQ). Although we know the sex-type for most of these pairs (see Figure 1), we will not assign zygosity until we receive a ZQ back from at least one of the twins. The 447 'unpaired' twins with an assigned pair id number are likely twins whose co-twins have already returned the ZQ. We are in the process of 'pairing up' these twins with the 2927 'unpaired' twins who have returned the ZQ (represented by the dashed double-headed arrow in Figure 1). Thus, in total, we have at least 2225 'complete' twin pairs in the PennTwins Cohort (i.e., 2102 pairs who have returned ZQs and 123 pairs who have not returned the ZQ), which represents 47.3% of the total PennTwins Cohort.

Bias Due to Nonresponse of Co-Twin

One method of testing for response bias using twin samples is to examine whether twins whose co-twins participate in the cohort (i.e., 'paired' twins) differ from twins whose co-twins have not participated (i.e., 'unpaired' twins). Table 2 presents the mean age, racial distribution, and marital status among these two groups, for the combined genders, as well as separately for males and females.[2] In contrast to the above geocode analyses, these analyses are based on twins' self-reports, so are representative of the characteristics of the twins themselves. The analyses have necessarily been restricted to the 7131 twins who returned the ZQ, as the age and demographic data were obtained through the questionnaire. We have excluded a total of 23 individuals who have missing or questionable data on gender, so the final N is 4190 'paired' twins versus 2918 'unpaired' twins.

Although the overall sample is split approximately equally among males and females (48.5% male, 51.5% female), twins whose co-twins have also participated are slightly more likely than twins whose co-twins have not participated to be female (54.7% of the 'paired' sample versus 46.9% of the 'unpaired' sample, $\chi^2_{(1)} = 42.26$, $p < .001$). As Table 2 indicates, twins whose co-twins participated are slightly younger ($F_{(1,7102)} = 228.0$, $p < .001$) and more likely to be Caucasian ($\chi^2_{(1)} = 66.4$, $p < .001$) than twins whose co-twins did not participate. These group differences were found for both male and female twins. There were also significant differences between paired and unpaired twins in marital status.

[2]All individual twins have been used in these analyses. For analysis of group differences in age, corrections for correlated observations within twin pairs were applied to obtain corrected standard errors and unbiased significant tests. For the chi-square analyses of nominal variables (i.e., race and marital status), Fisher's exact test was performed, which involves no parametric assumptions or assumptions about sample independence.

'Unpaired' twins are slightly more likely to be never married (50.7%) versus married (42.2%), whereas 'paired' twins are nearly twice as likely to be never married versus married (63.1% vs. 33.1%; $\chi^2_{(2)} = 115.03$, $p < .001$).

Gender Differences

Using the same sample of 7131 twins for whom we had valid self-report demographic information (see above), we also examined main effects of gender on the age, racial/ethnic background, and marital status of our participants (results not shown). Male participants are slightly older ($M = 38.3$, $SD = 6.0$) than female participants ($M = 35.6$, $SD = 5.6$; $F(1, 7102) = 312.8$, $p < .001$), and a slightly higher proportion of males (94.5%) are Caucasian compared to females (91.0%; $\chi^2_{(1)} = 38.3$, $p < .001$). Among male twins, approximately equal proportions report being 'never married' (47.2%) and 'married' (46.7%). In contrast, female twins are more than twice as likely (68.2%) to report that they have never been married than to report that they are currently married (27.7%; $\chi^2_{(2)} = 312.6$, $p < .001$).

Zygosity Differences

Table 3 presents the age, gender, race, and marital status distribution among the 1999 twin pairs ($N = 3998$ individuals) for whom we have confirmed zygosity (> 95.0% of the total sample of complete twin pairs). Means and percentages are presented separately for MZ, same-sex dizygotic (DZS), and opposite-sex dizygotic (DZO) twins. Analyses revealed that the three groups were highly similar with respect to age, marital status, and ethnicity. The only significant difference was with respect to gender distribution. Among same-sex twin pairs (i.e., MZ and DZS), there was a slight preponderance of female–female twin pairs relative to male–male twin pairs (55.8%–56.5% female–female), whereas the gender distribution among the complete DZO pairs was (by definition) split evenly by gender.

