Published in final edited form as:

J Clin Epidemiol. 2008 July ; 61(7): 671–678. doi:10.1016/j.jclinepi.2007.07.019.

Feasibility and validity were demonstrated of an online case–control study using the prototype of recent-onset systemic lupus erythematosus

Timothy McAlindon^{a,*}, Jun Wang^a, Margaret Formica^a, Ashley Kay^a, Hocine Tighiouart^b, Christine Chaisson^c, and Jeremiah Fletcher^a

^aDivision of Rheumatology, Tufts-New England Medical Center, 750 Washington Street, Box 406, Boston, MA 02111, USA

^bClinical Research Center, Tufts-New England Medical Center, Boston, MA, USA

^cData Coordinating Center, Boston University School of Public Health, Boston, MA, USA

Abstract

Objective—To test the feasibility and validity of the online case–control study design through the empirical deployment of a prototype study of recent-onset systemic lupus erythematosus (SLE).

Study Design and Setting—We conducted an Internet-based case–control study of SLE during 2003–2005. The source population comprised Google users searching on medical key terms, solicited using sponsored links. Cases fulfilled a self-administered algorithm for SLE diagnosed within 5 years. A subset underwent confirmation by medical record review. Controls were matched to cases using a propensity score.

Results—Four hundred and two cases and 693 control applicants finished the questionnaires, yielding 389 matched case–control pairs. Eighty-two percentage of the records documented a clinical diagnosis of SLE, and 61% documented 4 American College of Rheumatology criteria for SLE. Control applicants resembled Internet users, with the exceptions of comprising more women (86% vs. 52%) and fewer minority individuals (e.g., 5% vs. 9% for African-Americans). There was a broad representation of clinical manifestations. SLE was associated with miscarriage (odds ratio [OR] = 3.0, 95% confidence interval [CI] = 2.0–4.7), allergy to sulfonamides (OR = 2.2, CI = 1.5–3.2), hives (OR = 1.9, CI = 1.4–2.5), and shingles (OR = 2.3, CI = 1.4–3.7).

Conclusion—It is possible to perform case–control studies over the Internet using an internally valid design, obtain reliable information from participants, and confirm established associations.

Keywords

Case-control; Online; Internet; Lupus; SLE; Risk factors; Epidemiology

1. Introduction

The pursuit of causes of rare medical disorders is often obfuscated by the difficulty in assembling sufficient numbers of cases for robust etiologic investigations. Systemic lupus erythematosus (SLE), for example, is a serious autoimmune disorder whose epidemiology is

hard to study because of its low incidence. Development of new ways to enhance recruitment of cases in ways that conform to the design requirement of epidemiologic studies could facilitate studies of rare disorders such as SLE and stimulate progress in their fields.

An Internet-based approach for participant recruitment and data acquisition offers this possibility. The global expansion of Internet use and its technological refinement have made this interactive modality a powerful resource for performing questionnaire-based research. In principle, the Internet could greatly facilitate epidemiologic study of uncommon diseases by reaching and obtaining information from large numbers of geographically dispersed participants. An especially appealing application of the Internet is the possibility of performing case-control studies entirely online.

An important consideration that permits this approach is that methodologically sound case-control studies can be performed in special subsets of the population [1]. Nevertheless, the manner in which the complex methodological requirements of case-control studies might be accomplished using the Internet requires careful thought. In particular, the internal validity of the study design is contingent on being able to (i) define a source population of web users, (ii) recruit a representative sample of cases and controls, and (iii) operate a recruitment mechanism that is independent of the exposures of interest. Other important components include its performance in recruiting sufficient numbers of authentic cases and controls, and the reliability of their responses.

In addition to testing the feasibility of the online case-control study design through the empirical deployment of a prototype study, our objective was to evaluate three performance outcomes that provide insight into the internal, external, and construct validity of the approach. Specifically, we aimed to (i) evaluate the demographic characteristics of the applicants for the control group with reference to the source population, (ii) evaluate the representativeness of the clinical characteristics of the case participants, and (iii) use matched case-control analysis to test whether the Internet approach could confirm associations with a panel of known SLE epiphenomena.

2. Methods

2.1. Web site construction and data security

We constructed a Web site to solicit participants and conduct the study. This was hosted on an independent server within the Tufts-New England Medical Center domain. We kept the location of this Web site covert so that access would be restricted to individuals linking to it from our Google advertisements. The home page provided information about the study and links to the consent form, eligibility screening, and questionnaires. We protected participants' data with encryption, passwords, and a firewall. Information submitted by participants was added cumulatively to the database. Refinements included real-time error checking and generation of automated responses by e-mail.

