

## 연구논문

## 사이버공간과 사회조사: 온라인 사회조사의 쟁점과 과제

Cyberspace and survey research: The methodological implications and issues of online survey research

조 동 기\*

Dong-Gi Jo

사이버공간이 독특한 사회적 공간으로 자리잡고 이에 대한 경험적 자료의 필요성이 증대됨에 따라 컴퓨터 통신망을 이용한 온라인 사회조사가 늘어나고 있다. 본 연구는 사이버공간이 보여주는 기술적 가능성을 방법론적으로 검토해 봄으로써 온라인 사회조사의 발전가능성을 탐색하고 있다. 온라인 사회조사는 표집틀, 응답자 접촉방식, 질문지 형태 등에 따라 회원조사, 방문자조사, 전자우편조사, 전자설문조사로 구분될 수 있으며 기존의 조사방법에 비하여 조사비용의 저렴하고 조사의 실행이 용이하다는 특징을 가지고 있다. 그러나 체계적 오차와 관련하여 온라인 사회조사는 표본설계(모집단의 규정, 표집틀 오차, 추출오차), 측정(측정도구의 구성, 시스템의 안정성과 충실성, 비응답 및 응답오차), 조사윤리의 측면에서 많은 과제를 안고 있다. 이러한 방법론적인 쟁점에 대한 충분한 고려가 이루어질 때 온라인 사회조사는 본격적인 조사연구의 한 방법으로 자리잡을 수 있을 것이다.

This study investigates methodological implications of cyberspace and information technology for online survey research. Online survey method provides useful methodological tools for collecting empirical data for the study of cyberspace. In terms of sampling frame, selection method, and questionnaire form, four major types of the online survey research can be found: member survey, visitor survey, e-mail survey, and the electronic survey. Online survey research has the advantages of low cost and simple administration. However, systematic errors associated with sample design (population specification error, sampling frame error, selection error), measurement (measurement instrument bias, system-related processing error, nonresponse, and response bias), and research ethics (privacy and netiquette) pose important methodological issues for the survey. The realization of the technological possibilities of online survey will be possible only through a right consideration of these issues.

\* 서울대 사회발전연구소 연구원(infowell@infosoc.org)

## 1. 서론

### 1. 들어가는 말

정보기술의 발달에 발맞추어 사회조사를 수행하는 새로운 기법이 꾸준히 모색되어 왔다(Lee, 1995; Jones, 1999, Nesbary, 2000). 컴퓨터를 이용한 전화면접(CATI)이나 개인면접(CAPI), 무작위 번호호출(RDD)는 이미 오래 전부터 활용되어 왔고 보다 최근에는 통신망을 이용한 사회조사가 여러 분야에서 시도되고 있다. 특히 우리 나라 사회조사의 경우 조사비용이 늘어나고 응답거부율이 증가됨에 따라 새로운 방법의 모색이 시급한 실정이다.

최근 들어 급속하게 확장되고 있는 사이버공간은 사회조사의 기법과 대상에 대한 새로운 가능성을 보여 주고 있다. 발달된 정보처리 기술을 활용함으로써 자료수집 도구나 수집 절차를 혁신적으로 개선시킬 수 있게 되었고, 정보통신망을 적절히 활용함으로써 지리적인 경계와 시간적인 한계를 초월하여 조사대상과의 접촉하는 것이 가능해졌다. 또한 사이버공간이 독특한 사회적 공간으로 자리잡고 이 새로운 공간의 연구에 필요한 경험적인 자료의 필요성도 증가됨에 따라 온라인 사회조사의 활용이 늘어나고 있다.

정보통신망을 매개로 하는 온라인 사회조사는 많은 가능성을 가지고 있지만 이러한 기술적 대안들이 방법론적인 엄밀성을 갖추지 못하고서는 본격적인 조사연구의 한 방법으로 발전할 수가 없다. 특히 온라인 사회조사의 기술적인 편리함과 용이함이 지나치게 강조되는 경우에는 조사연구의 오용과 남용이라는 바람직하지 못한 현실이 나타날 수도 있다. 또한 비전문가들에 의해 방법론적인 원칙을 무시한 채 수행되는 많은 온라인 사회조사는 조사 자체에 대한 일반의 불신과 왜곡된 인식을 확산시키는 요인이 될 수 있다.

본 연구는 온라인 사회조사의 발전가능성을 모색하는 차원에서 현재 온라인 조사와 관련해서 제기되는 방법론적인 쟁점과 과제를 논의하고자 한다. 특히 온라인 사회조사의 유형과 특성, 표본설계 및 측정과 관

련된 체계적 오차를 주요 논의 대상으로 삼고자 한다.

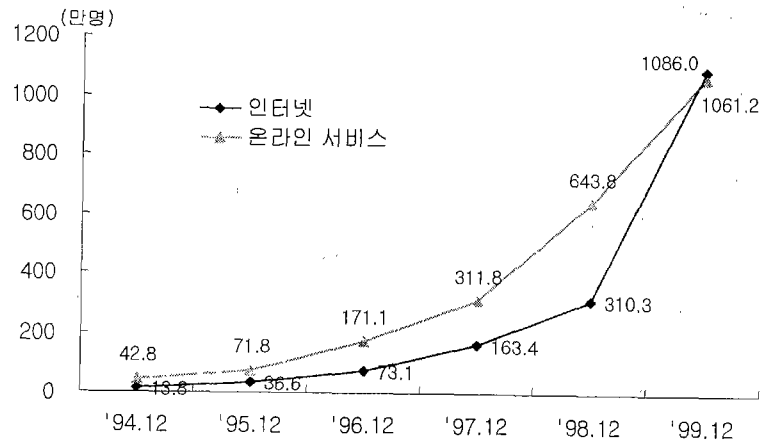
## 2. 사이버공간의 다양성과 사회연구

사이버공간은 일반적으로 컴퓨터와 통신망을 매개로 하여 사람들이 시간과 공간의 제약을 벗어나서 커뮤니케이션을 하는 가상적인 공간으로 정의된다(Jones, 1994; 1998; Rheingold, 1993). 이러한 사이버공간은 그 기술적 하부구조에 따라 다양한 사회적 특성을 지니게 된다(Kitchin, 1998). 사이버공간의 기술적 하부구조는 물리적인 접속장치(컴퓨터, 모뎀 등), 통신망대역(bandwidth: PSTN, ISDN, ADSL 등), 프로토콜(protocol: TCP/IP, PPP, POP 등), 사용자 인터페이스(interface: 문자중심, GUI 등) 등에 따라 정해지며, 사회적 구조는 사용자에게 대한 관리와 통제의 방식, 경계(boundary)와 개방성, 정체성의 단서, 상호작용의 방식 등에서 차이가 날 수 있다.

현재 우리나라의 사이버공간은 인터넷과 온라인 서비스(흔히 “PC 통신망”으로 칭해짐)로 대별될 수 있다. 일반적으로 통신망의 통신망으로 일컬어지는 인터넷이 수많은 컴퓨터들이 상호연결된 세계적이고 개방적인 통신망이라면, 온라인 서비스는 가입자 중심의 지역적이고 폐쇄적인 통신망이라고 할 수 있다.

우리 나라 사이버공간의 크기를 보면, 최근 몇 년 동안 인터넷과 온라인 서비스 모두 그 사용자 수가 급속하게 증가하여 왔다. 인터넷 인구에 대한 추계는 조사기관에 따라 심한 차이를 보이고 있지만(아이클릭, 1999; 인터넷 미트릭스, 1999; 한국인터넷정보센터, 1999), 한국인터넷정보센터에 따르면 99년 12월 현재 인터넷 이용자는 1,070만명에 육박하고 있으며, 이는 94년말의 약 14만명, 96년말의 73만명, 98년말의 310만명이라는 추세로 볼 때 폭발적으로 증가한 것이다.<sup>1)</sup> 최근 들어 그 증가추세가 약간 둔화되고 있는 온라인 서비스 가입자도 99년 12월 현재 1,060만명을

1) 인터넷 인구의 외형적 크기의 이러한 증가는 추계방식의 변화에서도 일부 기인한다.



자료: 인터넷: 한국인터넷정보센터 인터넷통계 월보(<http://www.knmc.net>).  
 온라인 서비스: 통계청 통계정보시스템(<telnet://www.nso.go.kr>).

〈그림 1〉 사이버공간의 성장

넘어서 98년말의 640여만명보다 크게 증가하였다.

사이버공간에서는 동기적인(synchronous) 실시간 대화에서부터 비동기적인 전자우편이나 게시판에 이르는 다양한 커뮤니케이션이 발생하고 있다. 커뮤니케이션의 기술적 특성은 그 과정에서 출현하는 사회적 의미와 문화적 특성까지도 관련되어 있다(Jones, 1998). 이는 연구대상 또는 연구의 장소로서의 사이버공간에 접근하는 방법 또한 다양해야 한다는 것을 의미한다.

사이버공간을 올바르게 이해하기 위해서는 다양한 사회적 구조의 어떤 측면에 초점을 맞출 것인가를 결정하고 거기에 적절한 연구의 방법을 선택하는 것이 중요하다. 온라인 대화에서 형성되는 사람들간의 관계나 사회적 의미의 출현에 관심이 있는 경우에는 참여관찰이나 대화내용에 대한 질적인 분석을 시도할 필요가 있고, 게시판에 등록된 글의 내용보다 글을 등록하는 사람의 수나 글을 읽는 사람의 수에 관심이 있는 경우에는 양적인 분석이 시도될 수 있다. 이 글에서는 사이버공간의 거주자(netizens)를 대상으로 하는 온라인 사회조사의 방법론적 쟁점과 과제를 논의해 보고자 한다.

## II. 온라인 사회조사의 유형과 특성

### 1. 온라인 사회조사의 과정과 유형

#### 1) 온라인 사회조사의 과정

온라인 사회조사는 인터넷, 온라인 서비스, 인트라넷 등 컴퓨터 통신망 상에서 이루어지는 사회조사를 통칭하는 용어이다. 온라인 사회조사의 기본과정은 응답자가 컴퓨터 화면장치를 통해 응답한 내용이 통신망을 통하여 조사기관의 서버(server) 컴퓨터에 전달되어 자료 파일로 저장되는 절차를 따른다. 자기기입식 질문지 조사나 면접조사와 비교해 볼 때 온라인 조사는 질문 내용이 질문지나 면접원의 구술이 아닌 컴퓨터 화면장치로 전달되며 응답내용의 처리도, 부호화과정과 자료입력 과정이 통합되어 컴퓨터에 의해 이루어진다는 특징을 가지고 있다. 이런 특성 때문에 온라인 사회조사에서는 사용자가 화면장치를 통하여 접하게 되는 클라이언트(client) 응용프로그램의 방식과 서버 컴퓨터의 자료 기록과정이 중요한 역할을 담당하게 된다.

