

Determining Zygosity in the Vietnam Era Twin Registry: An Update

Christopher W. Forsberg,¹ Jack Goldberg,^{1,2} Jennifer Sporleder,¹ and Nicholas L. Smith^{1,2,3}

¹ Vietnam Era Twin Registry, Seattle Epidemiologic Research and Information Center, Department of Veterans Affairs Office of Research and Development, Seattle, WA, United States of America

² Department of Epidemiology, University of Washington, United States of America

³ Group Health Research Institute, Group Health Cooperative, Seattle, WA, United States of America

Our work assessed the accuracy of the original zygosity classification in the Vietnam Era Twin (VET) Registry using new information from DNA markers on a subset of participants. We then constructed an updated zygosity classification algorithm. The VET Registry includes 7,375 male–male twin pairs who served in the military during the Vietnam era. During the mid-1980s 4,774 twin pairs completed a zygosity questionnaire of 20 items. Additionally, military record information, including blood group, was available. Items from the zygosity questionnaire and blood group were used in the original zygosity classification. Between 1990–2009 DNA was obtained from 612 twin pairs and concordance between co-twins was used to classify zygosity. Next logistic regression was used to construct predicted probabilities of zygosity using items from the zygosity questionnaire with this subsample. All twins were reclassified according to the new zygosity prediction model and compared with the original zygosity assignment. The original and new predicted probabilities of zygosity were highly correlated ($r = 0.962$) and concordance for the classification of zygosity was similarly high ($\kappa = 0.936$). Errors in the original zygosity assignment were primarily due to monozygotic twins that were misclassified as dizygotic based on military record blood group data. Removing the military record blood group data markedly improved the accuracy of the original classification. Zygosity assignment based on a zygosity questionnaire was highly predictive of DNA-based zygosity. Augmentation of such a zygosity classification from administrative data, military records, or other records, should be done with caution.

Keywords: Questionnaire Zygosity Determination, Vietnam Era Twin Registry, Twin Study

Twins provide a unique and valuable resource for research into the genetic and non-genetic contributions of a vast array of mental and physical health conditions. The Vietnam Era Twin (VET) Registry is composed of 7,375 adult male–male twin pairs born between 1939 and 1955 who both served on active duty during the Vietnam Era (1964–1975; Eisen et al.,

1989). The VET Registry serves as a national resource for studying the influence of military deployment on the health and wellbeing of Veterans of the Vietnam generation and for studying the health of aging males in general (Henderson et al., 1990). Numerous individual projects have studied the genetic and non-genetic factors of post-traumatic stress disorder (PTSD), substance use, cardiovascular disease, and aging using the VET Registry (Goldberg et al., 2002). Virtually all these studies require an accurate measure of twin zygosity.

Eisen and colleagues (1989) proposed a zygosity assignment method for the VET Registry very similar to one previously used by registries in Norway. The method uses answers reported by both twins for a series of twin-likeness questions to predict zygosity (Magnus et al., 1983). At the inception of the Registry, blood was not available from twins so the definitive gold-standard of zygosity assignment using molecular markers could not be used. The VET Registry used separate zygosity prediction models for whites and non-whites, and used blood group information from military records to override the prediction models and assign a twin pair as dizygotic (DZ) when the blood groups were discordant.

In the 20-plus years since the inception of the Registry, an array of additional data from the VET Registry members were collected by studies using the VET Registry as a source for twin research (Goldberg et al., 2002). Through these studies the VET Registry has acquired DNA on more than 600 twin pairs. The purpose of this paper is to use these DNA samples to validate the zygosity assignment in a subset of the VET Registry with molecular information on zygosity (referred to as DNA-based zygosity) and then to use this information to improve the method for assigning zygosity in the absence of molecular information.

Received 5 June, 2010; accepted 30 July, 2010.

Address for correspondence: Christopher W. Forsberg, Seattle Epidemiologic Research & Information Center, VA Puget Sound Health Care System, 1100 Olive Way, Suite 1400, Seattle, WA 98101. E-mail: Christopher.Forsberg@va.gov

Methods

Description of the VET Registry

The VET Registry consist of 7,375 male-male twin pairs born between 1939 and 1955 who both served on active duty in the US military during the Vietnam era (1964–1975). The Registry was constructed from a computerized search of Department of Defense military records and the Department of Veterans Affairs databases during the early 1980s. A complete description of the construction of the VET Registry is available elsewhere (Eisen et al., 1987).