2.2. Source population and participant recruitment

We defined the source population as individuals searching on Google using medical key terms, living within the United States, aged 18 years or older. We placed sponsored links on the Google search engine web page to solicit participants for our study. Their appearance was triggered when a user entered one of a set of prespecified key terms in the search field. The sponsored link pointed toward our study Web site.

2.2.1. Cases—To solicit potential cases, we linked our advertisement to SLE-related terms (SLE, lupus, and systemic lupus erythematosus), and screened applicants using a validated self-reported algorithm for SLE that has been developed in prior research [2]. This algorithm classifies individuals on the basis of responses to questions about a recent (<5 years) physician-given diagnosis of SLE, the use of appropriate SLE medications (steroids/corticosteroids, hydroxychloroquine/chloroquine/mepacrine, azothiaprime, cyclophosphamide, and methotrexate), and the nonuse of other rheumatic disease medications (gold, sulphasalazine, and penicillamine). In prior research, we found this algorithm to have a positive predictive value for SLE of 0.84 [2]. Our primary case group comprised individuals fulfilling this screening algorithm for SLE. To confirm the diagnosis, we also asked people reporting SLE for permission to obtain their medical records and send a criteria checklist to their physician(s). Among the cases for whom we obtained records, we evaluated each chart for documentation of each American College of Rheumatology (ACR) criterion for classification of SLE [3,4]. This was combined with information on the criterion checklist completed by each participant's physician. Classification of “definite” SLE required 4 positive criteria.

2.2.2. Controls—To solicit potential controls, we used terms derived from a list of medical disorders most frequently entered on Google. Initially, we selected the nine most commonly entered terms for diseases not known to share risk factors with SLE and not exclusive to men (e.g., migraine, hypertension, sinusitis, and fibroids). Later, we changed the approach to one in which we randomly selected 10 key terms every 4 weeks from a list of 80 common medical key terms. To reduce the bias potentially caused by the difference between demographic characteristics of case and control applicants, we performed a one-to-one case-control match on propensity score defined by gender, age, race, ethnicity, and region of residence.

2.3. Informed consent and medical record release

The Web site and online consent form included a detailed description of the study and the measures taken to protect confidentiality. Applicants asserted their consent to participate electronically; however, we required a signed hard copy of the medical record release. The study received ethical approval from the Institutional Review Board of Tufts-New England Medical Center.

2.4. The questionnaire

Participants completed an online questionnaire about demographic factors, medications, comorbidities, common exposures, and a panel of factors consistently associated with SLE (miscarriage [5], shingles [6–8], hives [6,9], and allergy to sulfonamides [10–13]). We used standardized questions derived from previous studies [2,14]. The questions on lifestyle factors and disease epiphenomena were constructed to ascertain lifetime exposure, current exposure, and exposure before SLE diagnosis. For control applicants, we assigned a reference year in lieu of date of diagnosis that was frequency matched to the distribution of diagnosis year among case applicants. This was accomplished in real time using an embedded computational subroutine.

2.5. Evaluation of data quality

We also used a panel of 13 duplicative questions to evaluate the internal consistency of the questionnaire responses. For example, we asked for both age and date of birth in the questionnaire and compared the consistency of the answers from these two questions. We also checked the accuracy of self-reported information for date of birth, body weight, and height, using data on these exposures documented in medical records.

2.6. External validity of the samples

We derived statistics on the demographic and socioeconomic characteristics of Internet users in the United States by combining information from two 2003 U.S. Census Bureau data sets; one on general population demographics [15] and the other on Internet-user characteristics [16]. We obtained data on the distribution of body mass index (BMI) in the U.S. population from the National Health and Nutrition Examination Survey of 1999–2002 [17], and on smoking and drinking from the National Center for Health Statistics [18]. We compared the age-adjusted distributions of these characteristics among the control applicants with the general population. As a basis for future case group comparisons, we also analyzed the clinical patterns of SLE by computing the sample prevalence of each of the manifestations that contribute to the ACR criteria set.

2.7. Statistical analysis

We used the propensity score–matching method [19] to generate a matched case–control data set from the case and control applicants. Briefly, we first used the multivariate logistic regression including gender, age, race, ethnicity, and region of residence as independent variables to compute each applicant's propensity score, which is the predicted probability of application to the case group. We then created a one-to-one propensity score–matched sample using greedy match algorithm and the nearest available pair-matching method. Cases were ordered and sequentially matched to the nearest unmatched control. If more than one control matched to a case, the control was selected at random. Once a match was made, the match was not reconsidered. This matching procedure was conducted using a user-written SAS Macro [20]. The distribution of the demographic and socioeconomic characteristics among matched cases and controls was compared using the Wilcoxon signed-rank test for continuous and ordinal variables, and McNemar's test for categorical variables. To evaluate the association between SLE and its epi-phenomena, we performed conditional logistic regression analysis with a discrete logistic model stratified by matching variables. We also confined the analyses to confirmed cases and their matched controls. For the analysis including all case and control applicants, we ran a multivariate unconditional logistic regression model, adjusting for gender, age, race, ethnicity, and region of residence.