클라이언트 응용프로그램의 방식을 일반적으로 인터페이스라고 하며 여기에는 사용자가 마우스와 같은 보조 입력장치를 이용하여 명령을 내리는 그래픽 방식(GUI)과 문자를 입력하여 명령을 내리는 텍스트 방식으로 구분된다. 인터넷상의 대부분의 서비스에서는 그래픽 방식이 통용되나 전자우편의 경우에는 텍스트 방식도 많이 사용되고 있다. 온라인 서비스는 전통적으로 텍스트 방식으로 운영되어 왔고 최근 들어 신규 이용자들을 중심으로 그래픽 방식이 급속히 보급되고 있는 추세지만 장기 이용자들 사이에서는 텍스트 방식이 주종을 이루고 있다.

인터넷의 CGI(Common Gateway Interface)나 온라인 서비스의 DBMS(database management system)는 응답자와 자료 데이터베이스간을 매개하면서, 응답자들에게 질문의 내용을 일정한 순서에 따라 제시하여 주고 응답의 내용을 데이터베이스에 기록하는 역할을 수행한다(Krol and Ferguson, 1995). 이러한 처리과정은 일반 사회조사에서 질문지를 회수하여 응답내용을 부호화시키고 컴퓨터에 입력하는 과정에 상응하는 것이다. 이

과정에서는 또한 응답의 비밀관성이나 무응답에 대한 확인작업도 수행되기 때문에 면접원의 역할도 일부 수행된다. 일반 사회조사에서 많은 시간과 노력이 투자되는 면접, 부호화, 및 입력 과정이 온라인 사회조사에서는 컴퓨터 시스템에 의해 처리되기 때문에 시간과 비용의 상당한 절감 효과가 나타나게 된다.

## 2) 온라인 사회조사의 유형

온라인 사회조사에서 <응답자 - CGI/DBMS - 자료 데이터베이스>라는 기본적인 과정은 동일하지만 조사대상자 선정의 기반이 되는 표집틀(sampling frame)의 형태, 응답자들과 접촉하는 방식, 질문지의 형태 등을 기준으로 하여 몇 가지로 유형화시켜 볼 수 있다. <표 1>에서 정리된 바와 같이 특정한 사이트나 메뉴에 접속하는 사람들이 자발적으로 참여하는 방문자조사, 미리 선정된 회원이나 가입자를 대상으로 하는 회원조사나 전자설문조사, 전자우편 사용자를 대상으로 하는 전자우편조사 등이 있다.<sup>2)</sup>

〈표 1〉 온라인 조사의 4가지 유형

조사의 유형	표 집 틀	응답자 접촉	질문지 형태 <sup>1)</sup>	사이버공간
회원조사	회원 DB	전자우편 (전화)	웹문서	인터넷
방문자조사	-	배너/링크	웹문서 전자설문	인터넷 온라인 서비스
전자우편조사	전자우편목록	전자우편	텍스트 문서 일반 문서	인터넷 온라인 서비스
전자설문조사	가입자 DB	전자우편 (전화)	전자설문	온라인 서비스

주: 1) 일반적으로 사용되는 질문지 형태이며, 이 밖에도 PDF(Portable Document Format)와 같은 이동성 문서도 사용될 수 있다.

2) CASRO(1999)와 Roth(1999)는 인터넷 조사의 유형을 전자우편조사, 웹조사, 인

## (1) 회원 조사

회원 조사(member survey)는 사전에 형성된 사용자들의 풀(pool)이나 가입자 데이터베이스(DB)를 표집틀로 사용하는 조사이다. 회원조사에서 사용되는 회원<sup>3)</sup> 또는 가입자 DB는 특정한 하나의 조사를 위한 것이 아니라 일반적인 목적을 위해 장기적으로 구축되며 성공적인 조사를 위해서는 이 DB의 포괄성과 충실성이 무엇보다도 중요하다. 즉 회원 DB는 통신망 이용자들의 특성을 대변할 수 있을 만큼 포괄적이어야 할뿐만 아니라 정확한 정보를 담고 있어야 한다.

회원조사에서는 이러한 DB를 표집틀로 하여 확률표집을 통해 조사대상을 선정하며 경우에 따라서는 조사대상을 일정한 특성에 따라 할당하는 것도 가능하다. 선정된 조사대상자들에게는 등록된 회원정보를 기초로 하여 전자우편이나 전화를 이용하여 조사참여를 공지하며 질문지는 웹문서로 작성된다.

이 방법에서 회원 DB 전체를 표본으로 삼지 않고 표집을 하는 논리는 회원 또는 가입자 DB가 방대할 경우 조사의 효율성이 떨어지기 때문이다. 모든 회원에게 조사를 공지하는 데에는 상당한 시간과 자원이 소요되며 응답자 집단에서 빈번이용자층이 과잉대표될 가능성도 높다. 회원 조사는 방문자조사에서 나타나는 중복응답자 문제나 전자우편조사에서 나타나는 조사대상의 특정화라는 문제를 극복할 수 있다. 또한 응답자에 대한 정확한 신분의 확인이 가능하기 때문에 비응답자에 대해서도 자세한 분석이 이루어질 수 있다.

---

터랙티브 프로그램조사로 구분하고 있다. 이 글에서 다루어지지 않는 인터랙티브 프로그램 조사는 미리 구조화된 질문 프로그램을 통신망을 통해 전송받아 질문에 응답하는 방법으로서, 질문 프로그램을 디스켓의 형태로 배포하고 회수하던 종전의 방식이 통신망을 통한 전송으로 변화된 것으로 일종의 CASI(컴퓨터를 이용한 자기기입식 조사)이다.

- 3) 회원 DB는 흔히 "패널(panel)"로 칭해지고 있으나, 이는 종단적 연구의 한 형태로서의 패널과 혼동될 수 있기 때문에 여기서는 회원DB라는 용어를 사용한다.

## (2) 방문자조사

방문자조사(visitor survey)는 특정한 사이트나 메뉴에 조사에 대한 링크나 배너를 만든 후 거기를 통해 들어온 사용자들이 자발적으로 조사에 참여하는 방식이다. 방문자조사에서 응답자의 표집은 “참여에 의한 선정”이라는 방법으로 이루어지며 이는 일종의 편의표집이다. 언론사의 웹사이트에서 흔히 볼 수 있는 간이조사(spot poll)나 긴급조사(quick poll)가 방문자조사의 대표적인 예라고 할 수 있다. 방문자조사에서 확률표집은 특정 사이트에 들어오는 모든 이용자들에게 조사를 노출시키는 것이 아니라 무작위적으로 선별하여 조사에 참여시키는 경우에 가능하다. 방문자조사의 질문지 형식은 인터넷의 경우 웹문서로 작성되는 것이 일반적이며 온라인 서비스에서는 텍스트 방식의 전자설문이 대부분이다.

방문자조사의 성패는 가능한 한 많은 사용자들이 참여할 수 있도록 적절한 경로를 통해 조사를 효과적으로 공지하는 방식에 의해 크게 좌우된다. 인터넷의 경우 많은 사용자들이 빈번하게 드나드는 검색 엔진이나 웹메일 서비스와 같은 소위 관문(portal) 사이트나 대중적인 언론 사이트가 될 수 있다. 조사 주제와 관련된 전자게시판이나 뉴스그룹에 공지하는 것도 유용한 방법이 될 수 있다. 온라인 서비스의 경우에도 사용자들이 자주 보게 되는 게시판에 조사를 공지하는 것이 가장 효율적이다.

방문자조사는 응답자들이 익명으로 참여할 수 있기 때문에 민감하거나 위협적인 질문에 대한 비응답이 적다는 장점이 있지만 중복응답자를 선별하고 처리해야 하는 한계를 지니고 있다. 방문자조사는 일반적으로 사전조사나 간이조사에 적합한 방식이다.

## (3) 전자우편조사

전자우편조사(e-mail survey)는 전자우편에 질문지를 포함시키거나 첨부시켜 조사 대상자들에게 발송하고 응답자들이 완성된 질문지를 다시 전자우편의 형태로 회신하는 방식이다. 이러한 실행과정은 일반적인 우편조사와 매우 유사하다. 다만 전자우편은 “달팽이”(snail mail)로 비유되



는 일반 우편보다 신속하게 오갈수 있고 대부분의 전자우편 시스템에서 가능한 메일링 리스트(mailing list) 기능을 활용하면 질문지를 다수의 수신자들에게 효과적으로 발송할 수 있다.

전자우편조사의 장점은 간편함과 용이함이다. 특히 간단한 텍스트형 질문지를 사용하는 경우에 질문지의 작성과 발송이 용이하고 단순하기 때문에 누구나 손쉽게 조사를 시작할 수 있으며, 문자처리 프로그램을 이용하면 회신된 응답내용의 처리도 비교적 쉽게 끝낼 수 있다(Swoboda *et al.*, 1997). 그러나 응답의 형태가 일정하지 않거나 개방형 질문이 많이 포함된 경우에는 다른 방법에 비하여 자료정선에 많은 시간과 노력이 들어간다. 또한 전자우편조사에서는 텍스트나 일반문서가 주로 사용되기 때문에 온라인 상에서 가능한 다양한 기법을 제대로 활용하지 못한다는 한계를 가지고 있다.

전자우편조사에서는 조사의 목적에 적합한 조사대상자들의 전자우편 주소를 확보하는 것이 중요하다. 특정한 단체에 가입된 성원들의 전자우편 주소나 연구 주제에 관련된 게시판이나 뉴스그룹에 글을 올린 사람들의 전자우편 주소가 중요한 원천이 될 수 있다. 그러나 조사내용이나 대상이 특정화되어 있는 경우를 제외하면 표집틀로 사용할 수 있는 포괄적인 전자우편 주소 명부를 확보하는 것이 거의 불가능하기 때문에 전자우편조사에서 표집은 대부분 편의표집의 방식으로 이루어지고 있다. 또한 다수의 사용자들에게 질문을 무차별적으로 발송하는 행위는 소위 스팸(spamming)으로 간주되어 비용답이 높아질 수 있다.

#### (4) 전자설문조사

“전자설문조사”(the electronic survey)는 온라인 서비스에서 실시되는 사회조사를 일반적으로 지칭하는 것으로, 현재의 조사수행 관행은 회원조사와 방문자조사의 중간적인 성격을 가지고 있다. 가입자 DB를 근거로 한다는 점에서 회원조사와 유사하지만 조사대상의 선정은 방문자조사와 마찬가지로 편의표집으로 이루어진다. 즉, 초기화면에서 조사의 공지에 접한 사용자들이 자발적으로 조사 포럼에 들어가서 응답을 하는 방식이

다. 기술적으로는 가입자 DB를 토대로 하여 확률표집을 하거나, 방문자 조사와 같이 무작위적으로 조사를 노출시키는 방식이 가능하다.