Discussion

The PennTwins cohort was developed using a population-based method of ascertainment to identify twins born in the state of Pennsylvania from 1959 through 1978. The initial ascertainment of 9341 twins reflects only 12.1% of the 77,012 individuals who were identified through birth records as part of a likely twin pair. However, when one discounts the number of twins for whom we could not obtain valid addresses and the small number of twins who were ineligible, the response rate rises to 40.8%.

Potential Bias in the PennTwins Cohort

A strength of this cohort is that we were able to obtain geocoded demographic data for both the participants in our cohort and the nonparticipants identified through PennDOT records (including the $N = 7535$ twins from the PennDOT data base who had invalid addresses), which allowed us to specifically look for indications of response bias. Analysis of the geocoded demographic data revealed very few differences of *practical effect* among participants and nonparticipants. Of 14 indices examined, only two group differences met the criterion for a 'small' effect (i.e., Cohen's $d \geq 0.20$). Overall, the average effect size was .14, reflecting a mean difference of approximately one seventh of a standard deviation unit. Importantly, there were no differences in the underlying racial/ethnic composition,

marital status, unemployment rates, or prevalence of psychiatric or medical diagnoses across groups. There was some indication that the twins who participated in our cohort came from slightly more advantaged socioeconomic environments than nonparticipants, as the effect sizes greater than .15 concerned annual per capita income among the census-block regions and average education levels of the census-block region. Again, however, these differences were small (i.e., one quarter to one fifth of a standard deviation). Thus, based on the analyses of the geocoded data, we could find no indication that the participants in our cohort differed meaningfully from nonparticipants.

Although the geocoded data are only approximations of an individual's specific demographic characteristics, we assume that they are largely reflective of the demographic characteristics of our participants and nonparticipants. Of potentially greater concern is the fact that we were unable to obtain geocode data for the substantial number of twins (N = 46,211) who were not initially linked to the PennDOT data base, and therefore never had the opportunity to receive our consent-to-contact package. Twins who moved from Pennsylvania, and had no record with PennDOT, could never be contacted. The developers of the PennTwins Cohort were not allowed to have names and addresses from the original birth records and, therefore, it was impossible that even a small number of these individuals could be traced and contacted. This suggests that our sample may be biased in terms of individuals with lower mobility. Likewise, potential participants who did not have a current state-issued driver's license or state ID card were not in the PennDOT database, and therefore could not be linked up with birth records. Thus, our subject pool may underrepresent those at the extreme lower ends of the poverty scale and the homeless. Finally, females who had married and changed their legal name on their PennDOT driver's license could not be contacted as we were not given permission by either PA-DOH or PennDOT to follow-up these individuals to obtain their current names. This latter constraint has almost certainly introduced a bias in our sample, as only 27.7% of the female twins in our current cohort who have completed self-reports of marital status reported that they are currently married, despite the fact that the average age among these female twins is approximately 37. Additionally, female twins were significantly more likely to report that they have never married compared with male twins in the cohort. Nevertheless, given our large initial sample size, we have nearly 600 'paired' females and 400 'unpaired' females in our sample who *are* currently married; thus, future analyses can address the extent to which differences in the distribution of marital status may affect our results.

Finally, we note that overall, there are very small differences between paired and unpaired twins with regard to basic characteristics such as age, gender, ethnicity, and marital status, and virtually no differences exist across the different zygosity groups. Moreover, compared to the standard 'two-thirds' rule observed in many twin studies (Lykken et al., 1987), we have nearly equal numbers of male and female participants in our cohort (i.e., 51.5% of the 7108 twins with known zygosity are female). Female–female twins are only slightly overrepresented among our same-sex twin pairs (56.1% of same-sex twin pairs are female), as are MZ twins compared to same-sex DZ twins (i.e., 59.0% of same-sex twins are MZ).

