

Khó khăn và cách khắc phục trong chọn mẫu ngẫu nhiên

TS. LƯU HỒNG MINH

Học viện Báo chí và Tuyên truyền; Email: minhpbctt@gmail.com

Nhận ngày 25 tháng 4 năm 2024; chấp nhận đăng tháng 7 năm 2024.

Tóm tắt: Các nghiên cứu xã hội học định lượng thường phải tiến hành chọn mẫu ngẫu nhiên, trong đó đòi hỏi quy trình chọn mẫu phải tuân thủ theo rất nhiều quy định chặt chẽ, có cơ sở khoa học, đặc biệt là toán xác suất thống kê. Thực tế xã hội đang thay đổi nhanh chóng, đa dạng, phong phú luôn đòi hỏi các nhà nghiên cứu khoa học phải tìm tòi, cập nhật nắm bắt được cách xử lý các tình huống mới trong quyết định thiết kế cách chọn mẫu ngẫu nhiên, nhưng vẫn phải dựa trên những kiến thức cơ bản của toán xác suất thống kê từ cách đây hàng trăm năm. Bài viết này nhằm làm rõ những khó khăn và đưa ra cách khắc phục trong quá trình thiết kế quy trình chọn mẫu ngẫu nhiên thường được áp dụng cho nghiên cứu xã hội học định lượng.

Từ khóa: chọn mẫu ngẫu nhiên, đơn vị chọn mẫu, khung chọn mẫu, các phương pháp chọn mẫu ngẫu nhiên, quy trình chọn mẫu.

Abstract: Quantitative sociological research often requires random sampling, which demands that the sampling process adhere to many strict, scientifically-based regulations, particularly statistical probability. The rapidly changing, diverse, and rich social reality always requires researchers to explore, update, and understand how to handle new situations in deciding on random sampling designs, while still relying on the basic knowledge of statistical probability from hundreds of years ago. This article aims to clarify the challenges and propose solutions to overcome difficulties in the random sampling design process commonly applied in quantitative sociological research.

Keywords: random sampling, sampling unit, sampling frame, random sampling methods, sampling process.

Mục đích các nghiên cứu xã hội học qua một số mẫu đã lựa chọn (nghiên cứu một bộ phận của tổng thể điều tra) song lại mong muốn có khả năng suy rộng cho tổng thể điều tra, phản ánh phù hợp những đặc trưng và cơ cấu của tổng thể điều tra. Tuy nhiên, do các khách thể của nghiên cứu xã hội học là con người, họ chỉ là một nhóm nào đó trong tổng thể xã hội và họ luôn di động trong quá trình sinh sống. Chính vì vậy, để so sánh, kiểm chứng kết quả nghiên cứu từ mẫu so với tổng thể điều tra thường là điều bất khả thi.

Nghiên cứu chọn mẫu khác với nghiên cứu tổng thể vì nó chỉ lấy một phần các đơn vị của tổng thể để nghiên cứu. Vậy những đơn vị không được nghiên cứu có giống các đơn vị đã được nghiên cứu hay không.

Đối với nghiên cứu định tính thường chọn mẫu điển hình (có thể có tên gọi là nghiên cứu theo kinh nghiệm và xét đoán/ phi xác suất) nghiên cứu này nhiều khi mang tính đại diện rất cao nhưng kết quả nghiên cứu không biết có đúng là đại diện cho tổng thể nghiên cứu hay không, vì nó bị ảnh hưởng rất nhiều vào sự lựa chọn chủ quan của nhà nghiên cứu.

Đối với nghiên cứu định lượng thường chọn mẫu ngẫu nhiên (có thể có tên gọi là chọn mẫu theo xác suất) kết quả nghiên cứu đại diện cho tổng thể nghiên cứu, nhưng khi suy rộng cho tổng thể thường có những sai số nhất định mà nhà nghiên cứu cần phải tính toán để biết được độ chính xác của kết quả nghiên cứu. Chính vì vậy, nhà nghiên cứu cần phải có sự hiểu biết nhất định về toán xác suất thống kê, hiểu được các phương pháp chọn mẫu ngẫu nhiên để lựa chọn thiết kế quy trình chọn mẫu phù hợp mục đích nghiên cứu và tình hình thực địa, đồng thời phải tính toán được cỡ mẫu và độ chính xác kết quả nghiên cứu thu được.