표집틀의 측면에서 볼 때 전자설문조사에서 사용되는 가입자 DB는 인터넷 상의 회원 DB와는 달리 포괄성과 충실성이 비교적 높은 편이다. 인터넷 상의 회원 DB는 인터넷 사용자의 일부만을 담고 있는 반면에 온라인 서비스 가입자 DB는 특정 온라인 서비스에 가입한 모든 사용자를 담고 있다. 따라서 확률표집을 활용하면 대표성이 높은 표본을 만드는 것이 가능하다. 사전표집을 하지 않는 경우에도 대부분의 온라인 서비스의 경우 하루에 수십만명의 가입자들이 일상적으로 이용한다는 점을 고려할 때, 초기화면에서 조사가 공지되는 경우에는 방문자조사 보다 대표성이 높은 표본을 구성할 수가 있다.

전자설문조사에서 질문지 작성은 온라인 서비스 이용자의 상당수가 텍스트 방식을 사용하고 있기 때문에 그래픽 방식의 질문지 구성이 어렵고 단순하게 구성해야 한다는 한계를 지니고 있다.

## 2. 온라인 사회조사의 특성

온라인 사회조사의 조사방법론적인 특징을 검토하기 위해서는 회원조사, 전자설문조사, 방문자조사, 전자우편조사를 일반적인 조사연구 방법들과 비교해 볼 필요가 있다. 편의상 <표 2>에서와 같이 비교의 영역을 크게 조사의 실행적 요인, 질문지 요인, 및 자료의 질적 요인으로 구분하여 검토하고자 한다.

### 1) 조사의 실행적 요인

조사의 실행과 관련된 요인으로 비용, 수집기간, 정리기간, 그리고 표본의 지리적 분포 등을 고려해 볼 필요가 있다. 우선 비용의 측면에서 온라인 조사는 비용이 비교적 적게 드는 조사연구 방법이다. 자료의 수집기간도 다른 방법에 비하여 단기적이라고 할 수 있다. 전화조사와 마찬가지로 응답이 바로 전달되고, 응답결과도 직접 기록되기 때문에 자료의 회수가 일반적인 조사방법에 비하여 단기간에 이루어진다. 응답자 집

단의 특성에 많이 의존하는 전자우편조사의 경우 1~4주 정도의 시일이 소요되며 회원조사나 방문자조사의 경우에는 이보다 더 단축될 수 있다.

이러한 측면은 자료정선(data cleaning) 기간의 단축과 밀접하게 관련되어 있다. 특히 CGI나 DBMS가 효율적으로 활용되는 회원조사나 전자설문조사, 그리고 방문자조사의 경우에는 비논리적인 응답에 대한 확인을 통하여 불성실 응답을 줄일 수 있고 응답 내용도 바로 전자적 매체에 기록되기 때문에 부호화과정에서 나타날 수 있는 입력오류도 배제시킬 수가 있다.<sup>4)</sup> 표본의 지리적 분포는 광역적일 수 있지만 현재 통신망 사용자의 분포는 대도시 중심적이다. 그러나 온라인 조사는 자기기입식 조사이기 때문에 조사의 대상이 공간적으로 집락화되어 있을 필요가 없다.

## 2) 질문지 요인

질문지와 관련해서는 질문지의 구성, 문항의 복잡성, 질문의 형태 및 내용, 보조 자료 활용 등과 같은 요인들이 검토될 수 있다. 우선 사용자들의 통신 이용 특성을 고려할 때 질문지의 길이는 회원조사나 전자설문조사의 경우는 15분~30분내에 응답이 마무리될 수 있도록 하는 것이 좋다(CASRO, 1999). 방문자조사의 경우에는 사용자들의 자발적인 참여에 크게 의존하기 때문에 질문지의 길이는 5~10이내에 마무리 될 수 있는 정도가 좋다. 전자우편조사의 경우에도 20분이내에 응답이 완료될 수 있는 정도로 질문지 길이를 한정할 필요가 있다.

질문지 구성의 복잡성이라는 측면에서 온라인 조사는 중간정도라고 할 수 있다. 자기기입식 질문지라는 점에서 단순하게 구성하는 것이 원칙이지만, 각종 기법을 활용하여 질문의 순서를 통제하는 것이 가능하기 때문에 일반적인 자기기입식 조사보다는 훨씬 더 복잡한 구조를 가질 수 있다. 전자우편조사를 제외한 나머지 온라인 조사에서는 질문을 내용에 따라 구역화시키거나 분할형(contingency) 질문을 효과적으로 사용함으로써

4) 다만 전자우편조사의 경우에는 질문지의 형태에 따라 자료의 다른 온라인 조사방법에 비하여 오래 걸릴 수 있다.

써 응답자들이 정해진 순서대로 응답하도록 유도할 수 있다. 그러나 온라인 조사에서 질문 문항 자체는 비교적 간단할 필요가 있다. 기본적으로 우편 조사와 마찬가지로 자기기입식 조사이기 때문에 복잡한 질문 형식은 응답편의(response bias)를 초래하기 쉽기 때문이다.

온라인 조사에서 개방형 질문의 사용은 양호한 편이다. 면접원이 개방형 질문에 대한 상세한 응답지침을 제공해 줄 수 있는 전화조사나 면접 조사에서처럼 뛰어나지는 않지만, 응답자가 스스로 편집해 가면서 응답 내용을 제시하는 것이 가능하기 때문에 일반적인 자기기입식 질문지 조사방법에 비하여 양호한 결과를 얻을 수가 있다.

시각적(visual) 자료의 활용은 온라인 조사가 가지는 가장 뛰어난 측면 중의 하나이다. 면접조사에서도 사진이나 도표 등과 같은 시각적 자료의 활용이 가능하지만 온라인 조사에서만만큼 원활하지는 못하다. 특히 웹문서에서는 그림이나 사진, 동화상 등 멀티미디어 자료도 활용할 수 있다. 가구나 개인에 관련된 기록 자료의 이용도 우수한 편이다. 전화조사와 마찬가지로 온라인으로 진행되는 조사이기 는 하지만 응답자 스스로가 응답 상황을 통제하기 때문에 기록자료의 활용이 전화조사보다 우수한 편이다.

전화조사나 면접조사와는 달리 면접원과의 대면적 또는 간접적 접촉이 결여되어 있기 때문에 래포(rapport)의 형성이 뛰어나지는 않지만, 상호작용적 기법이나 이미지 등을 활용하면 상당한 정도의 래포형성이 가능하다. 민감한 주제를 다루는 데에도 온라인 조사는 우수한 방법이 될 수 있다. 온라인 사회조사에서는 비면접성으로 인한 상대적 익명성의 효과가 있기 때문에 민감한 주제에 대한 응답을 비교적 용이하게 이끌어낼 수 있다.

〈표 2〉 온라인 사회조사의 특성 비교

주요 쟁점	조사방법		회원조사 전자 설문조사	방문자조사	전자우편조사	우편조사	전화조사	면접조사
	실행비용	자료수집기간						
실행 요인	실행비용	저비용	저비용	저비용	저비용	저비용·중간	저비용·중간	고비용
	자료수집기간	단기(1-2주)	단기(1-3주)	단기(1-4주)	장기(10주)	단기(2-4주)	중·장기(4-12주)	
	자료정선	단기	단기	단기·중기	장기	장기	단기·중기	중기·장기
	표본의 지리적 분포	광역(제한적)	광역(제한적)	광역(제한적)	광역	광역	광역	집락
	질문지의 길이	중간(15-30분)	짧게(5-10분)	중간(10-20분)	중간(4-12 쪽)	중간(15분-30분)	중간(15분-30분)	길게(30-60분)
	질문지 구성의 복잡성	복합적 구조 가능*	중간	간단하게	간단하게	복합적 구성 가능	복합적 구성 가능	복합적 구성 가능
	문항의 복잡성	비교적 간단하게	비교적 간단하게	짧고 단순하게	비교적 간단하게	짧고 단순하게	짧고 단순하게	복합적 질문 가능
	질문순서의 통제	우수	우수	열등	열등	열등	매우 우수	매우 우수
	개인형 질문 사용	양호	양호	양호	열등	열등	양호	우수
	시각적 자료 활용	매우 우수*	매우 우수	우수	우수	우수	대개는 불가능	매우 우수
자료의 질적 요인	가구·개인 기록자료 이용	우수	우수	우수	우수	매우 우수	양호	우수
	레포(rapport) 형성	양호	양호	양호	양호	양호	우수	매우 우수
	민감한 주제	우수	우수	우수	우수	우수	양호·우수	양호
	표집률 편기	대체로 높음	매우 높음	높음	대체로 낮음	낮음**	낮음**	낮음
	응답률	(다양)	계산불가능	(다양)	25%-70%	30%-70%	65%-95%	
	응답 편기	중간	높음	중간	중간	낮음	낮음	낮음
	거부자·미집속자 분석	(다양)	열등	(다양)	열등	열등	열등	양호
	응답상황 통제	열등	열등	열등	열등	열등	열등	우수
	기록된 응답의 질	우수	양호	양호·우수	양호	양호	매우 우수	매우 우수

주: \* 회원조사의 경우에만 해당 \*\*무작위번호추출(RDD)의 경우

### 3) 자료의 질적 요인

모집단에 대한 표본의 대표성을 보여주는 자료의 질(quality)은 표집틀, 표본 크기, 구체적인 표집방법 등에 의해 결정된다. 확률표집의 바탕이 되는 표집틀(sampling frame)은 포괄성(comprehensiveness), 표집요소의 추출확률(probability of selection) 계산가능성, 효율성(efficiency) 등의 측면에서 평가될 수 있다(Fowler, 1993; 홍두승, 1993). 그러나 온라인 조사에서는 이러한 기준에서 제대로 된 표집틀을 확보하는 것이 어렵기 때문에 다른 방법에 비하여 표집틀에 의한 편기가 크게 나타날 수 있다. 이러한 편기는 장기간에 걸쳐 구축된 회원 DB나 가입자 DB를 사용하는 회원조사나 전자설문조사에서보다 편의표집에 의존하는 방문자조사나 전자우편조사의 경우에 특히 문제가 될 수 있다.