Additional Caveats

Like the majority of other twin registries, the PennTwins Cohort is comprised primarily of twins who indicate that their ethnic/racial background is Caucasian (92.7%). Nevertheless, this figure is only slightly higher than the average 85% to 86% Caucasian statistic obtained through the geocode data, indicating that our cohort is reasonably representative of the racial/ethnic background of the state of Pennsylvania. An additional limitation stems from the fact that the PA-DOH did not allow us to directly contact any potential twins before informed consent had been obtained. In particular, over one half of twins with presumably valid addressees simply never responded to the consent-to-contact packets, and the PennTwins Program was not able to later contact nonresponders to ascertain if such individuals were truly uninterested in being involved in a twin studies program (i.e., were refusers) or would be willing, if asked through more personal contact, to be part of such a cohort but had simply not responded to the mailings for other reasons (e.g., were busy at the time of mailing, lost the form, etc.). Thus, if it had been possible to track or contact the twins directly, it is likely that we would have been able to achieve a higher response rate, consistent with the higher response rate in other population-based studies which have been able to locate twins directly (e.g., the Minnesota Twin Registry, Krueger & Johnson, 2002). We recognize, however, that these constraints on subject ascertainment and recruitment were due to PA-DOH's charge to protect the rights of the general population. These constraints currently exist in a large number of states in the United States and do limit the ability of investigators to develop population-based cohorts of twins.

Maintenance and Further Development of the Cohort

In the fall of 1999, the developer of the PennTwins Cohort (EFC) took a position in the Department of Psychiatry at the University of Chicago. Thus, all PennTwins Program activities are currently being conducted in Chicago, Illinois. In addition to the conduct of the current research studies described below, the PennTwins Program engages in two primary maintenance activities: address updates using the National Change of Address program; and the production of a PennTwins Newsletter that is posted to all PennTwins participants at least yearly. This newsletter includes information about twin research in general and also shares information from our enrolled twins about their experiences in our program. In addition to providing us with current addresses, the PennTwins Newsletter has also allowed us to add new participants to our cohort, primarily through word of mouth. Further developments continue to occur as we work to pair up more twins by contacting the existing unpaired twins in order to obtain 'consent to contact' from the other twin sibling. We are also working to resolve all uncertain zygosity determination through DNA analysis.

Current PennTwin Studies

The PennTwins Program is in the process of completing its first large-scale study (to be completed in 2007). This study (funded by NIMH RO1 MH063262) focuses on the behavioral genetics of aggression and impulsivity and examines these constructs using a multimodal–multitrait method. Part 1 of the study (data collection now complete) involved two questionnaire mailings to all twin pairs identified in the PennTwins Cohort by 2004. Overall response rate to the questionnaire was very good (74.3% for individual responses

and 57.6% for twin pairwise responses). We have currently received questionnaires from over 3072 individual twins. These questionnaires provide basic self-report information on personality, health, and common behaviors, with a specific focus on aggression and antisocial behavior, and further contain a series of detailed questions concerning demographic and family environmental factors. Part 2 of the study (ongoing at this time) involves studying same-sex twin pairs in our laboratory so that analogue laboratory measures of aggression and impulsivity, as well as platelet measures of serotonin uptake, serotonin transporter binding, and serotonin transporter DNA polymorphisms can be added to the psychometric data collection begun with the questionnaire mailer. At this time, more than 120 same-sex twin pairs have visited our laboratories here in Chicago for this part of the study. Other pilot studies regarding the behavioral genetics of social-emotional information processing and of fMRI neuronal circuitry activity are beginning now.

Acknowledgments

Dr E. F. Coccaro is supported by NIMH RO1 MH063262; Dr K. C. Jacobson is supported by NIH K01 MH-068484. The authors wish, first, to thank Cindy S. Bergeman, PhD, for her advice, expertise on twin studies, and her collaborative efforts in determining the zygosity of the PennTwins. The authors also wish to thank Len Murrelle, PhD, Matt McGue, PhD, and Greg Carey, PhD, for their advice on the development of Twin Cohorts and Crystal Cole and the PennTwins staff for their direct work with the PennTwins. Finally, the authors wish to thank Drs A. Busjahn and Y.-M. Hur for their comments and suggestions for revision.