3. Results

3.1. Recruitment performance

During the 25-month recruitment phase, 1,727 people applied to join as cases and 1,379 as controls, of whom 601 potential cases and 1,329 potential controls passed the preliminary eligibility screen. Four hundred and two of the case group (67%) and 693 of the control group (52%) consented to the study and completed the questionnaires. Among these, the one-to-one matching procedure generated 389 case–control pairs. The full flow of recruitment is enumerated in Fig. 1.

3.2. Consistency and reliability of data collected from Internet

Among the panel of 13 built-in duplicative questions, 4% of participants exhibited one inconsistent answer, and none exhibited more than one. In the medical records, we were able to find documentation of date of birth for 154 participants, height for 57, and weight (within 1 year) for 85. There were two (1.3%) discrepancies for date of birth, which we identified to be data entry errors. For height, 55 (96%) records corresponded to within 1 in, whereas two (4%) differed by 1–2 in. The mean difference between self-reported and documented body weight was 3.5% (SD 3.3%). The mean difference of weight among the nine participants with BMI > 40 was 2.1% (SD 1.4%).

3.3. Comparison of the control applicants to general Internet users and the U.S. population

The two groups of control applicants (i.e., those recruited using selected vs. random key terms) differed in age (mean 45 vs. 40 years, $P < 0.0001$), but were similar in terms of gender, race, ethnicity, region of residence, education, and household income (all $P \geq 0.1$). Because of their overall similarity, we combined them in further analyses.

Table 1 compares the demographic and socioeconomic characteristics of the control applicants to those of general U.S. Internet users. Because data on Internet use among some minority groups were unavailable, we applied to them the rate documented among African-Americans. Compared to general Internet users, the control applicant group had fewer minority individuals (e.g., 5.0% vs. 9.1% for African-Americans), fewer younger applicants (9.2% vs. 14.6% for age 18–24), more women (85.6% vs. 51.5%), and a possible skew to higher education level, but lower household income. We also compared the control applicant sample with the U.S. population sample with respect to BMI, smoking, and drinking (for which data are not available among general Internet users). These show substantial similarity between these two populations for women (Table 2). Less similarity was seen for male population, probably due to the small number in our male control sample.

3.4. The case group

We were able to obtain medical records for 170 cases, among whom 139 (82%) documented a clinical diagnosis of SLE, 104 (61%) documented 4, and 119 (70%) documented 3 ACR criteria for SLE. Twelve (7%) of the records indicated no lupus or other diseases. The remainder (11%) did not have enough information to make a diagnosis. Twenty-nine percentage of the confirmed cases had been diagnosed within 1 year. Compared to those who fulfilled the self-reported algorithm for SLE but did not provide records or the records did not give a confirmed diagnosis, the cases confirmed by medical record review had higher education (college graduate or higher, 50% vs. 34%, $P < 0.05$) and appeared to be older (median age 42 vs. 39, $P = 0.07$), but otherwise had similar demographic characteristics (gender, race, ethnicity, region of residence, and household income). Based on the chart review, there was also a broad representation of clinical manifestations among confirmed cases that included severe disease (Table 3).

3.5. Confirmation of factors associated with SLE

Because of our matching procedures, participants with SLE and their matched controls exhibited highly concordant distributions for demographic and socioeconomic characteristics (Table 4). Participants with SLE were more likely to have experienced miscarriages (odds ratio [OR] = 3.03, 95% confidence interval [CI] = 1.95–4.70), allergy to sulfonamides (OR = 2.19, 95% CI = 1.48–3.24), hives (OR = 1.87, 95% CI = 1.39–2.52), and shingles (OR = 2.25, 95% CI = 1.38–3.69) (Table 5). These results were similar when analyses were confined to confirmed cases and their matched controls. We also analyzed the above associations with all applicants included (402 cases and 693 controls) and adjusted for gender, age, race, ethnicity, and region of residence in the multivariate models. Results were very similar to the matched case–control analysis (data not shown).