온라인 조사에서 응답률은 응답자 선정, 질문지의 발송 및 회수 방식에 따라 다양하게 나타날 수 있다. 회원조사나 전자설문조사에서 사전 표집을 통하여 조사대상자를 미리 선정하는 경우에 비교적 높은 응답율을 기대할 수 있다. 그러나 편의표집이 사용되는 방문자조사의 경우에는 실제 응답율을 계산하는 것이 불가능하거나 무의미하며, 전자우편조사의 경우에도 정확한 응답율의 계산은 쉽지가 않다.

응답편기도 실제 사용된 방법에 따라 다양하게 나타날 수 있으며 일반적인 조사에 비하여 높다고 할 수 있다. 온라인 조사에서 응답편기는 주로 응답자의 기술적 특성에 의해 생겨난다. 예를 들면 일반문서가 첨부된 전자우편조사에서는 첨부된 파일을 처리할 수 없는 초보 이용자들의 응답이 저조할 수 있다. 응답자들의 자발적인 참여에 의지하는 방문자조사나 편의표집을 활용하는 전자설문조사의 경우에는 소량이용자들의 응답이 낮게 나올 수 있다. 회원조사의 경우에는 표집된 회원에 대한 통보과정에서 전화나 전자우편을 통하여 조사 대상자들과 접촉할 수가 있기 때문에 다른 방법에 비하여 응답편기가 낮은 편이다.

온라인 조사에서 응답편기의 다양성은 비응답자(nonresponses)나 응답거부자(refusals)에 대한 분석의 다양성과 관련되어 있다. 전자우편조사나 방문자조사의 경우에는 개별 조사대상자에 대한 정보가 없기 때문에 비응답자나 응답거부자에 대한 분석이 거의 불가능하다. 반면에 회원 조사

나 전자설문조사의 경우에는 표집된 조사대상자들의 정보를 활용하기 때문에 비응답자 분석이 가능하다. 면접자와의 접촉이 결여되어 있기 때문에 온라인 조사에서는 일반 우편조사와 마찬가지로 응답상황에 대한 통계가 거의 불가능한 편이다. 응답된 자료의 질은 대체로 양호하다고 할 수 있다.

### Ⅲ. 온라인 사회조사의 쟁점과 과제

온라인 사회조사와 관련된 방법론적인 쟁점은 비확률적 또는 체계적 오차를 중심으로 논의해 보고자 한다. 사회조사에서 비확률적 오차는 크게 표본설계에 관련된 것과 측정에 관련된 것으로 구분해 볼 수 있다. 이와 함께 온라인 사회조사에서 제기되는 윤리적인 쟁점도 논의될 필요가 있다.

#### 1. 온라인 사회조사의 표본설계

사회조사에서 표본설계(sample design)와 관련된 비확률오차는 모집단에 대한 불명확한 규정, 표집틀의 편기, 추출오차 등을 들 수 있다.

##### 1) 모집단 규정의 문제<sup>5)</sup>

온라인 사회조사에서 모집단 규정이 특히 문제가 되는 것은 모집단 자체가 비정형의 상태이기 때문이다. 최근 들어 급변하고 있는 사이버공간에 대한 적절한 규정을 내리기란 쉽지 않으며, 특히 인터넷의 경우에는 외적인 경계가 급속하게 확장되고 있는 국면이기 때문에 모집단에 대한

5) 온라인 사회조사에서 가장 흔히 제기되는 쟁점은 표본의 대표성이다. 그런데 현재 통신망 이용자들의 구성이 일반 모집단의 인구학적 특성을 반영하지 못한다는 것은 분명하기 때문에 표본의 대표성의 쟁점은 통신망 사용자 자체를 모집단으로 하는 경우로 한정시켜 논의될 필요가 있다.

규정이 어려운 문제가 되고 있다. 한국인터넷정보센터에서는 인터넷 사용자를 “모든 인터넷 서비스를 상시 이용가능한 인터넷 서비스 가입자” (한국인터넷정보센터, 1999)로 규정하고 있으나 사회조사를 위한 모집단 규정으로는 불명확한 정의이다. 반면에 온라인 서비스의 경우는 서비스에 가입한 사용자로 모집단을 규정하면 큰 문제는 없다.

모집단에 대한 규정은 분석의 단위(unit of analysis)와 그 경계를 명시함으로써 이루어진다. 연구의 목적에 따라 분석의 단위는 개인, 가구, 호스트(host) 또는 도메인(domain) 등이 될 수 있으며(December, 1996), 통신망 이용자들을 대상으로 한 대부분의 온라인 사회조사에서는 개인으로 설정된다. 일반적으로 개인이 분석의 단위가 되는 경우 모집단은 지역적 경계, 연구 모집단의 기본적 특성, 응답자의 연령, 그리고 주거단위 등에 따라 설정된다(Czaja and Blair, 1996). 온라인 사회조사에서 모집단에 대한 정의도 이러한 차원들을 중심으로 정의될 수 있으나 몇 가지에 유의할 필요가 있다.

첫째, 사이버공간의 경우 지리적 경계의 설정이 어렵다. 컴퓨터 통신망은 전세계 어느 곳이나 연결될 수 있기 때문에, 연구의 대상을 대한민국의 영토 내에서 접속하는 사람으로 한정시킬지 아니면 외국에서 접근하는 사람도 포함시킬 것인지를 결정해야 한다. 둘째, 연구모집단의 특성규정은 응답자의 기본적인 특성과 거주지역에 대한 명시인데, 통신망 사용자라면 반드시 거주지역에 따른 제한을 둘 필요는 없을 것이다. 셋째, 응답자의 연령은 연구의 성격에 따라 다양하게 규정될 수 있는데, 통신망 인구, 특히 온라인 서비스 이용자의 상당수가 중고등학생이라는 점을 고려할 필요가 있다.<sup>6)</sup> 마지막으로 일반적으로 주거단위는 표집틀과 관련하여 가구로 설정되나 온라인 조사의 경우에는 주거단위의 제약을

6) 인터넷 사용자의 연령에 대한 규정은 자료를 발표하는 기관에 따라 매우 다양하다. 한국인터넷정보센터(<http://www.knic.net>), GVU(<http://www.gvu.gatech.edu>), Media Metrix(<http://www.mediametrix.com>) 등에서는 사실상 연령의 제한이 없고, Harris Interactive(<http://www.harrisonline.com>) 19세이상, Commerce Net(<http://www.commerce.net>) 16세이상, 아이클릭(<http://www.eyeclick.co.kr>) 13~49세, 인터넷 미트릭스(<http://www.internetmetrix.com>) 15~49세 등이다.



받지 않는다는 점도 고려될 필요가 있다. 통신망 접속은 가구 내에서뿐만 아니라, 근무지, 학교, 집단시설 등 여러 장소에서 이루어질 수 있다.

요컨대 일반적인 사회조사의 비교적 정형적인 모집단과는 달리 급속한 팽창 상태에 있는 통신망 사용자 인구를 대상으로 하는 온라인 사회조사에서는 모집단에 규정이 중요한 고려 사항이 되어야 한다.

## 2) 표집틀의 문제

표집틀 오차는 표본추출의 기반이 되는 표집틀이 모집단의 특성을 제대로 반영하지 못하는 경우에 발생한다. 앞에서 언급되었듯이 표집틀을 평가하는 기준은 포괄성, 효율성, 동일한 추출확률 등인데, 어떤 체계적인 요인에 의해 모집단 구성요소의 일부가 누락되거나 배제되는 경우에 포괄성이 떨어진다. 표집틀의 효율성은 외래적 요소나 중복된 요소가 많이 포함되어 있는 경우에 문제가 될 수 있다. 또한 확률표집이 가능하기 위해서는 표집틀에 포함된 구성요소의 추출확률을 확인할 수 있어야 하는데 표집단위가 혼재하거나 중복된 요소가 많은 경우에는 추출확률의 계산이 불가능하게 된다.

회원조사나 전자설문조사에서는 DB가 표집틀로 사용되는데 두 조사의 DB는 상이한 특성을 가지고 있다. 즉, 가입자 DB는 온라인 서비스 가입자 전체를 포함하고 있는 반면에 회원 DB는 인터넷 사용자의 일부만 담고 있다. 따라서 회원조사에서 표집틀에 의한 오차가 크게 나타날 수 있으며, 이를 감소시키기 위해서는 회원 DB의 구축과정에서 정밀한 설계가 요구된다. 온라인 서비스의 가입자 DB는 타당한 표집틀이기는 하지만 현재의 조사 관행 하에서는 확률표집이 아닌 편의표집이 사용되기 때문에, 다량이용자의 과잉대표라는 선택편기(selection bias)가 나타날 수 있다(Berk, 1983; 남춘호, 1998 참조).

편의표집에 의존하는 일반적인 전자우편조사나 방문자조사에서는 표집틀의 개념을 적용시킬 수가 없다. 다만 조사대상이 특정화되어 있는 전자우편조사에서는 전체 조사 대상의 전자우편 주소 명부를 표집틀로 한 확률표집이 가능하다. 이러한 전자우편조사에서는 표집틀의 효율성을 주의깊게 살펴볼 필요가 있다. 한 사용자가 다수의 전자우편 계정을 보유

하는 경우가 많고<sup>7)</sup> 중복여부를 확인하는 것이 쉽지 않기 때문이다.

표집틀에 의한 체계적 오차를 줄이기 위해서는 DB의 구축이나 조사 노출 과정에서 관문(portal) 사이트나 특성이 다른 다양한 사이트를 활용할 필요가 있다. 엄밀한 회원가입 절차를 마련하거나, IP 주소나 전자우편 주소 등을 이용하여 응답자의 중복성을 선별함으로써 효율성도 높일 수가 있다.

전체 수준에서의 표집틀의 확보가 불가능한 경우에는 다단계의 층화 또는 집락 표집을 활용할 수 있다. 그러나 사이버공간에서 층화 또는 집락의 기준으로 사용될 수 있는 변수를 설정하는 것은 쉽지가 않다. 온라인 사회조사에서는 물리적 및 사회적 기준뿐만 아니라 ISP 유형, 접속시간대, 서비스 영역 등과 같은 기술적 특성도 고려되어야 한다.

### 3) 추출오차

추출오차(selection error)는 조사 참여자를 선정하는 과정에 체계적인 편기가 생김으로써 발생한다. 표집틀이 완벽하다고 하더라도 응답자의 선정이 부적절한 절차에 따라 이루어지면 추출오차가 커지게 된다.

