PERSPECTIVE ARTICLE
published: 30 April 2013
doi: 10.3389/fgene.2013.00061

The household contact study design for genetic epidemiological studies of infectious diseases

Catherine M. Stein[1,2]*, Noémi B. Hall[1], LaShaunda L. Malone[2] and Ezekiel Mupere[1,2,3]

[1] Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, OH, USA
[2] Uganda – Case Western Reserve University Research Collaboration, Kampala, Uganda
[3] Mulago Hospital, Makerere University School of Medicine, Kampala, Uganda

Edited by:
Karen T. Cuenco, University of Pittsburgh, USA

Reviewed by:
Peter A. Kanetsky, Perelman School of Medicine at the University of Pennsylvania, USA
Karen T. Cuenco, University of Pittsburgh, USA
Hui-Qi Qu, The University of Texas School of Public Health, USA

***Correspondence:**
Catherine M. Stein, Department of Epidemiology and Biostatistics, Case Western Reserve University, 2103 Cornell Road, Wolstein Research Building 1316, Cleveland, OH 44106, USA.
e-mail: catherine.stein@case.edu

Most genetic epidemiological study designs fall into one of two categories: family based and population-based (case–control). However, recent advances in statistical genetics call for study designs that combine these two approaches. We describe the household contact study design as we have applied it in our several years of study of the epidemiology of tuberculosis. Though we highlight its applicability for genetic epidemiological studies of infectious diseases, there are many facets of this design that are appealing for modern genetic studies, including the simultaneous enrollment of related and unrelated individuals, closely and distantly related individuals, collection of extensive epidemiologic and phenotypic data, and evaluation of effects of shared environment and gene by environment interaction. These study design characteristics are particularly appealing for current sequencing studies.

Keywords: extended pedigrees, genetic association, whole genome studies, cohort study, case-contact

INTRODUCTION

The advantages of family studies for genetic epidemiology have long been established (Stein and Elston, 2009). Early methods in genetic epidemiology utilized twins, sibling pairs, and other relative pairs to establish the relative recurrence risk of a disease. Segregation analysis and traditional linkage analysis can only be conducted using pedigree data. Concerns of population stratification are easily accounted for. In addition to these analytical issues, family studies have the advantage of investment of relatives; if someone in the family has a particular disease, family members are more likely to participate in research in order to somehow help their relative and others affected with the disease. Today, with the advent of whole exome and whole genome sequencing technologies, there are additional advantages of family studies, which we shall review below.

These advantages of family studies are further amplified for genetic epidemiological studies of infectious diseases. It was once believed that tuberculosis (TB) was a familial disease because it occurred within families. Once the disease was determined to be caused by a mycobacteria, the ideas surrounding the familial component recessed to the background. Now decades after the causal pathogen, *Mycobacterium tuberculosis* (Mtb), has been identified, many studies have shown that human genetic factors influence risk for development of TB infection and disease (Moller and Hoal, 2010; Stein, 2011). Development of TB infection and disease is essentially a phenotype resulting from a gene by environment interaction, so a well-constructed genetic epidemiological study must account for host genetics, shared environment, and gene x environment interaction. In this paper, we provide an overview of our household contact (HHC) study of TB and its advantages for genetic epidemiological studies, particularly in light of study designs best suited to identify rare genetic variants.

OVERVIEW OF THE HOUSEHOLD CONTACT STUDY DESIGN

In its natural history, TB is a two-stage process of infection followed by disease (Comstock, 1982). The household provides a natural setting to study TB because the genetic epidemiology of the two stages of infection and disease can be characterized. In our previous studies (Guwatudde et al., 2003), we defined a household as a group of people living within one residence and share meals together with a head of family who makes decisions for the household. Extensive epidemiological data are collected on individual risk factors, such as proximity and frequency of contact with the index case as well as other factors that may increase susceptibility, characteristics of the home that may increase the risk of transmission, as well as clinical data. Blood samples are obtained at baseline and longitudinally for genetic and immunologic studies.