Questionnaire Measures of Twin Similarity

In the mid-1980s a survey was mailed to VET Registry members that contained a 20-item zygosity questionnaire. Questions about zygosity can be grouped into four sections: (a) self-assessment of zygosity (1 item); (b) overall similarity as children, the ‘alike as peas in a pod’ question (1 item); (c) the frequency they were confused by family and others (6 items); and (d) feature and aptitude similarities as children (12 items). These questions were similar to those used by Magnus et al., (1983). A total of 4,774 twin pairs returned usable responses from both brothers.

Responses to the zygosity questionnaire were individually scored as follows: a response indicating similarity between the twins was coded as 1, a response indicating dissimilarity was coded as -1, and all other responses were coded 0. The intra-pair mean for each question item was then calculated and used in the development of the zygosity classification algorithm (Eisen et al., 1989).

DNA Determination of Zygosity

DNA was available from 612 complete twin pairs who had participated in an in-person VET Registry study in the time period 1990–2009. Using the DNA samples, zygosity was determined by comparing 15–28 microsatellite markers, all of which have at least 4 different alleles. If a pair matched at all the microsatellite markers it was assigned as a monozygotic (MZ), otherwise the twin pair was classified DZ. The minimum number of discordancies observed in a pair classified as DZ was five microsatellite markers. All DNA zygosity testing was done using an ABI 3100 Genetic Analyzer at the Geriatric Research Education and Clinical Center of the VA Puget Sound Health Care System.

Original Zygosity Classification from Questionnaire

The methodology for assigning zygosity based on questionnaire data by the VET Registry has been previously reported (Eisen et al., 1989). In summary, first the coefficients from the Norwegian twin registry were applied (Magnus et al., 1983) to the intra-pair mean scores from the 20 VET Registry zygosity question items. This was used to estimate the probability that a pair was MZ. A subsample of twin pairs at the extremes of the distribution of estimated probabilities was identified; those with a probability of > 0.88 were assumed to be MZ and those with a probability <

0.12 were assumed to be DZ. These pairs formed a ‘silver standard’ for the development of a new prediction algorithm for the VET Registry. Using a stepwise discriminant analysis to predict our silver standard zygosity, the questions that best predicted zygosity were identified. Separate algorithms for white and non-white twin pairs were developed.

The third step in the zygosity assignment process was to assign zygosity to all twins in the VET Registry based on the predicted probabilities from our algorithm. Those with MZ probabilities of >0.95 were classified MZ and those with MZ probabilities of < 0.55 were classified as DZ. We also had an indeterminate zygosity (IZ) group of twins who could not be classified as MZ or DZ, those with a probability between 0.55 and 0.95. This group constituted 3.1% of the twin pairs. As a final step we made use of the blood group (ABO and Rh factor) data available from the military records; when blood group was discordant, a pair was classified as DZ regardless of the predicted probabilities.

Updating Zygosity Classification

We developed a new predictive algorithm using responses to the zygosity questionnaire items and DNA-based zygosity among the pairs with DNA-based zygosity available. Logistic regression was used to determine which items from the zygosity questionnaire and the previously mentioned intra-pair scoring method were the most highly predictive of zygosity. The predictive algorithm was applied to all twin pairs with valid zygosity questionnaire responses. In the updated zygosity classification twin pairs with a probability of being MZ > 0.8 are classified as MZ, twin pairs with a probability of being MZ < 0.6 are classified as DZ, and those with probabilities between 0.8 and 0.6 are classified IZ. The results of the new algorithm are compared to both DNA zygosity and the 1993 algorithm using common statistical methods such as; sensitivity, specificity, overall agreement, and kappa.

Due to a limited number of non-white twin pairs with DNA available ($n = 26$) the new algorithm was only developed for white twin pairs from white twin pairs with DNA zygosity ($n = 586$). The analysis presented was conducted in Stata 11.0 (Stata Corporation, College Station, TX).

Results

Evaluation of the Original Zygosity Classification

Table 1 presents the comparison of the original zygosity classification to the DNA-based zygosity for the 612 twin pairs. The original zygosity had an overall accuracy of 92.6% (95% CI: 90.6–94.7), sensitivity for MZ of 89.3% (95% CI: 85.7–92.1), and specificity for MZ of 98.2% (95% CI: 95.3–99.4). The vast majority of the misclassification was MZ twin pairs being classified as DZ. Of the 265 pairs classified as DZ originally, 41 or 15.5% turned out to be MZ based on the DNA-based zygosity. In sharp con-

trast, of the pairs classified as MZ originally, only 4 pairs or 1.1% were actually DZ.