#### (1) 표집틀의 주기성과 응답자 적격성

회원조사나 전자설문조사에서 DB를 표집틀로 한 계통표집이 이루어지는 경우 자료의 배열에 어떤 주기성(periodicity)이 있는지를 확인해야 한다. 표집간격(sampling interval)과 자료의 주기성이 결합되어 추출오차가 크게 나타날 수 있기 때문이다.<sup>8)</sup> 이러한 문제는 전자우편 주소 명부를 표집틀로 하여 확률표집하는 전자우편조사에서도 나타날 수 있다. 추출오차가 보다 크게 문제가 되는 경우는 방문자조사이다. 사이트 특성, 요일, 시간대 등에 따라 특정 사이트에 접속하는 이용자의 수와 특성이 상

7) 한 조사에 따르면 국내 통신망 사용자 중 2개이상의 전자우편 계정을 가지고 있는 사람의 비율이 88%에 이른다고 한다(<http://www.daum.net>).

8) 표집틀에서 자료가 주기성을 보이지 않고 단순히 분류(sort)되어 있는 경우에는 오히려 표집오차를 감소시키는 효과를 얻을 수 있다(Rossi et al., 1983).

당히 다를 수 있기 때문에 정밀한 표본설계가 수반되지 않으면 추출오차가 크게 나타날 수 있다.

추출오차와 관련하여 제기되는 온라인 조사의 또 다른 쟁점은 응답자의 적격성(eligibility) 확인이다. 익명성이 상대적으로 높고 정체차용 impersonation)의 경향이 강한 사이버공간에서는 사용자의 신분을 확인할 방법이 제한되어 있기 때문에 자격의 미비나 불일치, 또는 중복응답에 따른 자료의 타당성이 문제가 될 수가 있다(Smith and Kollock, 1998). 특히 중복응답자의 처리는 온라인 사회조사에서 해결되어야 할 중요한 쟁점이다. 회원조사나 전자설문조사의 경우에는 DB를 활용하기 때문에 중복응답은 사전에 선별될 수 있으나 사전 선별이 불가능한 방문자조사나 전자우편조사에서는 중복응답이 자료의 질을 크게 손상시킬 수 있다.<sup>9)</sup>

방대한 사이버공간에서 다양한 층의 사용자집단을 효과적으로 포착하기 위해서는 일정한 영역을 포괄하는 다수의 조사 사이트를 연계하여 측점을 다원화시킬 필요가 있다. 특히 언론사, 검색엔진, 쇼핑몰 등 사이트의 종류에 따라 이용자의 특성이 다르기 때문에 측점의 다원화를 통해 사이트 특성이나 접속 시간대에 따른 추출오차를 감소시킬 수가 있다.

## (2) 표본의 크기와 응답율

일반 사회조사에서 표본의 크기는 조사비용이나 조사기간과 직접적인 관련이 있기 때문에 표본설계에서 중요한 고려사항이 되지만, 온라인 사회조사에서는 표본단위당 조사비용이 크지 않기 때문에 표본의 크기가 결정적인 제약 요인은 아니다. 물론 분석에 필요한 최소한의 목표 표본 수는 확보되어야 하겠지만 표본의 크기는 유동적으로 결정될 수 있다.

온라인 조사에서 표본의 크기는 조사의 기간과 밀접한 관련을 가지고

9) 응답자에 대한 사전 정보가 없는 경우 IP 주소가 중복응답자를 선별할 수 있는 하나의 단서가 될 수 있다. 그러나 IP 주소를 이용한 중복응답자 판별은 몇 가지 한계를 지니고 있다. 사용자들이 가정에서 ISP나 학교를 통해 PPP로 인터넷에 접속하는 경우에는 IP가 가변적일 수 있고 고정된 IP라고 하더라도, 학교나 직장의 컴퓨터 실습실에 비치된 컴퓨터와 같이, 여러 사람이 사용하는 컴퓨터의 경우에는 IP가 개별 사용자를 구분하는 적절한 근거가 되지 못한다.

있다. 네티즌의 통신행태는 요일에 따라 그리고 하루 동안에도 시간대에 따라 주기를 보이는 것으로 알려져 있다(골드뱅크, 1997; 광고단체연합, 1999; 아이클릭, 1999; 인터넷 미트릭스, 1999; 정보문화센터, 1999; Jones, 1994; Smith and Kollock, 1998). 특히 주중과 주말의 사용자 특성이 다를 수 있고 하루 중에서도 주간과 야간의 접속자 특성이 다를 수 있다. 이러한 측면을 감안하여 온라인 사회조사는 24시간을 개방하여 최소한 일주일 동안 실시될 필요가 있다. 일반적으로 일주일의 기간이면 목표 표본수가 충분히 확보될 수 있으나 그렇지 못한 경우에는 다시 일주일 단위로 조사기간을 연장시킬 필요가 있다.<sup>10)</sup>

응답률은 조사의 질을 보여주는 중요한 척도로서 높을수록 조사의 질이 높아진다. 온라인 조사에서 응답률의 계산은 표집의 방식에 따라 달라질 수 있다. 회원조사나 전자설문조사와 같이 표집틀을 이용한 확률표집이 이루어지는 경우에는 응답율이 조사의 질에 대한 유용한 지표가 될 수 있다. 그러나 편의표집에 의존하는 방문자조사나 전자우편조사, 그리고 사전에 표집을 하지 않는 회원조사나 전자설문조사의 경우에는 일반적인 의미에서의 응답률의 계산은 불가능하거나 그 자체가 무의미하다.

이러한 조사에서는 응답율을 대신할 수 있는 새로운 개념을 도입할 필요가 있다. 하나는 “사후표집율(post sampling rate)”로서, 이는 목표 표본수를 일정한 기간 동안 조사에 참여한 사례수로 나눈값으로 계산된다. 사후표집율은 일정 기간 동안 조사에 참여한 사례 중에서 일부(흔히 목표 표본수)를 사후적으로 표집하여 사용하는 경우에 적용될 수 있는 개념이다.

이와 함께 “응답밀도(response density)”라는 개념도 유용할 수 있다. 응답밀도는 주어진 기간 동안 응답자들의 참여가 얼마나 집중적으로 이루어지는가에 대한 지표로서 일간 응답밀도와 시간당 응답밀도로 구분해 볼 수 있다. 주중과 주말에 따라 일간 응답밀도의 차이가 날 수 있고, 시

10) 조사기간을 일주일 단위로 편성하는 경우 목표 표본수보다 많은 응답자들이 조사에 참여할 가능성이 높다. 그러나 그 크기가 지나치게 클 경우에는 목표 표본수만큼만 사후적으로 추출하여 사용하는 방법이 효과적이다.

간당 응답밀도는 하루 중 주야간의 응답 집중도의 차이를 보여줄 수 있다.

4) 대표성의 평가

〈표 3〉은 자료간 비교가 가능한 변수를 중심으로 인터넷과 온라인 서비스의 회원 DB 특성 및 온라인 조사의 응답자 특성을 요약한 것이다. 인터넷의 회원 DB(A)와 DB(B)는 회원조사의 방식으로 온라인 조사를 전

〈표 3〉 온라인 사회조사의 응답자 특성 비교

(단위: %)

구분	자료명 <sup>1)</sup>	회원 DB (A)	회원 DB (B)	MBC	광고단체 연합 <sup>2)</sup>	가입자 DB(C)	가입자 DB(D)	센서스
	작성시기	99.8	99.10	99.6	99.5	99.9	99.10	95.11
	전체(명)	-	-	20,030	12,685	-	-	44,608,726
성별	남 자	79.6	72.6	64.5	79.9	76.3	70 <sup>4)</sup>	50.2
	여 자	20.4	27.3	35.5	20.1	23.7	30 <sup>4)</sup>	49.8
	계	100	100	100	100	100	100	100
연령별	10~19세	6.0	12.2	6.1	12.8	12.8	12.9	20.1
	20~29세	63.7	59.3	68.8	58.1	43.2	43.7	22.4
	30~39세	27.1	21.2	21.1	29.2	28.9	27.4	22.2
	40~49세	3.2	4.7	4.0		11.7	13.1	14.7
	50세이상		2.6			3.3	3.0	20.5
	계	100	100	100	100	100	100	100
지역별	서울	38.3	35.2	(n.a.)	35.8	38.4	32.8	22.9
	광역시	25.5	25.3		24.1 <sup>3)</sup>	20.9 <sup>3)</sup>	26.1	25.0 <sup>3)</sup>
	기타	36.2	39.5		40.1	40.7	41.2	52.1
	계	100	100		100	100	100	100

주: 1) 회원 DB 및 가입자 DB는 비공개자료 (전체크기 생략).  
 2) 연령대의 구분이 다른 자료와 다름: 20세이하, 21-30세, 31세이상.  
 3) 울산광역시 제외.  
 4) 추정치.

문으로 실시하는 두 조사기관의 DB에 반영된 회원들의 특성 정보이다. MBC와 광고단체연합의 자료는 방문자조사 방식으로 실시된 조사결과에 반영된 응답자 특성으로, MBC 자료는 어떤 TV 프로그램에 대한 온라인 시청자조사의 결과이고(<http://www.mbc.co.kr>), 광고단체연합의 자료는 연합회 소속 47개 기관의 웹사이트에 조사용 배너를 연결하여 실시한 인터넷 이용자 특성에 관한 조사의 결과이다(<http://www.adic.co.kr>). 온라인 서비스의 경우 가입자 DB(C)와 DB(D)는 온라인 서비스 제공사의 현재 가입자 특성 정보이다. 회원 DB와 가입자 DB는 조사의 결과가 아닌 회원 또는 가입자 전체에 대한 정보임에 유의할 필요가 있다.

회원 또는 가입자 DB의 구축시기와 방식, 조사대상의 규정 등이 약간씩 다르기 때문에 엄밀한 비교는 불가능하지만 인구학적 특성에 대한 대략적인 비교는 가능하다. 성별로 보면 남성 사용자가 70~80% 가량을 차지하고 있다. 성비는 인터넷과 온라인 서비스 사이에서만 아니라 내부에서도 약간씩 차이를 보이고 있다. 비슷한 시기에 구축된 두 인터넷 회원 DB간에 상당한 차이를 볼 수 있으며, 방문자조사 방식으로 조사된 MBC 자료와 광고단체연합의 자료도 차이를 보이고 있다. 언론사 웹사이트 경우 전반적으로 여성과 20대의 방문빈도가 다른 사이트에 비하여 높은 점을 감안하더라도(한국광고단체연합, 1999) 두 자료간에는 상당한 차이가 있다.