4. Discussion

In contemplating how a case–control study design might be operated over the Internet, we identified several factors that predicate the utility and validity of such an approach. One of the fundamental methodological considerations was whether we could define a population of web users among whom we could solicit cases and controls using a unified recruitment mechanism. Our use of Google met this need and avoided special-interest Web sites and

their potential biases. Further, the resources offered by Google (sponsored links and campaign reports) supported our recruitment activities.

We met our goal in online recruitment, and subsequent authentication, of over 100 individuals with recently diagnosed SLE—even using constrained solicitation strategy. Although the period of recruitment was fairly long, the effort involved was considerably less than that in traditional approaches [21], and our advertisement campaign was not particularly vigorous. This suggests that studies of rare diseases may be a particularly appealing application for the Internet-based design, especially where the procedures required for disease classification are simpler.

The primary function of the control group in a case-control study is to provide a representation of the population-generating cases. Demographic data on the strictly defined source population were unavailable, so we compared our control applicant data with data on Internet users inferred from census statistics. The comparison suggests that our operational definition of the source population as “Google™-users searching on medical key terms” may have delineated a subsample of Internet users whose characteristics differed from those of general Internet users in some respects, especially gender. Compared to general Internet users, our control applicant group had substantially more women (85.6% vs. 51.5%). However, it has been reported that women are more likely than men to use the Internet to search for health information—85% among female vs. 75% among male Internet users—and tend to be older and have higher educational levels [22]. The alternative is that biases in recruitment occurred, perhaps as a result of greater volunteerism among women. In either case, it is likely that our use of medical key terms generated differences in characteristics between the control applicants and Internet users. Differences in educational levels might also explain the lower numbers of minority applicants in our study. There is, however, some uncertainty in our data with respect to the extent of minority group underrepresentation because a proportion of applicants declined to provide race/ethnicity information. If disinclination to provide information of race is greater among minority groups, the actual difference will be smaller than what our data suggested. It is also encouraging that the disparity in Internet use by minorities and individuals of lower socioeconomic status is rapidly diminishing [23].

One of our goals was to evaluate the representativeness of clinical SLE manifestations in the case group. Among the chart-confirmed cases, the prevalence of the different clinical manifestations encompassed within the ACR criteria set [3,4] fell within the ranges observed among clinical series [24], with the exception of renal involvement, which was less common in our sample.

Because the accuracy of information entered by participants into the web-based questionnaires is critical to the validity of this approach, we examined indicators of internal consistency and compared self-reported height and weight with documented findings in medical records. As in prior Internet-based studies [25–28], we found a high degree of reliability in participants' responses. Indeed, the concordance of self-reported with recorded weight among individuals with high BMI exceeded that documented in more traditional approaches [29,30]—perhaps reflecting the sense of anonymity afforded by the Internet [31–33].

A major concern in conducting a case-control study is that cases differ from controls in characteristics other than those targeted for study. One approach to control for such potential confounding is to match the controls to cases so that they are similar in key characteristics. We accomplished this using a propensity score-matching method, a well-tested approach that is in increasing use to reduce bias or confounding from covariates in observational

studies [19]. We used the matched case-control data set to assess the construct validity of our study design. We tested for associations with a panel of factors documented in prior studies to be associated with SLE, including herpes zoster [6,7,34,35], urticaria [6,9], allergy to sulfonamides [10–12], and miscarriage [5]. Our analyses demonstrated associations that were consistent with prior studies and showed strong concordance with respect to effect size. For example, our OR of 2.2 (95% CI = 1.5–3.2) for allergy to sulfonamides is similar to the previous-reported OR of 2.4 (95% CI = 1.2–4.7) in a case-control study of SLE [11]. The OR of 1.9 (95% CI = 1.4–2.5) for the history of hives is similar to the previous-reported 1.8 (95% CI = 1.1–3.0) [6]. The OR of 2.3 (95% CI of 1.4–3.7) for history of herpes zoster is also approximate to the reported ranges of OR = 2.98 ($P < 0.003$, 95% CI unknown) [34] and OR = 4.7 (95% CI = 1.1–21.7) [6]. Of note, an unmatched case-control analysis (i.e., using the entire control applicant sample) yielded almost identical strengths of associations, suggesting that the size and range of characteristics among the samples provided robustness to the analyses.

Our study design brings with it some unique complexities, the most striking of which is the abstractness of the source population. On the other hand, most aspects of the study have parallels in traditional approaches. For example, as in studies based in special clinics, we do not expect our sampling procedure to identify the entirety of cases. This can limit generalizability, but should not affect the validity of the associations found for the case sample. Also, considerably more individuals visited the study Web site than those who applied to participate. We were not able to obtain direct data regarding the eligibility of these visitors, therefore, could not evaluate the participation rate among eligible visitors. This problematic issue is also pervasive in accepted traditional approaches, such as random digit dialing, where nonparticipation rates are likely to be high, yet often not transparent [36]. We also did not have data to show whether eligible SLE cases who participated in this study were different from those who were eligible but did not participate. However, the spectrum of SLE clinical manifestations in our case group suggests that they were likely to be fairly representative.