In our HHC study, the first TB patient is identified in the household and referred to as the index case. Thereafter individuals who reside in the same household with the index case for a certain period prior to the diagnosis of the index case are identified and screened for TB as HHCs. Each HHC is also evaluated clinically for latent Mtb infection with the tuberculin skin test (or interferon-γ response assay in the future). Individuals who are tuberculin skin test negative have repeated skin tests several times over the 2-year study follow-up. Thus, the HHC evaluation is efficient in identification of individuals with different phenotypes or stages of TB infection in a household including: (1) exposed and uninfected, (2) exposed and infected without disease, (3) recent infection, and (4) active TB. These different household phenotypes or categories can provide the basis to compare genetic factors associated with TB infection and disease. As all of these stages of infection and disease are diagnosed, both the index case and his/her contacts receive appropriate clinical care and treatment, which is an immediate benefit to all study participants.

The design of the HHC study is ideal for evaluating genetic susceptibility to TB (Stein et al., 2003, 2005, 2007, 2008). The family structure and the ability to identify sibling pairs can form the basis for linkage analysis studies. Evaluation for new candidate genes for TB can be done through conduct of association studies such as case–control, family based, and/or case–parent studies. Heritability to TB can be determined using standard quantitative genetic approaches which can be based on host immune responses as intermediate phenotypes (Stein et al., 2005; Tao et al., 2013). Studies of HHCs have demonstrated that young children are at greater risk for developing TB and the clustering of cases within families does give hint at a familial susceptibility (Brailey, 1940; Puffer et al., 1952).

In sum, the essence of an HHC design is the recruitment of an entire household through an index case/proband, and collection of extensive clinical and epidemiological data. All age ranges and relative pair types are enrolled, and the entire spectrum of disease is captured. There is flexibility for collection of biological samples and a longitudinal component to observe changes in phenotypes and biomarkers.

ADVANTAGES OF THE HHC DESIGN FOR CURRENT GENETIC EPIDEMIOLOGICAL STUDIES

RECRUITMENT AND PHENOTYPE COLLECTION

As summarized above, the household is ascertained through an index case with TB (aka proband). Thus, as long as each individual in the household provides informed consent (or assent in the case of children), an entire family is enrolled in the study. Sometimes, there is another individual with TB in the household at the time of enrollment (co-prevalent case). In some households, another individual develops TB later on during the course of study follow-up (incident case). In this respect, no additional recruitment efforts are needed to identify additional affected individuals. The longitudinal component of the HHC design is valuable, especially for TB, where individuals have a 5–10% lifetime risk of developing active disease after exposure. In our studies, we have observed incident cases develop 2 years after initial enrollment of the household. If related individuals are desired for analytical and study design reasons (see "Analytical Considerations" below), the HHC design allows for easier enrollment of relatives, particularly in settings where literacy is low and roads are impassable (Bennett et al., 2002). Since both HIV co-infected and uninfected individuals may live within the same household, both will be enrolled in the study; this enables the examination of gene by HIV interaction effects (Stein et al., 2007). Finally, the ideal setting for a case-contact study is where the balance of household vs. community spread of disease is in favor of the household (Hill and Ota, 2010).

Both pediatric and adult TB cases may be diagnosed because the HHC design does not restrict enrollment by age. Studies suggest that the genetic influences on pediatric vs. adult TB differ (Malik et al., 2005; Alcais et al., 2010) and the HHC study design is an efficient method for ascertaining both types of cases. By contrast, studies that focus solely on recruitment of pediatric TB cases are challenging – school-based studies are limited because children living in poverty may not have access to education, and hospital- and clinic-based studies may also miss out on enrolling children because many babies are born at home in developing countries and families in poverty who are most at risk for developing TB may not have access to medical care. Door-to-door case finding strategies would require a great number of resources in order to identify a sufficient number of pediatric cases.

The HHC design also enables the enrollment of appropriate "controls." For a proper case–control study, controls must be similar in every way to the cases except that they do not have the disease of interest. For infectious diseases like TB, this is especially true, and in order for an individual to have the opportunity to become a case, he/she must have been exposed to an infectious TB case. This is particularly important for TB, because clinical status of the controls determines whether observed genetic associations are with susceptibility to latent infection or progression to active disease (Stein, 2011). By virtue of the HHC design, all the household members have been exposed to the index case. The selection of appropriate controls in community-based studies of TB is problematic (Hill and Ota, 2010).