연령별 분포를 보면 20대와 30대가 주류를 이루고 있는데, 특히 20대가 40~60%를 차지하고 있다. 지역적으로는 서울이 대부분의 30%이상을 차지하고 있다. 분류방식의 일관성 결여로 여기에 제시되지는 않았지만 대부분의 자료에서 전문기술직과 사무직 종사자가 큰 비중을 차지하고 있고 교육수준 별로는 대학졸업이상이나 대학(원) 재학생의 비율이 높게 나타났다.<sup>11)</sup>

11) 인구센서스 자료와 비교해 볼 때 현재의 사이버공간은 남성, 20대의 연령, 서울지역이 과잉대표되고 있다. 한편 표본의 특성분포에서 아무리 많은 유사성을 보이더라도 모집단을 추정하는 값에서는 커다란 차이가 날 수 있다는 점도 고려할 필요가 있다(Robinson, 1989).

온라인 서비스의 가입자 DB 자료는 각 서비스 업체의 전체 가입자를 대상으로 한 것이기 때문에 온라인 서비스 사용자의 특성을 대체로 잘 반영하고 있는 것으로 볼 수 있다. 그러나 두 업체의 가입자 특성이 약간씩 다르게 나타나고 있기 때문에 한 업체의 특성이 전체 온라인 서비스 가입자의 특성이라고 단정짓기는 힘들다. 인터넷 사용자 특성은 모집단 자체가 애매한 만큼 자료별 차이도 크게 나타나고 있다. 현재 인터넷 사용인구는 급속히 증가하는 추세에 있고 그에 따라 사용자들의 특성 분포도 상당히 유동적이다. 사이트 유형에 따라 사용자의 특성도 다르게 나타난다(한국광고단체연합회, 1999). 따라서 온라인 서비스와 마찬가지로 어떤 하나의 자료가 우리나라 인터넷 전체 인구의 특성을 제대로 반영하는 것으로 보기는 힘들다.

요컨대 각 자료는 특정 사이트 또는 서비스를 이용하는 사용자들의 특성을 대표할 수는 있어도 어느 하나가 인터넷 전체 또는 온라인 서비스 전체를 대표한다는 단서는 발견하기 힘들다. 통신망 인구가 급속하게 늘어나고 있기 때문에 시간의 경과에 따라 사용자 인구 전체의 성숙에 따른 특성변화는 나타나겠지만, 현시점에서 각 사이트에 접근하는 그 수가 증가되더라도 그 구성은 크게 달라지지 않을 것이다.

## 2. 온라인 사회조사의 측정

사회조사에서 측정(measurement) 과정에서 체계적 오차를 발생시키는 요인에는 측정도구, 자료의 기록 및 입력과정, 면접원, 비응답 등이 있다.

### 1) 측정도구의 구성 문제

#### (1) 사용자 인터페이스

온라인 조사는 기본적으로 자기기입식 조사이기 때문에 응답자들로 하여금 긴 질문에 답하도록 설득하거나 혼란스러운 부분을 설명해 줄 면접자가 존재하지 않는다. 따라서 질문지의 형식이나 구성 방식이 응답과정에 영향을 줄 수가 있기 때문에 응답자들이 흥미를 가지고 편하게 응답

을 할 수 있도록 질문지를 구성하는 것이 매우 중요하다(Schuman and Presser, 1981; Rea, 1992).

일반적으로 텍스트나 전자설문의 형태보다 웹문서로 작성된 GUI방식의 질문지가 사용자 편의적인 환경을 만들어낼 수 있다(Nesbary, 2000). 특히 웹문서는 이미지나 오디오와 같은 멀티미디어 기능을 활용한 질문 구성이 가능하기 때문에 전통적인 방식으로는 불가능했던 형태의 조사도 시도할 수가 있다.

일반적으로 GUI 방식이 선호되지만 기존의 텍스트 방식에 익숙해져 있는 사용자들이나 GUI 방식을 사용하기 어려운 경우도 고려되어야 한다. 텍스트 방식이 새로운 기술의 유용한 측면들을 제대로 활용하지 못하는 한계는 있지만, 아직도 많은 사용자들에게는 협소한 통신대역(bandwidth)이 현실적인 제약이 되고 있기 때문에 조사대상의 특성을 충분히 고려하여 사용자 인터페이스를 설정할 필요가 있다.

긴 질문지의 경우 주요 내용에 따라 구역으로 나누고 질문지를 구역단위로 제시함으로써 응답의 효율성과 응답자 편의성을 높일 수 있다. 구역화는 질문내용의 논리적 전개에 따라 자연스럽게 응답이 이루어질 수 있도록 하는 데에 일차적인 목적이 있지만, 응답과정에서 지루함을 덜어줄 뿐만 아니라 질문지의 로딩/loading)시간의 단축에 따른 심리적 전송속도의 향상도 기대할 수 있다.

비일관적인 응답이나 무응답에 대한 확인도 구역단위로 수행하여 응답자들이 질문내용을 제대로 기억하고 응답을 수정하거나 재입력할 수 있게 하는 것이 좋다. 나아가서 질문의 순서에 따른 응답효과(response effect)가 예상되는 경우에는 질문 문항을 무작위적으로 배치(randomization)할 필요가 있는데, 이 경우에도 구역화는 효과적인 질문지 구성 전략이 될 수 있다.

## (2) HTML 태그의 선택과 자동값의 설정

질문지 구성에 관련된 대부분의 HTML 태그(tag)은 그 기능이 비교적 분명하지만 <RADIO>와 <SELECT>의 경우에는 경우에 따라서 상호 교환적으로 사용될 수 있다. 일반적으로 선택지의 지문이 간단한 경우에 사



용되는 <SELECT> 태그는 질문을 먼저 드러내고 선택지는 응답할 경우에만 제시되기 때문에, 질문의 취지를 보다 잘 드러낼 수 있고 응답자들이 보다 신중한 선택을 할 수 있게 해 준다. <RADIO> 태그의 경우는 일반적인 질문지에서 사용되는 선택지 구성을 웹에서 그대로 할 수 있게 해주는 태그로서 특히 선택 지문이 긴 경우에 효과적이다. <표 4>는 웹문서 형태의 질문구성에서 흔히 사용되는 HTML 태그를 정리하고 있다.

이러한 태그의 사용과 관련하여 유의해야 할 부분은 자동값(default)의 설정이다. HTML의 <FORM> 관련 태그는 자동값을 설정할 수 있으나 질문지 구성에서 자동값의 설정은 피해야 한다. 자동값이 설정되는 경우 특정한 방향으로 응답을 유도할 수도 있고 차후에 문항별 무응답을 확인하는 것도 불가능하기 때문이다. 자동값을 설정하지 않고 자료의 기록과정에서 무응답이 나타나면 CGI나 Java 스크립트를 통하여 오류 메시지가 나타나도록 하면 조사가 보다 효과적으로 수행될 수 있다.

<표 4> 질문지 구성에 사용되는 HTML 태그\*

용 도	HTML TAGs	비 고
구역의 구분	<FORM> ... </FORM> <INPUT TYPE = SUBMIT> <INPUT TYPE = RESET>	
배타적 선택지 리커트형 척도	<INPUT TYPE = RADIO> <SELECT><OPTION> ... </SELECT>	
중복 선택지	<INPUT TYPE = CHECKBOX>	
문자입력	<INPUT TYPE = TEXT> <TEXTAREA> ... </TEXTAREA>	
분할형 선택지	<INPUT TYPE = SUBMIT>	<value="선택지">
기타	<INPUT TYPE = HIDDEN> <INPUT TYPE = PASSWORD>	고유번호(IDNO) 전달

\* HTML 표준규정 2.0 기준(<http://www.w3.org/MarkUp/html-spec>).

## 2) 자료처리 시스템의 문제

### (1) 시스템의 안정성과 보안성

온라인 사회조사의 경우 조사과정에서 핵심적인 역할을 수행하는 것이 바로 서버 컴퓨터이다. 서버 컴퓨터는 응답자들에게 질문을 제시하고 질문내용을 기록하는 역할을 수행하기 때문에 일반적인 사회조사에서의 면접자, 부호자, 자료입력자의 역할을 동시에 수행한다. 따라서 자료처리 시스템은 측정과 관련된 자료입력 및 면접원 오차와 밀접하게 관련되어 있다.

앞에서도 언급되었듯이 통신망 사용자들은 사용 시간대와 요일에 따라 다양한 특성을 보이기 때문에 온라인 사회조사는 24시간 전시간대를 포괄해야 하며 기간도 적어도 일주일이상 되어야 하는데, 조사의 전기간 동안 서버가 중단없이 운영될 수 있도록 해야 한다. 현재의 기술적인 한계로 통신망의 장애는 드물지 않게 일어나고 있으며 예고없는 정전이나 하드웨어의 오작동도 예상될 수 있는 문제이다. 또한 입력의 집중에 따른 과부하로 인하여 시스템의 장애가 발생할 수도 있다. 철저한 사전점검이 이루어져야 하고 조사가 수행되는 동안에서도 수시로 점검하여 시스템의 안정성이 유지될 수 있도록 해야 한다. 시스템의 이상으로 일정 기간 동안 조사가 진행되지 못하면 그 시간대에 자주 이용하는 사용자의 배제로 선택편기가 발생하게 된다.

나아가서는 시스템의 보안성도 고려해야 한다. 보안성은 호환성과 개방성을 지향하는 통신망의 내재적인 문제로서, 외부자의 시스템에 대한 침입에서부터 응답정보의 유출과 같은 다양한 쟁점과 관련되어 있다. 특히 민감한 사안이나 응답자의 사적인 정보가 포함된 조사에서는 시스템의 보안성이 중요하게 고려되어야 한다.

### (2) 소프트웨어의 충실성

하드웨어의 안정성과 함께 소프트웨어의 충실성도 반드시 고려해야 한다. 앞에서 보았듯이 온라인 조사에서 응답 결과를 처리하는 CGI/DBMS 나 Java 스크립트의 충실성은 자료의 질과 직결되는 중요한 요인이다. 일

반적으로 프로그램의 문법(syntax) 오류는 프로그램 작성과정에서 비교적 쉽게 수정될 수 있지만 보다 문제가 되는 것은 논리적인 (logical) 오류나 자료기록 과정상의 문제이다. 사례나 문항의 불명확한 구분, 조건부 문항 처리의 오류, 복수 응답 문항의 잘못된 처리 등이 흔히 발생하는 프로그램 오류들이다.

온라인 사회조사에서 자료처리를 위한 프로그램은 작성은 전문적 프로그래머가 담당하게 되는데 이 과정에서 연구자와 프로그래머간의 효과적인 커뮤니케이션이 매우 중요하다.<sup>12)</sup> 소프트웨어의 오류는 개발단계부터 여러 차례의 검증을 확인하고 수정을 해야 한다. 사전조사 과정에서도 프로그램의 정상 작동 여부를 반드시 확인해야 한다. 또한 본격적인 조사가 진행되는 과정에서도 예기치 못한 프로그램 오류가 발생할 수 있기 때문에 기록된 자료를 수시로 점검하여 문제를 발견되면 즉시 수정할 수 있도록 해야 한다.