1. Một số vấn đề khó khăn thường gặp trong chọn mẫu ngẫu nhiên hiện nay.

1.1. Cần có kiến thức cơ bản về toán xác suất thống kê

Trong chọn mẫu ngẫu nhiên nhà nghiên cứu cần phải có những kiến thức cơ bản về toán xác suất thống kê, nếu không có thể dẫn đến nhiều sai lầm trong quá trình thực hiện chọn mẫu và suy rộng kết quả nghiên cứu mà nhà nghiên cứu không nhận biết được. Việc lựa

chọn quy trình lấy mẫu phụ thuộc vào rất nhiều yếu tố, không có nguyên tắc giản đơn nào hướng dẫn người nghiên cứu lựa chọn theo một sơ đồ mẫu cụ thể. Để có thể quyết định lựa chọn mẫu như thế nào, trước hết ta thấy nó phụ thuộc vào nội dung của vấn đề nghiên cứu (quyết định đối tượng và khách thể nghiên cứu). Từ nội dung này có thể có một số sơ đồ lấy mẫu phù hợp, người nghiên cứu phải tự cân nhắc quyết định mình lựa chọn theo sơ đồ nào, thông thường ở đây dựa vào kinh nghiệm các cuộc nghiên cứu và các ưu nhược điểm của từng phương pháp chọn mẫu ngẫu nhiên cơ bản. Thực tế cho thấy mỗi cuộc nghiên cứu không chỉ giới hạn xem xét một vấn đề mà thường rất nhiều vấn đề, hơn nữa tuy có thể là một vấn đề cần nghiên cứu nhưng thực tế nội dung đó lại được chia ra làm nhiều vấn đề nhỏ hơn theo các biến số hay theo các khung lý thuyết... Mỗi vấn đề này có thể cần một sơ đồ lấy mẫu và cỡ mẫu tương ứng, do vậy người nghiên cứu cần có sự thỏa hiệp và quyết định sự chọn mẫu của mình. Thông thường chỉ dựa vào một số vấn đề chủ yếu có tính quyết định tới cuộc nghiên cứu mà thôi (biến số phụ thuộc). Đối với vấn đề nào tần suất xuất hiện nhiều thì có thể chọn mẫu kích thước nhỏ ngược lại vấn đề nào tần suất xuất hiện ít thì phải chọn mẫu kích thước lớn. Nếu đặc trưng của vấn đề nghiên cứu phân tán tại một số khu vực nhưng lại tập trung tại một khu vực khác khi đó ta nên chọn mẫu phân tổ. Còn lại phần lớn các trường hợp khác thường là ta tiến hành chọn mẫu chùm hay chọn mẫu ngẫu nhiên có hệ thống để các đơn vị được chọn mẫu mang tính đại diện, nhưng cần chú ý tới danh sách chọn mẫu mà ta có thể tiếp cận được, về đặc điểm của danh sách ta sử dụng (có thể có sẵn hay phải làm lại) buộc ta phải chọn mẫu phân tầng hay chùm hoặc danh sách đó ta có theo vần A; B; C... theo khu vực địa lý; theo số lượng các đơn vị chọn mẫu... để sao cho đảm bảo tính ngẫu nhiên không bị sai số hệ thống. Như vậy, muốn đưa ra được quy trình chọn mẫu cụ thể người nghiên cứu ngoài kiến thức về toán xác suất thống kê họ còn phải có kiến thức hiểu biết về tổng thể điều tra và thông tin ban đầu có liên quan (ví dụ thông tin về tổng dân cư và ước lượng số khách thể trong tổng dân cư phân bố theo từng khu vực nghiên cứu).