Finally, studies of large pedigrees often include extensive and highly detailed phenotype information (Wijsman, 2012). This is extraordinarily useful for infectious diseases such as TB for a number of reasons. As the natural history of Mtb infection and disease follows a two-stage process, the longitudinal HHC design captures all of these stages, and progression from one stage to another. Furthermore, the HHC design can also include collection of extensive immunological data. The HHC design therefore is flexible enough to analyze immunological correlates of the natural history of TB (Whalen et al., 2006; Mahan et al., 2012), and also genetic influences on the immune response to Mtb (Stein et al., 2007, 2008). Omics technologies, such as gene expression and proteomic arrays, can also be incorporated into a study that has an established blood draw protocol and rigorous clinical classification. Finally, as we describe later, data are also collected on important epidemiological factors, which can be incorporated as covariates as well as in gene by environment interaction models.

ANALYTICAL CONSIDERATIONS

One unique aspect of HHC studies is that households may contain all sorts of relationship types – nuclear families, extended relatives, and unrelated individuals. Half-siblings are common in African settings where polygamy is practiced (Bennett et al., 2002). Similarly, adoption by extended relatives is common when children are orphaned, which may be particularly relevant in areas with a heavy AIDS burden.

A few studies have developed strategies for jointly analyzing family based and case–control/population-based data (Chen and

Lin, 2008; Gray-McGuire et al., 2009; Lasky-Su et al., 2010; Zheng et al., 2010; Mirea et al., 2012). Though they differ in how they combine data from these two different study designs – some analyze them all together, and some combine *p*-values or test statistics – there are some common themes. First, joint analysis of data from these two different study designs results in increased power due to increased sample size, enabling the detection of smaller effect sizes. Second, family based data have the advantage of controlling for population substructure, which alleviates this common concern of population-based studies.

There have been many recent reports detailing the usefulness of extended pedigrees for the analysis of sequence data and detection of rare variants. Cirulli and Goldstein (2010) explain how the analysis of distantly related, co-affected individuals is an economical design, because there will be fewer genetic variants in common, thereby reducing the search space for rare variants. Stringent filtering could use identity-by-descent sharing to capitalize on this biological phenomenon (Akula et al., 2011). Large pedigrees also have increased power to detect linkage, even in the presence of linkage heterogeneity among families, and are enriched for variants of interest (Wijsman, 2012). Linkage analysis with pedigree data can be used as a filtering strategy of chromosomal regions, and can guide the selection of subjects to sequence (Wijsman, 2012). In addition, linkage analysis may be conducted to examine co-segregation between the trait and variant(s) of interest (Clerget-Darpoux and Elston, 2007; Ziegler and Sun, 2012). Consanguineous marriages are common in West Africa, which increases the power to detect rare recessive alleles (Bennett et al., 2002). To summarize, all of the relationship types that are useful for the identification of rare variants are easily obtainable in the HHC design.

IMPACT OF ENVIRONMENT

A well-designed HHC study includes vast epidemiologic data about environmental risk factors for transmission of disease within homes. For TB, these include factors related to ventilation and crowding within the home, poverty, clinical characteristics of the index case that make him/her more infectious, and proximity to the index case that increase degree of contact (Stein et al., 2005; Mandalakas et al., 2012). Risk of infection by Mtb is determined by a number of epidemiological risk factors (Guwattude et al., 2003; Lienhardt et al., 2003; Mandalakas et al., 2012), and many variables associated with high risk of TB transmission are automatically present in the HHC design. Analysis of foster relationships as seen in adoptions may be useful for the estimation of effects due to shared environment (Bennett et al., 2002), and many such relationships occur in HHC studies in the developing world.