### 3) 비응답오차와 응답오차

사회조사에서 비응답(nonresponse)은 조사대상자가 응답을 하지 않는 경우를 지칭하는데 이는 여러 가지 요인 때문에 나타나며, 비응답오차는 응답에 참여한 사례와 비응답 사례간에 체계적인 차이가 존재하는 경우에 발생한다.

응답자의 자발성에 의존하는 온라인 사회조사에서는 비응답오차가 크게 나타날 수 있다. 특히 전자우편조사나 방문자조사에서 비응답오차가 문제가 될 수 있으며, 전자우편조사에서 비응답은 주로 전자우편 리스트의 결함에서 기인한다. 사용자명이나 메일 서버가 잘못 입력된 것은 전송이 불가능하다. 제대로 입력되었다고 하더라도 해당 사용자가 더 이상 그 메일 서버를 사용하지 않거나 조사 당시에 서버에 일시적인 장애가

12) 프로그래밍에 익숙하지 못한 연구자들이 스스로 프로그래밍을 하는 경우에는 PERL (Practical Extraction and Retrieval Language)이라는 스크립트를 사용하면 편리하다. 특히 PERL은 사용하기가 쉽고 공개된 프로그램으로서 다양한 코드들이 인터넷상에 공개되어 있기 때문에 약간의 수정으로 자신의 목적에 맞는 프로그램을 작성할 수 있다.

발생할 경우에도 비응답이 발생할 수 있다.<sup>13)</sup>

비응답자에 대한 분석은 응답편기를 발견하는 중요한 절차이다. 비응답자에 대한 정보가 부족하면 이들이 자료의 질에 미치는 효과를 분석할 수 없게 된다. 따라서 온라인 사회조사에서는 ISP나 IP 주소, 운영체제, 브라우저 등 통신망 상에서 자동적으로 수집가능한 기술적 특성정보를 수집하여 비응답자 분석에 활용할 필요가 있다.

응답오차(response error)는 응답자가 고의적으로 또는 사실에 대한 불충분한 이해로 왜곡되거나 정확하지 못한 정보를 제공하는 경우에 발생한다. 사이버공간에서는 익명성과 정체차용의 특성 때문에 응답오차가 크게 나타날 수 있다. 또한 일부 간이조사에서 행해지는 조사결과의 노출도 응답오차를 크게 만들 수 있다. 조사가 진행되는 도중에 결과를 노출시키는 관행은 사회조사의 기본적인 원칙을 위배하는 것으로 자료의 질에 치명적인 문제를 가져올 수 있다(Sudman and Bradburn, 1982). 물론 이러한 관행은 본격적인 조사가 아닌 시사적인 쟁점을 흥미위주로 다루는 간이조사에서 주로 행해지고 있고 고객서비스 차원에서 결과가 공개되고 있다. 그러나 이러한 사이트에 접속하는 사용자 수가 많다는 점을 고려할 때 이러한 조사관행은 온라인 조사에 대한 일반인들의 부정적인 인식으로까지 이어질 수 있다.<sup>14)</sup>

온라인 사회조사에서 비응답 및 응답오차는 조사 참여에 대한 인센티브의 제공구조와 밀접한 관련을 가지고 있다. 회원조사에는 일반적으로

- 
- 13) 사용자나 서버의 불명확(unknown)하거나 일시적인 장애를 겪고 있는 경우에는 오류 메시지를 보내오기 때문에 재시도를 할 수 있다(사실 서버 장애의 경우 인터넷 메일은 수차례의 자동 재전송을 시도한다). 가장 문제가 되는 것은 사실상 “죽은” 계정, 즉 주소나 서버에는 이상이 없지만 해당 사용자가 더 이상 그 계정을 사용하지 않는 경우이다. 이 경우 메일은 정상적으로 수신자 서버로 전달되지만 읽혀지지 않은 채 서버에만 남아있게 된다.
- 14) 방법론적인 원칙에 대한 고려가 없이 비전문가들에 의해 조잡하게 구성된 질문이 자주 등장하기 때문에 강화될 수도 있다. 경우에 따라서는 단순히 조사에 참여한 응답자 수를 공개하는 것도 응답효과를 초래할 수 있다(Smith and Kollock, 1999 참조).

응답에 참여하는 모든 회원들에게 인센티브를 제공하고 있고, 전자우편 조사나 방문자조사에서는 전체 참여자 중 일부를 추첨하여 답례를 제공하는 방식이 흔히 사용된다. 조사내용이 까다로운 문항을 포함하고 있거나 시간이 많이 소요되는 경우에는 인센티브 제공 방식도 달라져야 한다. 그러나 일반적인 조사에서와 마찬가지로 인센티브가 지나치게 강조되면 응답편기가 나타날 수 있다.<sup>15)</sup>

### 3. 온라인 사회조사의 윤리적 쟁점

사회조사에서 응답자와 관련된 윤리적인 문제는 주로 프라이버시의 보호, 자발적 동의, 그리고 비밀보장이라는 세 가지 원칙을 중심으로 논의된다(Sieber, 1992). 온라인 사회조사의 윤리적 쟁점은 프라이버시 보호와 통신예절의 존중을 중심으로 논의될 수 있다.

#### 1) 응답자 프라이버시 보호

온라인 조사에서는 전자우편주소나 회원정보 등과 같은 개인의 신분 확인이 가능한 정보가 기록되기 때문에 프라이버시 보호와 비밀보장이 문제가 될 수 있다(Smith and Kollock, 1998). 온라인 사회조사에서는 면접성의 결여로 인하여 일반적인 조사에서보다 응답자들의 상대적인 익명성이 더 많이 확보되는 측면이 있지만, 응답내용이 응답자의 정체와 쉽게 연결될 수 있는 온라인 조사에서는 비밀의 보장은 중요한 과제이다. 특히 회원조사와 전자설문조사의 경우 응답자의 실제적 정체성과 연결되어 있기 때문에 응답내용에 대한 비밀유지가 절대적으로 중요하다. 전자우편조사에서도 주소 자체가 개인의 정체성에 대한 간접적인 단서가 되기 때문에 동일한 문제가 제기된다.

조사 실행과정에서 수집된 응답자의 개인정보가 조사이외의 목적으로

15) 참고로 필자가 통신망 이용자들을 대상으로 수행한 초점집단면접에서 경품을 받기 위해 가족의 정체를 차용하여 응모하는 경우가 빈번하다는 것이 확인되었다.

전용되는 경우에는 프라이버시에 대한 심각한 침해가 야기될 수 있다. 회원조사나 전자설문조사에서 사용되는 회원DB나 가입자 DB의 정보는 쉽게 다른 목적으로 사용될 수가 있다. 방문자조사나 전자우편조사에서 수집되는 전자우편주소와 같은 개인 정보도 다른 용도로 전용될 수가 있다. 특히 사이버공간에서 주요한 피해의 하나가 되고 있는 광고우편(UCE)나 스팸(spamming) 등에 이러한 개인 정보가 이용될 소지가 크다.

전자우편조사의 경우 질문지 발송과정에서 뿐만 아니라 전자우편 명부의 확보과정에서도 프라이버시 침해가 발생할 수 있다. 전자우편 리스트를 이용한 대량 우편의 발송은 수신자의 동의가 없이 이루어진다는 점에서 이용자의 프라이버시를 침해할 수 있다. 또한 정당한 절차를 거치지 않고 전자우편 주소를 확보하는 과정에서도 프라이버시가 발생할 수 있다. 응답자의 정체성을 확인하지 않는 단순한 방문자조사의 경우에도 조사의 필요에 따라 쿠키(cookie)<sup>16)</sup>의 형태로 전달되는 기본적인 정보나, 응답자 사례의 필요성 때문에 수집하는 개인정보도 프라이버시 침해로 이어질 수 있다.

어떠한 종류이건 간에 조사의 목적으로 수집된 정보는 다른 목적으로 사용되어서는 안된다. 결국 응답자들의 프라이버시에 대한 보호는 연구자나 조사수행 주체의 의지에 상당히 좌우되는 문제라고 할 수 있다.

## 2) 통신예절의 존중

온라인 조사는 일반적으로 응답자들의 자발적인 참여를 전제로 한다. 일반적인 조사에서와 마찬가지로 조사의 목적, 후원기관 및 수행주체, 응답자 선정, 결과의 활용 등에 대한 친절한 안내가 제공되어야 할뿐만 아니라, 온라인 조사는 사이버공간의 문화를 존중하는 방식으로 수행되어야 한다. 특히 면접성의 결여와 익명성으로 특징지어지는 사이버공간에

16) 쿠키정보는 웹브라우저를 이용하여 특정한 사이트에 접속할 때 자동적으로 저장되는 사용자의 컴퓨터에 관한 기본 정보이다. 쿠키는 온라인 조사에서 응답자의 확인, 구역간의 이동, 응답유형의 분석 등에 필요한 중요한 정보를 담고 있다.

서도 사용자들에게 요구되는 예절, 즉 네티켓(netiquette)이 존재하며(Jones, 1994; 조동기, 1997; Smith and Kollock, 1998), 온라인 조사는 각 상황에서 요구되는 적절한 통신예절을 준수하여야 한다. 응답자의 인격이나 프라이버시의 보호라는 윤리적인 차원에서뿐만 아니라 응답자들과의 성공적인 래포의 형성을 통한 응답의 신뢰성 확보를 위해서도 통신예절은 존중되어야 한다.

특히 전자우편조사의 경우 조사의 취지나 목적에 대한 충분한 설명이 없이 이루어지는 질문지의 발송은 그 자체가 스팸밍으로 간주될 수 있다. 따라서 질문지를 발송하기전에 조사 대상자들의 동의를 받는 것이 중요하다(CASRO, 1999). 조사의 취지나 목적을 설명하고 전자우편 주소 명부를 만든 절차와 그 출처를 밝혀줌으로써 더 높은 응답률을 기대할 수 있다.

요컨대 온라인 사회조사에서는 응답자의 프라이버시를 보호하고 사이버공간의 문화를 존중하는 연구자의 책임이 강조된다. 연구자는 조사의 목적이나 결과활용에 대한 일반적인 안내와 더불어, 응답내용에 대한 비밀유지, 개인정보 보호, 사이버공간의 예절준수 등에 대한 적극적인 의지를 조사대상자들에게 보여주어야 한다.