We examined the data further to determine the reasons for the relatively high misclassification of MZ zygosity as DZ, looking specifically at the blood group data from the military record. Table 2 presents the comparison of the original zygosity classification algorithm without using the military record blood group data to the DNA-based zygosity. Without the blood group data 231 twin pairs were classified DZ, and only 10 or 4.3% were incorrectly classified as DZ. Among those classified as non-IZ the original algorithm without the blood group data has an overall accuracy of 97.4% (95% CI: 96.1–98.6) and a kappa = 0.944 (p value < .0001). The sensitivity for MZ was 97.4% (95% CI: 95.1–98.7), and specificity for MZ was 97.4% (95% CI: 94.1–98.9).

Updated Classification Algorithm from DNA Analysis

A new predictive algorithm for VET Registry white twin pairs was developed using the DNA-based zygosity by stepwise exclusion. The following items were included in the prediction model: ‘alike as peas in a pod’, confusion by strangers, similarity in eye color, similarity in hair type, and similarity in hair color. For non-white twin pairs the original classification algorithm without the military record blood-group-overwrite was used. Table 3 presents the updated zygosity classification based on the regression model compared with DNA-based zygosity. When this classification is compared with the DNA-based zygosity among those classified as non-IZ the overall accuracy is 97.5% (95% CI: 96.3–98.8) with a kappa of 0.948 (p value < 0.0001). The sensitivity for MZ was 97.4% (95% CI: 95.1–98.7), and specificity for MZ was 97.8% (95% CI: 94.7–

Table 1

Comparison of Original VET Registry Zygosity Classification to DNA-Based Zygosity

Original zygosity classification	DNA zygosity		Total <i>N</i>
	MZ <i>N</i> (%)	DZ <i>N</i> (%)	
MZ	343 (89.3)	4 (1.8)	347
DZ	41 (10.7)	224 (98.2)	265
Total	384	228	612

Table 2

Comparison of Original VET Registry Zygosity Classification Without Blood Group Data to DNA-Based Zygosity

Original zygosity classification	DNA zygosity		Total <i>N</i>
	MZ <i>N</i> (%)	DZ <i>N</i> (%)	
MZ	373 (97.1)	6 (2.6)	379
DZ	10 (2.6)	221 (96.9)	231
IZ	1 (0.003)	1 (0.004)	2
Total	384	228	612

Table 3

Comparison of Updated Zygosity Classification to DNA-Determined Zygosity

Updated zygosity classification	DNA zygosity		Total <i>N</i>
	MZ <i>N</i> (%)	DZ <i>N</i> (%)	
MZ	372 (96.9)	5 (2.2)	377
DZ	10 (2.6)	223 (97.8)	233
IZ	2 (0.005)	0 (0.0)	2
Total	384	228	612

Table 4

Distribution of VET Registry Zygosity Among Twin Pairs Using the Updated Classification Algorithm and DNA-Based Zygosity

	DZ	MZ	IZ
	<i>N</i> (row %)	<i>N</i> (row %)	<i>N</i> (row %)
White	1733 (39.1)	2601 (58.6)	101 (2.3)
Non-white	127 (37.5)	206 (60.8)	6 (1.8)
Total	1860 (39.0)	2807 (58.8)	107 (2.2)

99.2). The results are remarkably similar to the original zygosity classification without the blood group data. For the twin pairs used to develop the updated classification algorithm there is 99.2% agreement between the original without blood-group-overwrite and updated classifications. Table 4 gives the current zygosity classification of the 4,774 complete twin pairs in the VET Registry as determined by the updated classification algorithm and DNA information.

Discussion

Our results suggest that the original determination of zygosity in the VET Registry was highly accurate. The principal error in the original zygosity classification was caused by incorporating data on blood group from the military records. During the development of the original zygosity classification we assumed that the military record blood group data was measured without error. However, the current analysis of DNA-based zygosity demonstrated that there was substantial error in the military record data. If we had applied the zygosity algorithm without the blood group data we would have achieved 97.1% accuracy. This is comparable to previously reported error rates in zygosity classification using questionnaires among twin pairs which range from 91–98% (Song et al., 2010; Christiansen et al., 2003; Jackson et al., 2001; Peeters et al., 1998; Sarna & Kaprio, 1980). However, unlike those studies the VET Registry did not have a subsample with serological markers or other molecular based zygosity to initially develop or test the classification algorithm when assigning zygosity to the twins within the VET Registry.

We have no readily available way to know why the blood group data were incorrect for some individuals.

The possibilities include errors in recording the information at the time of enlistment to errors in transcribing what was contained in the records. The latter seems more probable. We suspect that the errors in recording or transcription were random. However, our use of the blood group data did generate a systematic error for zygosity assignment since we reassign pairs from MZ to DZ based on blood group discordance.