Our experience in this study highlights several considerations for investigators planning similar endeavors. The most critical of these concerns the ability of the study design to determine the similarity of the sample to the source population. Approaches to this could include linking recruitment to a benchmark, such as information from a disease registry and deploying a design that is less susceptible to this issue, such as a family-based approach. Secondly, the need for secondary confirmation of disease status, especially a disease with a panoply of manifestations, adds complexity, burden, and possible biases. Thirdly, if key terms are used to recruit *controls*, the study should collect detailed information about potential confounders to facilitate a stringent matching procedure. Finally, there may be a favorable trade-off of advertising cost and intensity with duration of recruitment.

In summary, our prototype Internet-based case-control study exhibited several obvious measures of success. We were able to recruit a large number of individuals with recently diagnosed SLE (a rare medical disorder) and controls using an internally valid design. We obtained reliable information from participants, and were able demonstrate convergent construct validity by confirming associations with a panel of disease epiphenomena. Although certain aspects would benefit from further development, especially the approach to recruitment of samples for which the source population has measurable characteristics, the approach holds promise as a way to perform epidemiologic research into rare disorders, and adds to the portfolio of study designs that have been deployed over the Internet [25–28].

Acknowledgments

This work was supported by grant P60 AR47785 from the National Institutes of Health and National Institute of Arthritis and Musculoskeletal and Skin Diseases.

References

1. Rothman, KJ.; Greenland, S. Case-control studies.. In: Rothman, KJ.; Greenland, S., editors. Modern Epidemiology. Lippincott Williams & Wilkins; Philadelphia: 1998. p. 93-114.
2. McAlindon TE, Formica M, Palmer JR, Lafyatis R, Rosenberg L. Assessment of strategies for identifying diagnosed cases of systemic lupus erythematosus through self-report. *Lupus*. 2003; 12:754-9. [PubMed: 14596424]
3. Hochberg MC. Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus. *Arthritis Rheum*. 1997; 40:1725. [letter]. [PubMed: 9324032]
4. Tan EM, Cohen AS, Fries JF, Masi AT, McShane DJ, Rothfield NF, et al. The 1982 revised criteria for the classification of systemic lupus erythematosus. *Arthritis Rheum*. 1982; 25:1271-7. [PubMed: 7138600]
5. Kitridou, RC.; Goodwin, TM. The fetus in systemic lupus erythematosus.. In: Wallace, DJ.; Hahn, BH., editors. Dubois' lupus erythematosus. Lippincott Williams & Wilkins; Philadelphia: 2002. p. 1023-40.
6. Strom BL, Reidenberg MM, West S, Snyder ES, Freundlich B, Stolley PD. Shingles, allergies, family medical history, oral contraceptives, and other potential risk factors for systemic lupus erythematosus. *Am J Epidemiol*. 1994; 140:632-42. [PubMed: 7942763]
7. Moutsopoulos HM, Gallagher JD, Decker JL, Steinberg AD. Herpes zoster in patients with systemic lupus erythematosus. *Arthritis Rheum*. 1978; 21:789-802.
8. Wang F, Chua CT, Bosco J. Herpes zoster in patients with systemic lupus erythematosus. *Singapore Med J*. 1983; 24:218-20. [PubMed: 6648552]
9. Yell JA, Mbuagbaw J, Burge SM. Cutaneous manifestations of systemic lupus erythematosus. *Br J Dermatol*. 1996; 135:355-62. [PubMed: 8949425]
10. Freni-Titulaer LW, Kelley DB, Grow AG, McKinley TW, Arnett FC, Hochberg MC. Connective tissue disease in southeastern Georgia: a case-control study of etiologic factors. *Am J Epidemiol*. 1989; 130:404-9. [PubMed: 2787600]
11. Petri M, Allbritton J. Antibiotic allergy in systemic lupus erythematosus: a case-control study. *J Rheumatol*. 1992; 19:265-9. [PubMed: 1629825]
12. Pope J, Jerome D, Fenlon D, Krizova A, Ouimet J. Frequency of adverse drug reactions in patients with systemic lupus erythematosus. *J Rheumatol*. 2003; 30:480-4. [PubMed: 12610805]
13. Rubin, RL. Drug-induced lupus.. In: Wallace, DJ.; Hahn, BH., editors. Dubois' lupus erythematosus. Lippincott Williams & Wilkins; Philadelphia: 2002. p. 885-916.
14. Cooper, GS.; Gilkeson, GS.; Dooley, MA.; St. Clari, EW.; Treadwell, EL.; Pandey, JP., et al. Medical history risk factors for the development of SLE: results from a population-based case-control study.. American College of Rheumatology Annual Meeting; 1999.
15. American Community Survey. U.S. Census Bureau: 2003.
16. A nation online: entering the broadband age. U.S. Department of Commerce, Economics and Statistics Administration, National Telecommunications and Information Administration; Sep. 2004
17. Hedley AA, Ogden CL, Johnson CL, Carroll MD, Curtin LR, Flegal KM. Prevalence of overweight and obesity among US children, adolescents, and adults, 1999-2002. *JAMA*. 2004; 291:2847-50. [PubMed: 15199035]
18. Health, United States. With chartbook on trends in the health of Americans. National Center for Health Statistics; Hyattsville, MD: 2005. 2005
19. D'Agostino RB Jr. Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Stat Med*. 1998; 17:2265-81. [PubMed: 9802183]