Genetic substrains of Mtb may differ in their transmissibility. All of these factors relate to the risk of an individual to acquire infection, and develop disease, and thus are important in epidemiological characterization of affected individuals. Furthermore, recent studies have also suggested that substrains of Mtb have synergistic effects with host genes, thus resulting in gene x environment interaction effects related to TB risk (Caws et al., 2008). Case-only designs can be nested within HHC studies to examine these gene x environment effects (Bennett et al., 2002). Because exposure to the index case is generally highest, and in turn exposure to that individual's strain of Mtb, the HHC design provides a natural setting to test both transmissibility, gene x environment interaction, and role of shared environment.

Nutrition and nutritional status are also important factors in TB-related outcomes (Jaganath and Mupere, 2012; Mupere et al., 2012a). We have shown that nutritional status of a patient may be an indicator on how the food basket is shared in the household and the subsequent macro- and micronutrient intake (Mupere et al., 2012b). Because of the shared environmental and genetic components of diet and obesity (or in the case of TB, malnutrition), the HHC design provides a robust setting to test the role of nutritional status on infectious disease outcomes.

EXAMPLES FROM OUR STUDIES

Our genetic association studies have taken the approach by Gray-McGuire et al. (2009). We identified the first reported association between TNFR1 gene and TB and also a gene by HIV interaction for this same gene (Stein et al., 2007). Our genome-wide linkage scan (Stein et al., 2008) and subsequent fine mapping studies (Baker et al., 2011) replicated previously a novel set of genes on chromosome 20, CTSZ, and MC3R. We have also identified novel chromosomal regions linked to a unique resistance phenotype (Stein et al., 2008); we are uniquely able to clinically and epidemiologically characterize this phenotype because of our solid study design. Our future plans will incorporate structural equation modeling (SEM to multivariately analyze the influences of host genetics, immunology, and environment on clinical outcome; this shall be done using a SEM approach that jointly models familial relationship and covariance among variables (Morris et al., 2011).

CONCLUSION

Certainly HHC designs may be expensive to implement, because they include repeated clinical visits, longitudinal data collection, and travel to the homes. However, the wealth of data collected through HHC studies is invaluable for genetic epidemiological studies, as described here. HHC study designs offer unique advantages for genetic epidemiological studies, including the presence of related and unrelated individuals, and the ability to quantify environmental factors that are important for both shared environmental influences on the phenotype as well as gene x environment interaction. Though our focus has been primarily on studies of TB, this study design has advantages for the study of infectious diseases in general (Hill and Ota, 2010).

ACKNOWLEDGMENTS

We would like to thank Drs. Christopher Whalen and W. Henry Boom for their key roles in the design and implementation of the Household Contact Studies in Kampala, Uganda.

Funding for this work was provided by the Tuberculosis Research Unit (grant N01-AI95383 and HHSN266200700022C/N01-AI70022 from the NIAID), and National Heart Lung and Blood Institutes (NHLBI) grants R01HL096811, R01HL-10566113, and T32HL007567.

REFERENCES

Akula, N., Detera-Wadleigh, S., Shugart, Y., Nalls, M., Steele, J., and McMahon, F. J. (2011). Identity-by-descent filtering as a tool for the identification of disease alleles in exome sequence data from distant relatives. *BMC Proc.* 5(Suppl. 9):S76. doi: 10.1186/1753-6561-5-S9-S76

Alcais, A., Quintana-Murci, L., Thaler, D. S., Schurr, E., Abel, L., and Casanova, J. L. (2010). Life-threatening infectious iseases of childhood: single-gene inborn errors of immunity? *Ann. N. Y. Acad. Sci.* 1214, 18–33.

Baker, A. R., Zalwango, S., Malone, L. L., Igo, R. P. Jr, Qiu, F., Nsereko, M., et al. (2011). Genetic Susceptibility to Tuberculosis Associated with CTSZ Haplotype in Ugandan Household Contact Study. *Hum. Immunol.* 72, 426–430.

Bennett, S., Lienhardt, C., Bah-Sow, O., Gustafson, P., Manneh, K., Del Prete, G., et al. (2002). Investigation of environmental and host-related risk factors for tuberculosis in Africa. II. Investigation of host genetic factors. *Am. J. Epidemiol.* 155, 11, 1074–1079.