References

Cohen, J. Statistical power analysis for the behavioral sciences. 2. Hillsdale, NJ: Lawrence Earlbaum Associates; 1988.

Eisen S, Neuman D, Goldberg J, Rice J, True W. Determining zygosity in the Vietnam Era Twin Registry: An approach using questionnaires. Clinical Genetics. 1989; 35:423–432. [PubMed: 2736790]

Lykken DT, McGue M, Tellegen A. Recruitment bias in twin research: The rule of two-thirds reconsidered. Behavior Genetics. 1987; 17:343–362. [PubMed: 3675526]

Krueger RF, Johnson W. The Minnesota Twin Registry: Current status and future directions. Twin Research. 2002; 5:488–492. [PubMed: 12537882]

Magnus P, Berg K, Nance WE. Predicting zygosity in Norwegian twin pairs born 1915–1960. Clinical Genetics. 1983; 24:103–112. [PubMed: 6577993]

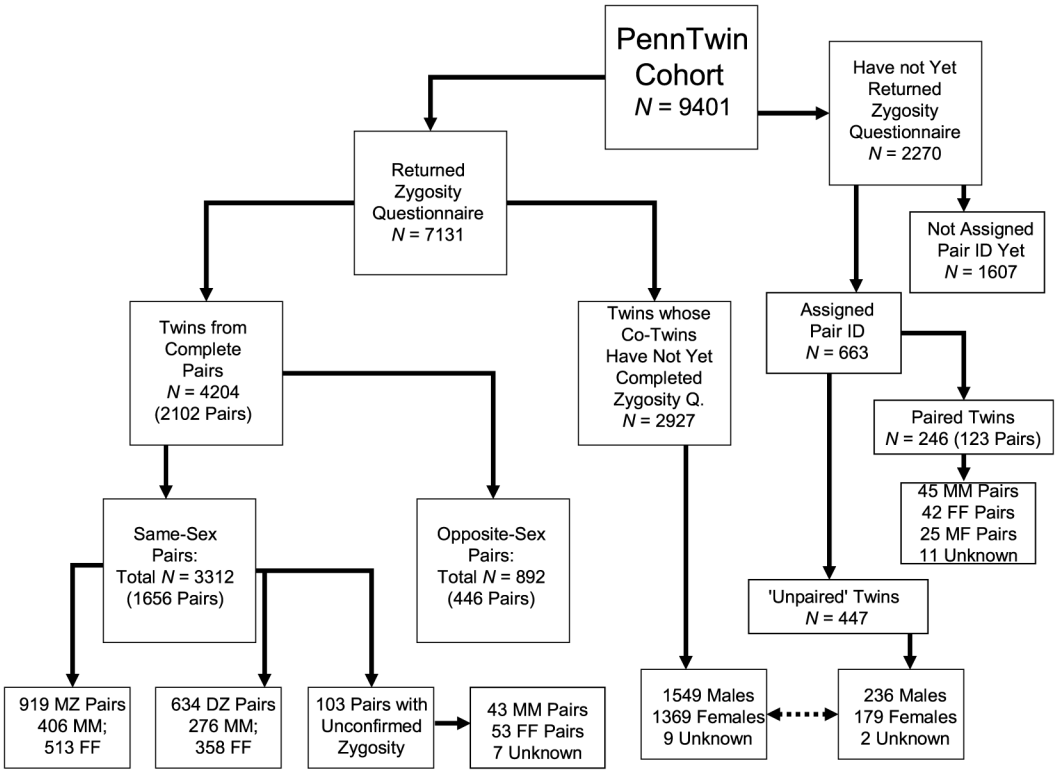

Figure 1.
Graphical depiction of the current PennTwins cohort.