20. Parsons, LS. Reducing bias in a propensity score matched-pair sample using greedy matching techniques.. Proceedings of the Twenty-sixth Annual SAS Users Group International (SUGI) Conference.; Cary, NC: SAS Institute Inc.. 2001;
21. Cooper GS, Dooley MA, Treadwell EL, St Clair EW, Gilkeson GS. Hormonal and reproductive risk factors for development of systemic lupus erythematosus: results of a population-based, case-control study. *Arthritis Rheum.* 2002; 46:1830–9. [PubMed: 12124867]
22. Fox S, Fallows D. Internet health resources. Pew Internet & American Life Project. 2003:1.
23. Latest trends: who's online. Pew Internet & American Life Project. 2005
24. Wallace, DJ. The clinical presentation of systemic lupus erythematosus.. In: Wallace, DJ.; Hahn, B., editors. *Dubois' lupus erythematosus*. Lippincott Williams & Wilkins; Philadelphia: 2002. p. 621-8.
25. McAlindon T, Formica M, LaValley M, Lehmer M, Kabbara K. Effectiveness of glucosamine for symptoms of knee osteoarthritis: results from an internet-based randomized double-blind controlled trial. *Am J Med.* 2004; 117:643–9. [PubMed: 15501201]
26. Zhang Y, Chaisson CE, McAlindon T, Woods R, Hunter DJ, Niu J, et al. The online case-crossover study is a novel approach to study triggers for recurrent disease flares. *J Clin Epidemiol.* 2007; 60:50–5. [PubMed: 17161754]
27. Zhang Y, Woods R, Chaisson CE, Neogi T, Niu J, McAlindon TE, et al. Alcohol consumption as a trigger of recurrent gout attacks. *Am J Med.* 2006; 119:800, e13–8. [PubMed: 16945617]
28. Etter JF, Perneger TV. A comparison of cigarette smokers recruited through the Internet or by mail. *Int J Epidemiol.* 2001; 30:521–5. [PubMed: 11416075]
29. Roberts RJ. Can self-reported data accurately describe the prevalence of overweight? *Public Health.* 1995; 109:275–84. [PubMed: 7667492]
30. Rowland ML. Self-reported weight and height. *Am J Clin Nutr.* 1990; 52:1125–33. [PubMed: 2239790]
31. Locke SE, Kowaloff HB, Hoff RG, Safran C, Popovsky MA, Cotton DJ, et al. Computer-based interview for screening blood donors for risk of HIV transmission. *JAMA.* 1992; 268:1301–5. [PubMed: 1507376]
32. Turner CF, Ku L, Rogers SM, Lindberg LD, Pleck JH, Sonenstein FL. Adolescent sexual behavior, drug use, and violence: increased reporting with computer survey technology. *Science.* 1998; 280:867–73. [PubMed: 9572724]
33. Baer A, Saroiu S, Koutsky LA. Obtaining sensitive data through the Web: an example of design and methods. *Epidemiology.* 2002; 13:640–5. [PubMed: 12410004]
34. Pope JE, Krizova A, Ouimet JM, Goodwin JL, Lankin M. Close association of herpes zoster reactivation and systemic lupus erythematosus (SLE) diagnosis: case-control study of patients with SLE or noninflammatory musculoskeletal disorders. *J Rheumatol.* 2004; 31:274–9. [PubMed: 14760796]
35. Kahl LE. Herpes zoster infections in systemic lupus erythematosus: risk factors and outcome. *J Rheumatol.* 1994; 21:84–6. [PubMed: 8151595]
36. Massey JT, O'Connor D, Krotki K. Response rates in random digit dialing (RDD) telephone surveys. Proceedings of the Survey Research Methods Section of the American Statistical Association. 1997:707–712.