Brailey, M. (1940). Mortality in the children of tuberculous households. *Am. J. Public Health Nations Health* 30, 816–823.

Caws, M., Thwaites, G., Dunstan, S., Hawn, T. R., Lan, N. T., Thuong, N. T., et al. (2008). The influence of host and bacterial genotype on the development of disseminated disease with *Mycobacterium tuberculosis*. *PLoS Pathog.* 4:e1000034. doi: 10.1371/journal.ppat.1000034

Chen, Y. H., and Lin, H. W. (2008). Simple association analysis combining data from trios/sibships and unrelated controls. *Genet. Epidemiol.* 32, 520–527.

Cirulli, E. T., and Goldstein, D. B. (2010). Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nat. Rev. Genet.* 11, 415–425.

Clerget-Darpoux, F., and Elston, R. C. (2007). Are linkage analysis and the collection of family data dead? Prospects for family studies in the age of genome-wide association. *Hum. Hered.* 64, 91–96.

Comstock, G. (1982). Epidemiology of tuberculosis. *Am. Rev. Respir. Dis.* 125, 8–15.

Gray-McGuire, C., Bochud, M., Goodloe, R., and Elston, R. C. (2009). Genetic association tests: a method for the joint analysis of family and case–control data. *Hum. Genomics* 4, 2–20.

Guwattude, D., Nakakeeto, M., Jones-Lopez, E. C., Maganda, A., Chiunda, A., Mugerwa, R. D., et al. (2003). Tuberculosis in household contacts of infectious cases in Kampala, Uganda. *Am. J. Epidemiol.* 158, 887–898.

Guwatudde, D., Zalwango, S., Kamya, M. R., Debanne, S. M., Diaz, M. I., Okwera, A., et al. (2003). Burden of tuberculosis in Kampala, Uganda. *Bull. World Health Organ.* 81, 799–805.

Hill, P. C., and Ota, M. O. (2010). Tuberculosis case-contact research in endemic tropical settings: design, conduct, and relevance to other infectious diseases. *Lancet Infect. Dis.* 10, 723–732.

Jaganath, D., and Mupere, E. (2012). Childhood tuberculosis and malnutrition. *J. Infect. Dis.* 206, 1809–1815.

Lasky-Su, J., Won, S., Mick, E., Anney, R. J., Franke, B., Neale, B., et al. (2010). On genome-wide association studies for family-based designs: an integrative analysis approach combining ascertained family samples with unselected controls. *Am. J. Hum. Genet.* 86, 573–580.

Lienhardt, C., Fielding, K., Sillah, J., Tunkara, A., Donkor, S., Manneh, K., et al. (2003). Risk factors for tuberculosis infection in sub-Saharan Africa: a contact study in The Gambia. *Am. J. Respir. Crit. Care Med.* 168, 448–455.

Mahan, C. S., Zalwango, S., Thiel, B. A., Malone, L. L., Chervenak, K. A., Baseke, J., et al. (2012). Innate and adaptive immune responses during acute M. tuberculosis infection in adult household contacts in Kampala, Uganda. *Am. J. Trop. Med. Hyg.* 86, 690–697.

Malik, S., Abel, L., Tooker, H., Poon, A., Simkin, L., Girard, M., et al. (2005). Alleles of the *NRAMP1* gene are risk factors for pediatric tuberculosis disease. *Proc. Natl. Acad. Sci. U.S.A.* 102, 12183–12188.

Mandalakas, A. M., Kirchner, H. L., Lombard, C., Walzl, G., Grewal, H. M., Gie, R. P., et al. (2012). Well-quantified tuberculosis exposure is a reliable surrogate measure of tuberculosis infection. *Int. J. Tuberc. Lung Dis.* 16, 1033–1039.

Mirea, L., Infante-Rivard, C., Sun, L., and Bull, S. B. (2012). Strategies for genetic association analyses combining unrelated case–control individuals and family trios. *Am. J. Epidemiol.* 176, 70–79.

Moller, M., and Hoal, E. G. (2010). Current findings, challenges and novel approaches in human genetic susceptibility to tuberculosis. *Tuberculosis (Edinb.)* 90, 71–83.