## V. 맺는 말

정보통신망을 매개로 하는 온라인 사회조사는 최근 들어 급속하게 확장되고 있는 사이버공간에 대한 경험적인 자료를 수집하는 새로운 방법으로 그 활용이 늘어나고 있다. 그런데 조사의 실질적 필요성보다 정보기술의 수단적 편의성이 강조되면 조사의 오남용으로 이어질 수 있다. 이 글은 온라인 사회조사의 발전가능성을 탐색하기 위하여 그 유형과 특성을 검토하고 체계적 오차와 관련하여 제기되는 방법론적인 쟁점을 분석하였다.

온라인 조사는 회원조사, 전자설문조사, 방문자조사, 전자우편조사로 구분될 수 있으며, 현재로서는 인터넷상의 회원조사와 온라인 서비스상

의 전자설문조사가 상대적으로 효과적인 방법이라 할 수 있다. 일반적인 조사연구 방법에 비하여 온라인 사회조사는 비용이 적게 들고 그 실행도 비교적 간편한 편이다. 또한 정보기술을 적절히 활용하면 상호작용성이 높고 사용자 편의적인 질문지를 구성할 수가 있다. 그러나 표집틀과 관련된 자료의 질은 상대적으로 떨어지는 편이다.

이 글에서는 온라인 사회조사의 방법론적인 쟁점을 표본설계, 측정, 윤리적 쟁점으로 정리해 보았다. 표본설계와 관련하여 모집단의 규정, 표집틀의 확보, 추출방법 등이 중요한 쟁점으로 논의되었다. 비정형적이고 급속한 팽창의 상태에 있는 통신망 사용자 인구를 대상으로 하는 온라인 사회조사에서는 모집단에 대한 규정 자체도 중요한 고려 사항이 되어야 한다.

대표성 있는 표집은 적절한 표집틀의 확보를 전제로 한다는 점에서, 편의표집의 경향이 강한 현재의 온라인 조사에서는 표집틀 편기가 크게 나타날 수 있다. 인터넷에서는 포괄성, 효율성, 추출확률의 계산가능성의 측면에서 편기가 적은 표집틀의 확보가 시급하고, 온라인 서비스에서는 기존의 표집틀을 보다 효과적으로 활용할 수 있는 방안이 모색되어야 한다. 표집틀의 문제와 함께 부적절한 추출방법은 다량 사용자의 과잉대표라는 문제를 가져온다. 다양한 사이트를 포괄할 수 있는 표집설계가 요구되며 측정의 다원화에 따른 응답자의 중복과 적격성을 확인할 수 있는 장치도 마련되어야 한다. 표집틀의 가용여부와 표집의 방법에 따라 응답률은 다르게 규정될 수 있으며 응답률의 계산이 불가능한 경우에는 사후 표집율이나 응답밀도 등과 같은 개념도 유용할 수 있다.

온라인 사회조사는 자기기입식이기 때문에 사용자 편의적인 질문지 구성이 요구된다. 웹문서는 멀티미디어의 활용이 가능하고 상호작용적 질문지 구성이 가능하기 때문에 많은 이점을 가지고 있다. 그러나 웹문서 방식을 수용하지 못하거나 텍스트 방식을 선호하는 사용자층이 상당한 부분을 차지하기 때문에 질문지 인터페이스의 선택은 조사대상의 특성을 고려해야 된다. 질문지 구성은 또한 구역화를 효과적으로 활용할 필요가 있으며, HTML 택의 형태와 같은 상이한 기술적 처리가 가져올 수 있는 응답효과에 대한 연구도 요구된다.



온라인 조사의 과정은 컴퓨터 시스템에 크게 의존하기 때문에 시스템의 안정성과 보안성, 소프트웨어의 충실성도 중요한 고려사항이 되어야 한다. 조사기간 동안 시스템이 안정적으로 운영되기 위해서는 전력 공급 중단이나 과부하의 집중 등에 대한 대비가 필요하다. 자료 처리에 필요한 소프트웨어의 충실성은 단순한 문법 오류보다 논리적 오류가 더 치명적이기 때문에 프로그래머와 연구자간에 효과적인 커뮤니케이션이 이루어져야 하고 다양한 과정을 통해 오류가 확인되어야 한다.

응답자의 자발성에 의존하는 온라인 사회조사에서는 익명성과 정체차용에 따른 비응답오차 및 응답오차가 크게 나타날 수 있기 때문에 이를 줄일 수 있는 방안이 모색되어야 한다. 온라인 조사에서 비응답은 인센티브의 제공구조와 밀접한 관련을 가지고 있는데, 조사의 내용과 분량에 따라 적절한 형태의 인센티브가 효과적으로 제공되어야 하겠지만 인센티브에 대한 지나친 강조는 오히려 응답편기를 크게 할 수 있다. 조사결과를 사전에 노출시키는 조사관행은 응답오차를 유발시키고 조사에 대한 부정적인 인식을 확산시킬 수 있다.

온라인 조사에서는 응답자에 대한 전자적 흔적을 여러 가지 형태로 남기기 때문에 응답내용에 대한 비밀의 유지와 응답자의 프라이버시 보호가 민감한 쟁점으로 등장한다. 조사의 목적으로 수집된 개인 정보는 다른 용도로 전용되어서는 안되며 사이버공간에서 요구되는 네티켓도 준수될 필요가 있다. 이를 위해서는 응답자의 프라이버시를 보호하고 사이버공간의 통신예절을 존중하려는 연구자의 의식과 책임이 강조된다.

정보기술을 매개로 하는 온라인 사회조사는 많은 가능성을 보여주고 있고 앞으로 사이버공간이 확장됨에 따라 그 활용이 더욱 늘어날 것이 분명하다. 표본설계, 측정, 및 윤리적 쟁점에 대한 엄밀한 방법론적인 성찰을 통해 온라인 조사는 면접조사나 전화조사와 같은 본격적인 조사연구의 한 방법으로 자리잡게 될 것이다.

## 참고문헌

- 골드뱅크. 1997. "골드뱅크 1차 설문조사: 인터넷 이용 현황조사." (주)골드뱅크 (Goldbank) 커뮤니케이션즈 미간행 자료.
- 남춘호. 1998. "사회학적 연구에서 표본선택편의." 《한국사회학》 제32집 봄호.
- 아이클릭. 1999. "인터넷 이용자 실태조사" (제1차). (주)아이클릭(Eyeclick) 조사보고서.
- 인터넷 미트릭스. 1999. "인터넷 센서스" (제2차). (주)인터넷 미트릭스 (Internet Metrix) 조사보고서.
- 조동기. 1996. "컴퓨터 통신망을 이용한 사회조사의 가능성과 한계." 서울대 사회발전연구소 Working Paper 시리즈 #2.
- \_\_\_\_\_. 1998. "인터넷을 통한 사회조사의 특성과 쟁점". 인터넷전략연구소 세미나 발표논문.
- 통계청. 1995. 《1995 인구주택총조사》(CD-ROM). 통계청.
- \_\_\_\_\_. 1997. 《1997 상반기 사회통계조사보고서: 정보와 통신, 안전부문》. 통계청.
- 한국인터넷정보센터. 1999. 《인터넷 이용현황 통계보고서》(12월). 한국전산원.
- 한국정보문화센터. 1999. 《국민생활 정보화실태 및 정보화인식 조사》. 한국정보문화센터.
- 한국광고단체연합회. 1999. 《KNP 보고서》(1999년 상반기). 한국광고단체연합회.
- 홍두승. 1992. 《사회조사분석》(제2판). 다산출판사.
- Berk, Richard A. 1983. "An Introduction to Sampling Selection Bias in Sociological Data." *American Sociological Review* 48: 386-398.
- Council of American Survey Research Organization(CASRO). 1999. Frequently Asked Questions about Online Research: New Methodologies for Traditional Techniques. <http://www.casro.org>.
- Czaja, Ronald, and Johnny Blair. 1996. *Designing Surveys: A Guide to*

- Decisions and Procedures*. Thousand Oaks, CA: Pine Forge Press.
- December, John. 1996. "Units of Analysis for Internet Communication." *Journal of Communication* 46(1), Winter: 14-38.
- Fowler, Floyd J., Jr. 1993. *Survey Research Methods: Applied Social Research Methods Series* Vol. 1(2nd Edition). Newbury Park, CA: Sage Publications.
- Jones, Steve (ed). 1994. *CyberSociety: Computer-Mediated Communication and Community*. Thousand Oaks, CA: Sage.
- \_\_\_\_\_. 1998. *CyberSociety 2.0: Revisiting Computer-Mediated Communication and Community*. Thousand Oaks, CA: Sage.
- \_\_\_\_\_. 1999. *Doing Internet Research: Critical Issues and Methods for Examining the Net*. Thousand Oaks, CA: Sage.
- Kitchin, Rob. 1998. *Cyberspace: The World in the Wires*. New York, NY: John Wiley and Sons.
- Krol, Ed., and Paula Ferguson. 1995. *The Whole Internet: Users Guide and Catalog*. Sebastopol, CA: O'Reilly and Associates.
- Lee, Raymond M.(ed). 1995. *Information Technology for the Social Scientist*. London: UCL Press.
- Nesbary, Dale K. 2000. *Survey Research and the World Wide Web*. Boston, MA: Allyn and Bacon.
- Rea, Louis M., and Richard A. Parker. 1992. *Designing and Conducting Survey Research: A Comprehensive Guide*. San Francisco, CA: Jossey-Bass Publishers.
- Rheingold, Howard. 1993. *The Virtual Community: Homesteading on the Electronic Frontier*. New York, NY: Addison-Wesley.
- Robinson, J. P. 1989. Poll Review: Survey Organization Differences in Estimating Public Participation in the Arts. *Public Opinion Quarterly* 53 (3): 397-414.
- Rossi, Peter H., James D. Wright, and Andy B. Anderson(eds). 1983. *Handbook of Survey Research*. New York, NY: Academic Press.

- Roth, Marty. 1999. Taking Market Research Online. *OnTheInternet*, September: 28-33.
- Schuman, H. and S. Presser. 1981. *Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording, and Context*. Orlando, FL: Academic Press.
- Sieber, J. 1992. *Planning Ethically Responsible Research: Developing an Effective Protocol*. Newbury Park, CA: Sage.
- Smith, Marc, and Peter Kollock (ed). 1998. *Communities in Cyberspace*. New York, NY: Routledge.
- Sudman, Seymour, and Norman M. Bradburn. 1982. *Asking Questions: A Practical Guide to Questionnaire Design*. San Francisco, CA: Jossey-Bass Publishers.
- Swoboda, Walter J., N. Muhlberger, R. Weitkunat, and S. Schneeweiß . 1997. Internet Surveys by Direct Mailing: An Innovative Way of Collecting Data. *Social Science Computer Review* Vol. 15 No. 3: 242-255.