What is new?

- An Internet-based approach can facilitate epidemiologic studies of uncommon diseases by reaching large numbers of geographically dispersed participants.
- Internet-based case-control studies with an internally valid design are feasible, and can demonstrate convergent construct validity.
- Potential challenges include recruitment of samples representing the source population and confirmation of disease status.
- There is potential to propagate the Internet-based approach to online genetic studies and family-based studies.

* applicants were screened positive for the self-reported screening algorithm for SLE

Fig. 1.
Flow chart of study participant recruitment.

Table 1

Comparison of demographic and socioeconomic characteristics between the general U.S. Internet users and the overall sample of individuals who applied to join as a control

Demographic and socioeconomic characteristics	Prevalence (% in U.S. Internet users ^a)	Prevalence (n [%] in control applicants)
Gender		
Female	51.5	593 (85.6)
Male	48.4	100 (14.4)
Race ^b		
White	81.6	593 (88.0)
African-American	9.1	34 (5.0)
Asian	4.4	7 (1.0)
American Indian	na	8 (1.2)
Native Hawaiian	na	1 (0.1)
More than one race	na	31 (4.6)
Ethnicity ^c		
Non-Hispanic	91.2	641 (96.7)
Hispanic	8.8	22 (3.3)
Age ^d		
Mean (SD)	na	41.3 (12.2)
18–24	14.6	62 (9.2)
25–49	56.6	423 (62.9)
50 and above	28.8	187 (27.8)
Region of residence ^e		
Northeast	19.5	171 (24.7)
Midwest	23.6	154 (22.2)
South	34.0	222 (32.0)
West	22.9	146 (21.1)
Education level (age ≥ 25) ^f		
High school graduate or less	27.5	80 (13.3)
Some college	32.7	240 (40.0)
College graduate	25.0	147 (24.5)
Professional or graduate school	14.7	133 (22.2)
Household income ^g		
Less than \$25K	16.1	157 (25.5)
\$25K–\$50K	26.4	197 (32.0)
\$50K–\$100K	37.4	182 (29.6)
\$100K+	20.0	79 (12.9)

^aStatistics were derived from two 2003 U.S. Census Bureau data sets; one on general population demographics [16], and the other on Internet-user characteristics [17].

^bNineteen (2.7%) control applicants chose “Would rather not say.” Percentages in the table were calculated within participants who reported a race.

^cThirty (4.3%) control applicants chose “Would rather not say.” Percentages in the table were calculated within participants who reported an ethnicity.

^dPercentages among the U.S. Internet users were calculated within people aged 18 and over.

^eRegion of residence was defined as follows: Northeast region: Maine, New Hampshire, Vermont, Massachusetts, Connecticut, Rhode Island, New Jersey, New York, and Pennsylvania; South region: Maryland, Delaware, West Virginia, Virginia, Kentucky, Tennessee, North Carolina, South Carolina, Georgia, Florida, Alabama, Mississippi, Arkansas, Louisiana, Oklahoma, and Texas; Midwest region: North Dakota, South Dakota, Nebraska, Kansas, Missouri, Iowa, Minnesota, Wisconsin, Illinois, Michigan, Indiana, and Ohio; West region: Washington, Idaho, Montana, Wyoming, Oregon, California, Nevada, Utah, Colorado, Arizona, New Mexico, Alaska, and Hawaii.

^fThirty-one (4.5%) control applicants did not report education level.

^gSeventy-eight (11%) control applicants did not report household income.

Table 2

Comparison of the body mass index (BMI), smoking, and alcohol status between control applicants and general U.S. population

	<u>Age-adjusted prevalence (%)</u>	
	U.S. general population	Control applicants
Overweight or obesity (BMI 25, age 20) ^a		
Female	61.6	60.7
Male	68.8	63.3
Current smoking ^b		
Female	19.4	25.6
Male	23.7	40.5
Current drinking ^b		
Female	55.2	58.0
Male	67.1	52.6

^aThe distribution of BMI in the U.S. population was from the National Health and Nutrition Examination Survey of 1999–2002 [18]. To make the data comparable, estimates were age adjusted by the direct method to the 2000 U.S. Census population, using the age groups 20–39, 40–59, and 60 years.