Morris, N. J., Elston, R. C., and Stein, C. M. (2011). A framework for structural equation models in general pedigrees. *Hum. Hered.* 70, 278–286.

Mupere, E., Malone, L., Zalwango, S., Chiunda, A., Okwera, A., Parraga, I., et al. (2012a). Lean tissue mass wasting is associated with increased risk of mortality among women with pulmonary tuberculosis in urban Uganda. *Ann. Epidemiol.* 22, 466–473.

Mupere, E., Parraga, I. M., Tisch, D. J., Mayanja, H. K., and Whalen, C. C. (2012b). Low nutrient intake among adult women and patients with severe tuberculosis disease in Uganda: a cross-sectional study. *BMC Public Health* 12:1050. doi: 10.1186/1471-2458-12-1050.:1050-12

Puffer, R. R., Zeidberg, L. D., Dillon, A., Gass, R. S., and Hutcheson, R. H. (1952). Tuberculosis attack and death rates of household associates; the influence of age, sex, race, and relationship. *Am. Rev. Tuberc.* 65, 111–127.

Stein, C. M. (2011). Genetic epidemiology of tuberculosis susceptibility: impact of study design. *PLoS Pathog.* 7:e1001189. doi: 10.1371/journal.ppat.1001189

Stein, C. M., and Elston, R. C. (2009). Finding genes underlying human disease. *Clin. Genet.* 75, 101–106.

Stein, C. M., Nshuti, L., Chiunda, A. B., Boom, W. H., Elston, R. C., Mugerwa, R. D., et al. (2005). Evidence for a major gene influence on tumor necrosis factor-alpha expression in tuberculosis: path and segregation analysis. *Hum. Hered.* 60, 109–118.

Stein, C. M., Zalwango, S., Chiunda, A. B., Millard, C., Leontiev, D. V., Horvath, A. L., et al. (2007). Linkage and association analysis of candidate genes for TB and TNFalpha cytokine expression: evidence for association with IFNGR1, IL-10, and TNF receptor 1 genes. *Hum. Genet.* 121, 663–673.

Stein, C. M., Zalwango, S., Malone, L. L., Won, S., Mayanja-Kizza, H., Mugerwa, R. D., et al. (2008). Genome scan of *M. tuberculosis* infection and disease in Ugandans. *PLoS ONE* 3:e4094. doi: 10.1371/journal.pone.0004094

Stein, C., Guwatudde, D., Nakakeeto, M., Peters, P., Elston, R. C., Tiwari, H. K., et al. (2003). Heritability analysis of cytokines as intermediate phenotypes of tuberculosis. *J. Infect. Dis.* 187, 1679–1685.

Tao, L., et al. (2013). Genetic and shared environmental influences on interferon-gamma production in response to *Mycobacterium tuberculosis* antigens in a Ugandan population. *Am. J. Trop. Med. Hyg.* (in press).

Whalen, C. C., Chiunda, A., Zalwango, S., Nshuti, L., Jones-Lopez, E., Okwera, A., et al. (2006). Immune correlates of acute *Mycobacterium tuberculosis* infection in household contacts in Kampala, Uganda. *Am. J. Trop. Med. Hyg.* 75, 55–61.

Wijsman, E. M. (2012). The role of large pedigrees in an era of high-throughput sequencing. *Hum. Genet.* 131, 1555–1563.

Zheng, Y., Heagerty, P. J., Hsu, L., and Newcomb, P. A. (2010). On combining family-based and population-based case–control data in association studies. *Biometrics* 66, 1024–1033.

Ziegler, A., and Sun, Y. V. (2012). Study designs and methods post genome-wide association studies. *Hum. Genet.* 131, 1525–1531.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 30 January 2013; accepted: 05 April 2013; published online: 30 April 2013.

Citation: Stein CM, Hall NB, Malone LL and Mupere E (2013) The household contact study design for genetic epidemiological studies of infectious diseases. Front. Genet. ***4****:61. doi: 10.3389/fgene.2013.00061*

This article was submitted to Frontiers in Applied Genetic Epidemiology, a specialty of Frontiers in Genetics.