Trong phần lớn các trường hợp nghiên cứu định lượng phải tiến hành thu thập thông tin nhanh, trên một phạm vi rộng với nguồn kinh phí và nhân lực hạn chế. Điều đó bắt buộc nhà nghiên cứu phải lựa chọn thiết kế một quy trình chọn mẫu cụ thể sao cho kết quả nghiên cứu thu được với một độ chính xác nhất (được đo bằng

sai số chấp nhận trong một khoảng có thể biết trước) khi suy rộng đối với tổng thể điều tra nhưng có nguồn chi phí nhỏ nhất

1.2. Thách thức khi nghiên cứu với tổng thể điều tra lớn và phức tạp

Phần lớn các nghiên cứu định lượng nhà nghiên cứu mong muốn hướng đến một khối dân cư mục tiêu trên phạm vi rộng (phạm vi từ quận/huyện, tỉnh/thành phố trở lên), tuy nhiên, không phải mọi người dân ở đó đều là khách thể nghiên cứu. Họ thường ở một độ tuổi nào đó (để đảm bảo có thể trả lời nhiều câu hỏi được đặt ra), họ cần có những đặc điểm nào đó liên quan đến nội dung nghiên cứu. Ví dụ công chúng truyền hình liên quan đến những người có xem truyền hình, những người yêu thích hay tham gia các hoạt động liên quan đến âm nhạc, thể thao... Muốn chọn mẫu ngẫu nhiên bắt buộc các nhà nghiên cứu phải có danh sách những người cần nghiên cứu này, nhưng không ai có thể cung cấp danh sách cho nhà nghiên cứu trong thời gian ngắn với chi phí hạn hẹp.

Tính phức tạp ảnh hưởng đến chọn mẫu ngẫu nhiên còn thể hiện sự phân bố dân cư không đồng đều giữa các vùng khác nhau: có những vùng tập trung đông dân cư, mật độ dân cư cao như chung cư cao tầng, các đô thị lớn, □

Tính phức tạp còn thể hiện người dân thường xuyên ở và làm việc tại nhiều khu vực khác nhau có thể dẫn đến tại nơi này đã được chọn vào mẫu nghiên cứu sang nơi khác họ lại được chọn vào mẫu nghiên cứu.

Tính phức tạp có thể do khoa học công nghệ và kinh tế thị trường tác động đến đặc điểm dân cư: Nhiều ngành nghề mới xuất hiện và tương ứng cũng có nhiều ngành nghề khác mất đi từ đó dẫn tới nhiều người giàu lên nhanh chóng và nhiều người mất việc làm trở nên nghèo khó, lối sống gắn theo các nhóm này cũng bị thay đổi, hình thức sở hữu tư liệu sản xuất cũng bị thay đổi (sở hữu ô tô cá nhân vừa dùng cho gia đình nhưng cũng được dùng để kinh doanh).

Đặc biệt, hiện nay có nhiều điều tra nghiên cứu xã hội học thường chọn mẫu qua mạng xã hội, qua Internet khi đó mỗi người dân có thể tự đặt tên cho mình theo các nhóm/ hội tham gia hay theo các loại mạng xã hội, theo số điện thoại, theo tài khoản đăng ký... chính điều này có thể dẫn đến mỗi một người trong xã hội thực nhưng lại có rất nhiều tên khác trong xã hội [ảo]. Điều này có thể dẫn đến nguyên tắc chọn ngẫu nhiên và xác suất được lựa chọn không đảm bảo: có người được lựa chọn vào mẫu nhiều lần nhưng có người không được

lựa chọn, thậm trí không có cơ hội được lựa chọn (không có tên trên mạng xã hội).

Tính phức tạp còn có thể do các quy định trong xã hội hiện đại thông qua các thiết chế xã hội luôn có sự thay đổi để phù hợp với sự phát triển của xã hội. Các mâu thuẫn giữa các giá trị, chuẩn mực cũ, truyền thống với các giá trị, chuẩn mực mới, hiện đại, chúng đang tác động mạnh mẽ đến các hành vi và ứng xử giữa các cá nhân hay nhóm này với các cá nhân hay nhóm xã hội khác. Tính đa dạng của các loại nhóm xã hội này cũng làm cho vấn đề chọn mẫu ngẫu nhiên cần đảm bảo các đặc tính khác nhau, đa dạng trở lên rất khó khăn trong chọn mẫu với tổng thể điều tra lớn.