^bThe prevalence of current smoking and drinking in the U.S. population was from National Center for Health Statistics [19]. To make the data comparable, estimates were age adjusted by the direct method to the 2000 U.S. Census population, using the age groups 18–24, 25–34, 35–44, 45–64, and 65 years.

Table 3

Clinical features of confirmed systemic lupus erythematosus (SLE) participants based on chart review

Prevalence in confirmed SLE cases (<i>n</i> = 104)		
Clinical features	<i>n</i>	%
Arthritis	89	86
Malar rash	59	57
Discoid rash	14	13
Photosensitivity	53	51
Oral ulcers	37	36
Serositis	39	38
Renal disease	11	11
Neurologic disorder	5	4.8
Psychosis	2	1.9
Hematologic disorder	50	48
Immunologic disorder	77	74
Antinuclear antibody	100	96

Table 4

Demographic and socioeconomic characteristics of matched cases and controls

Demographic and socioeconomic characteristics	Cases (389)	Controls (389)	<i>P</i> ^a
Gender			
Female	369 (95%)	368 (95%)	0.8
Male	20 (5%)	21 (5%)	
Race			
White	321 (83%)	330 (85%)	0.9
African-American	30 (8%)	27 (7%)	
Asian	4 (1%)	6 (2%)	
Others	7 (2%)	4 (1%)	
More than one race	20 (5%)	17 (4%)	
Do not say	7 (2%)	5 (1%)	
Ethnicity			
Non-Hispanic	362 (93%)	365 (94%)	0.9
Hispanic	17 (4%)	17 (4%)	
Do not say	10 (3%)	7 (2%)	
Age			
Mean (SD)	39.3 (11.0)	39.5 (11.4)	0.5
Region of residence			
Northeast	70 (18%)	68 (17%)	0.4
Midwest	61 (16%)	60 (15%)	
South	171 (44%)	176 (45%)	
West	87 (22%)	85 (22%)	
Education level ^b			
High school graduate or less	66 (17%)	59 (16%)	0.1
Some college	169 (45%)	147 (40%)	
College graduate	79 (21%)	93 (25%)	
Professional or graduate school	64 (17%)	71 (19%)	
Household income ^c			
Less than \$25K	97 (27%)	86 (25%)	0.8
\$25K–\$50K	111 (31%)	124 (36%)	
\$50K–\$100K	101 (29%)	97 (28%)	
\$100K+	45 (13%)	39 (11%)	

^a*P*-value calculated using Wilcoxon rank test for continuous and ordinal variables and McNemar's test for categorical variables.^bEleven (3%) of cases and 19 (5%) of controls did not report education level.^cThirty-five (9%) of cases and 43 (11%) of controls did not report household income.

Table 5

Associations between history of shingles, hives, allergy, miscarriage, and SLE

All 389 case-control pairs		Odds ratio (OR) (95% confidence interval [CI])		One hundred and four confirmed case-control pairs	
Cases (%)	Controls (%)			Cases (%)	OR (95% CI)
History of shingles (herpes zoster) ^a					
No	318 (85)	352 (93)	1.0 (ref)	77 (79)	94 (91)
Yes	55 (15)	27 (7)	2.25 (1.38–3.69)	21 (21)	9 (9)
History of hives (urticaria) ^b					
No	127 (34)	188 (50)	1.0 (ref)	33 (33)	48 (47)
Yes	246 (66)	190 (50)	1.87 (1.39–2.52)	68 (67)	54 (53)
Allergy to sulfonamides ^c					
No	222 (70)	264 (84)	1.0 (ref)	64 (73)	74 (87)
Yes	94 (30)	51 (16)	2.19 (1.48–3.24)	24 (27)	11 (13)
History of miscarriage ^d					
No	89 (35)	104 (60)	1.0 (ref)	26 (38)	25 (54)
Yes	163 (65)	69 (40)	3.03 (1.95–4.70)	42 (62)	21 (46)

^aSixteen cases (4%) and 10 (3%) controls (six cases and one control in the confirmed subset) answered “Do not know” to history of shingles.^bFifteen cases (4%) and 10 (3%) controls (three cases and two controls in the confirmed subset) answered “Do not know” to history of hives. One case-control pair (zero confirmed case-control pairs) was excluded because the case had a history of hive missing.^cSeventy-three cases (19%) and 74 (19%) controls (16 cases and 19 controls in the confirmed subset) answered “Do not know” to allergic to sulfa drugs.^dOne-hundred and eleven cases (31%) and 190 (52%) controls (27 cases and 49 controls in the confirmed subset) had never been pregnant. Twenty-six case-control pairs (nine confirmed case-control pairs) were excluded because one or two of each pair had pregnancy data missing.