Table 1

A Comparison of Age and Demographic Information Obtained through Geocoded Data Among Participants and Nonparticipants in the PennTwin Cohort

	Participants (*N* = 9341)	Nonparticipants (*N* = 20,309)	Cohen's *d* effect size
Age	37.0 ± 6.0	38.1 ± 5.6	0.19
Racial category			
% Caucasian	86.8 ± 22.7	85.0 ± 27.2	0.07
% African–American	8.5 ± 20.2	11.0 ± 25.2	0.11
% Hispanic and other	4.8 ± 8.5	4.0 ± 8.2	0.10
Marital status			
% never married	27.0 ± 11.2	26.8 ± 9.7	0.02
% currently married	55.5 ± 13.7	54.7 ± 13.4	0.06
% other	17.5 ± 6.7	18.5 ± 6.9	0.14
Unemployment rate (%)	2.7 ± 2.0	3.0 ± 2.3	0.13
Educational attainment			
% without H.S. degree	22.7 ± 12.5	25.3 ± 12.9	0.21
% H.S. degree, only	36.8 ± 12.2	38.9 ± 11.0	0.19
% some college	40.6 ± 20.0	35.8 ± 18.2	0.25
Household income	$43,529 ± 22,723	$ 40,313 ± 21,347	0.15
Per capita income	$21,175 ± 11,318	$19,180 ± 10,077	0.19
Psychiatric diagnoses (per 1000)	3.6 ± 0.4	3.6 ± 0.4	0.02
Medical diagnoses (per 1000)	8.8 ± 2.2	8.7 ± 1.9	0.04

Note: The 'other' category in race includes Hispanics, Asians, and any other racial/cthnic category that is not Caucasian or African–American. The 'other' group in marital status includes individuals who are separated, divorced, or widowed.

Table 2

Comparison of Age, Race, and Martial Status Among Paired and Unpaired PennTwin Twins with Zygosity Questionnaire Data

	'Unpaired' twins (*N* = 2918)			'Paired' twins (*N* = 4190)		
	Combined gender	**Males (*N* = 1549)**	**Females (*N* = 1369)**	**Combined gender**	**Males (*N* = 1896)**	**Females (*N* = 2294)**
Age	38.3 ± 5.6	39.3 ± 5.5	37.0 ± 5.5	36.0 ± 6.0	37.4 ± 6.2	34.8 ± 5.5
Race						
Caucasian	89.6%	92.2%	86.7%	94.8%	96.3%	93.5%
African–American	7.1%	4.8%	9.7%	3.8%	2.4%	5.0%
Other	3.3%	3.0%	3.6%	1.4%	1.3%	1.5%
Marital status						
Never married	50.7%	38.7%	64.4%	63.1%	54.1%	70.4%
Married	42.2%	53.3%	29.6%	33.1%	41.2%	26.5%
Other	7.1%	8.0%	6.0%	3.8%	4.7%	3.1%

Note: The 'other' category in race includes Hispanics, Asians, and any other racial/ethnic category that is not Caucasian or African–American. The 'other' group in marital status includes individuals who are separated, divorced, or widowed.

Table 3

Comparison of Age, Gender, Race, and Martial Status Across MZ and DZ Twins

	Same-sex MZ twins (MZ) (*n* = 1838)	Same-sex DZ twins (DZS) (*n* = 1268)	Opposite-sex DZ twins (DZO) (*n* = 892)	Comparisons[1]
Age (years)	35.9 ± 6.0	36.1 ± 6.1	35.9 ± 5.9	MZ = DZS = DZO
Gender				
Male	44.2%	43.5%	50.0%	DZO > MZ = DZS
Female	55.8%	56.5%	50.0%	
Race				
Caucasian	96.0%	94.2%	94.7%	MZ = DZS = DZO
African–American	3.0%	4.0%	4.2%	
Other	1.0%	1.8%	1.1%	
Marital Status				
Never married	62.7%	63.1%	65.4%	MZ = DZS = DZO
Married	33.5%	33.2%	30.9%	
Other	3.8%	3.7%	3.7%	

Note: The ‘other’ category in race includes Hispanics, Asians, and any other racial/ethnic category that is not Caucasian or African–American. The ‘other’ group in marital status includes individuals who are separated, divorced, or widowed.

[1] Differences between groups determined by post-hoc testing ($p < .05$) after ANOVA or chi-square statistics revealed an overall $p < .05$ where statistically significant.