1.3. Khó khăn trong việc đảm bảo tính đại diện và tính khách quan chọn mẫu ngẫu nhiên thực sự

Ngay từ cuối thế kỷ XIX đầu thế kỷ XX nhiều nhà nghiên cứu khoa học đã quan tâm đến chọn mẫu ngẫu nhiên trong nghiên cứu định lượng. Hai tiêu chuẩn gắn với nghiên cứu định lượng là □ xác định rõ yếu tố nghiên cứu □ và □ khả năng đo lường được □ đây là những yếu tố đầu tiên nhà nghiên cứu cần xác định trước khi chọn mẫu⁽¹⁾. Trong nghiên cứu định lượng nhà nghiên cứu thường phải đo lường các vấn đề xã hội thông qua các chỉ báo với các thang đo cụ thể gắn liền với các biến số nghiên cứu, từ đó xây dựng bảng hỏi với nhiều câu hỏi có sẵn các phương án trả lời. Để xác định được các chỉ báo và thang đo đối với các nghiên cứu về nhiều vấn đề xã hội là khó khăn với rất nhiều nhà nghiên cứu khi mới bước vào nghiên cứu xã hội học. Họ phải có một nền tảng những kiến thức cần thiết về vấn đề nghiên cứu để từ đó xây dựng được một khung phân tích thể hiện mối quan hệ giữa nhiều biến số nghiên cứu dựa trên cơ sở lý luận và thực tiễn. Đối với khả năng đo lường được trong nghiên cứu đòi hỏi nhà nghiên cứu cần lập luận giải thích cơ sở để đo lường những khái niệm trừu tượng, đa nghĩa thành khái niệm thực nghiệm đảm bảo có thể xác định được các chỉ báo, biến số, thang đo. Dựa trên nguyên tắc đó các nội dung nghiên cứu của xã hội học được xác định rõ và có mối liên hệ chặt chẽ phản ánh được bản chất của hiện tượng xã hội được nghiên cứu. Cũng dựa trên nguyên tắc này trong nghiên cứu định lượng nhà nghiên cứu cần phải xác định rõ khách thể nghiên cứu: họ là ai, họ ở đâu, làm sao có thể tiếp cận được với họ.

Hai tiêu chuẩn tiếp theo là khi chọn mẫu nghiên cứu định lượng cần đảm bảo tính □ đại diện □ và □ khách quan □, mẫu nghiên cứu đảm bảo tính đại diện có nghĩa là các thông tin về đặc điểm của mẫu nghiên cứu phản