The consequences to previous VET Registry research due to the misclassification of zygosity are likely minor. Fortunately, the systematic error would have biased results in a conservative manner. For the VET Registry studies based on the original zygosity classification their findings are unlikely to be altered. The relatively small number of MZ twins who were errantly placed into the DZ group in the original zygosity assignment would have likely increased the DZ twin correlations for most phenotypes. This would have deflated the observed genetic effects derived from classical twin study analyses. Further, this misclassification is unlikely to have altered the results from co-twin control studies, especially for studies within twins classified as MZ, due to the high MZ specificity observed.

Our updated DNA-based zygosity classification algorithm did not offer a major improvement over the original algorithm. Both algorithms had misclassification rates of less than 3%. This suggests that the original approach (Eisen et al., 1989), in the absence of a gold standard DNA-based or other serological zygosity reference sample, was reasonable and should be considered as a cost effective approach that could be used by other population-based twin samples.

There are some limitations to our work. One is that though DNA-based zygosity is treated as the definitive zygosity; there remains some possibility of laboratory error in the determination of zygosity. Additionally, the lack of a sufficient number of non-white twin pairs with DNA-based zygosity limited our reanalysis to white twin pairs within the VET Registry.

At its inception it was not feasible for the VET Registry to determine zygosity based on DNA. Instead, by using a zygosity questionnaire augmented by military record blood group data, twins were classified as MZ or DZ. The current report revisited our original zygosity classification and updates the methodology to produce a new zygosity classification. This new classification, now based on a subsample with DNA-based zygosity, refines the original methodology and corrects errors introduced by using the military record blood group data.

Acknowledgements

The United States Department of Veterans Affairs has provided financial support for the development and maintenance of the Vietnam Era Twin (VET) Registry. The views expressed in this article are those of the authors and do not necessarily represent the views of the Department of Veterans Affairs. Numerous organi-

zations have provided invaluable assistance in the conduct of this study, including: Department of Defense; National Personnel Records Center, National Archives and Records Administration; the Internal Revenue Service; National Institutes of Health; National Opinion Research Center; National Research Council, National Academy of Sciences; the Institute for Survey Research, Temple University. Most importantly, the authors gratefully acknowledge the continued cooperation and participation of the members of the VET Registry and their families. Without their contribution this research would not have been possible.

References

- Christiansen, L., Frederiksen, H., Schousboe, K., Skytthe, A., von Wurmb-Schwark, N., Christensen, K., & Kyvik, K. (2003). Age- and sex-differences in the validity of questionnaire-based zygosity in twins. *Twin Research*, 6, 275–278.
- Eisen, S. A., True, W. R., Goldberg, J., Henderson, W. G., & Robinette, C. D. (1987). The Vietnam era twin (VET) registry: Method of construction. *Acta Geneticae Medicae et Gemellologiae*, 36, 61–66.
- Eisen, S. A., Neuman, R., Goldberg, J., Rice, J., & True, W. R. (1989). Determining zygosity in the Vietnam era twin (VET) registry: An approach using questionnaire. *Clinical Genetics*, 35, 423–432.
- Goldberg, J., Curran, B., Vitek, M. E., Henderson, W. G., & Boyko, E. J. (2002). The Vietnam era twin registry. *Twin Research*, 5, 476–481.
- Henderson, W.G., Eisen, S. A., Goldberg, J., True, W. R., Barnes, J. B., & Vitek, M. E. (1990). The Vietnam era twin (VET) registry: A resource for medical research. *Public Health Reports*, 105, 368–373.
- Jackson R. W., Snieder H., Davis H., & Treiber F. A. (2001). Determination of twin zygosity: A comparison of DNA with various questionnaire indices. *Twin Research*, 4, 12–18.
- Magnus, P., Berg, K., & Nance, W. E. (1983). Predicting zygosity in Norwegian twin pairs born 1915–1960. *Clinical Genetic*, 24, 103–112.
- Peeters, H., Van Gestel, S., Vlietinck, R., Derom, C., & Derom, R. (1998). Validation of a telephone zygosity questionnaire in twins of known zygosity. *Behavior Genetics*, 28, 159–163.
- Sarna, S., & Kaprio, J. (1980). Use of multiple logistic analysis in twin zygosity diagnosis. *Human Heredity*, 30, 71–80.
- Song Y. M., Lee D., Lee M. K., Lee K., Lee H. J., Hong E. J., Han B., & Sung J. (2010). Validity of the zygosity questionnaire and characteristics of zygosity-misdiagnosed twin pairs in the healthy twin study of Korea. *Twin Research & Human Genetics*, 13, 223–230.