ánh giống như tổng thể nghiên cứu. Đây là tiêu chuẩn rất khó đảm bảo khi nghiên cứu với tổng thể điều tra lớn trên một phạm vi rộng, ví dụ nghiên cứu trên phạm vi cả nước, mẫu cần có người ở miền Bắc, miền Trung, miền Nam; cần có người ở nông thôn, đô thị; cần có người ở miền núi, trung du, đồng bằng, ngoài đảo; các dân tộc khác nhau; ngoài ra còn giới tính, độ tuổi và nhiều đặc điểm liên quan đến nội dung nghiên cứu. Đối với tiêu chuẩn khách quan có nghĩa là khi chọn mẫu cần đảm bảo cơ hội cho mọi thành viên (đơn vị nghiên cứu) có cơ hội được lựa chọn vào mẫu nghiên cứu. Trong đó cơ hội được hiểu là có khả năng được lựa chọn vào mẫu có thể nhiều hay ít nhưng nếu cách chọn của nhà nghiên cứu mà dẫn đến có một số đơn vị nghiên cứu của tổng thể điều tra không có cơ hội nào được lựa chọn là không đảm bảo tiêu chuẩn này. Hai tiêu chuẩn □ đại diện □ và □ khách quan □ thường dẫn đến khó xử lý trong chọn mẫu vì nếu tập trung vào chọn mẫu đại diện theo một số chỉ báo nêu ra sẽ rất khó đảm bảo khách quan trên phạm vi rộng, đặc biệt đối với các thành viên có đặc điểm khác hay các đặc điểm được nghiên cứu nhiều. Để đảm bảo hai tiêu chuẩn này trong nghiên cứu định lượng cần ưu tiên xác định tính khách quan trước để đưa ra các phương án chọn mẫu, sau đó tùy theo mỗi phương án và thử nghiệm thực hiện để xác định các mẫu nghiên cứu có đầy đủ các trường hợp đại diện cho tổng thể cần nghiên cứu hay không. Nếu trong mẫu nghiên cứu không có hay quá ít các trường hợp cần nghiên cứu thì mẫu có thể không đảm bảo tính đại diện và có thể loại bỏ nghiên cứu này. Nhưng ngược lại, do chọn đảm bảo tính khách quan có thể dẫn đến các kết quả tổng hợp của mẫu có thể sai khác so với tổng thể điều tra (dựa trên một số chỉ báo về đặc điểm nhân khẩu học của tổng thể điều tra có được từ kết quả điều tra dân số của Tổng cục Thống kê) khi đó nhà nghiên cứu phải kiểm tra lại toàn bộ quá trình chọn mẫu để xác định có nên gia trọng lại số liệu hay hủy bỏ kết quả nghiên cứu.

1.4. Khó khăn trong tiếp cận người cung cấp thông tin

Trong xã hội hiện đại quyền con người ngày càng được tôn trọng, đặc biệt họ có quyền từ chối tham gia nghiên cứu hay từ chối trả lời một số câu hỏi do nhà nghiên cứu xã hội học đặt ra. Chính điều này đang dẫn đến tỷ lệ không muốn tham gia nghiên cứu ngày càng tăng. Khi mới bắt đầu có các nghiên cứu xã hội học ở Việt Nam cách đây trên 30 năm tỷ lệ người từ chối không tham gia nghiên cứu thường không quá 5% nhưng gần đây tỷ lệ tăng cao khá nhanh có khoảng 10 -

20% người từ chối đặc biệt tại các đô thị lớn như Hà Nội, Thành phố Hồ Chí Minh số lượng từ chối có thể lên 30 - 40%. Tỷ lệ này càng gia tăng khi các chủ đề nghiên cứu có liên quan đến chính trị hay thông tin người trả lời không muốn cung cấp như thu nhập, tài sản□

Khả năng tiếp cận với người được nghiên cứu cũng là vấn đề ngày càng gây khó khăn cho nhà nghiên cứu xã hội học. Ngoài khó khăn về khách thể nghiên cứu của xã hội học thường là những người khó nhận biết được họ trong xã hội như: người hành nghề mại dâm, nghiện ma túy/ chất gây nghiện, đồng tính, trốn lậu thuế□ Do sự phát triển của khoa học kỹ thuật, Internet xuất hiện nhiều người thường xuyên di động trong xã hội như không có việc làm ổn định (nhảy việc), không có nơi ở ổn định, làm nghề điều khiển các phương tiện giao thông□ gây khó khăn cho nhà xã hội tìm cách tiếp cận người cần nghiên cứu. Khi cách chọn mẫu ngẫu nhiên xuất hiện nhiều người không tiếp cận được có thể dẫn đến kết quả khi suy rộng cho tổng thể bị sai lệch nhiều không kiểm soát được. Thông thường trong quá trình thu thập thông tin người tổ chức nghiên cứu thường tăng số lượng mẫu dự bị để bù đắp cho số lượng mẫu không tiếp cận được, trong trường hợp này họ đã giả định người không tiếp cận được giống những người đã tiếp cận được. Thực tế có thể những người không tiếp cận được có thể có nhiều đặc điểm rất khác với những người có tiếp cận được (vì một số lý do nêu trên). Vì vậy tại một nước châu Âu các cơ quan nghiên cứu xã hội học thường phải nghiên cứu bằng được những người này hoặc làm rõ đặc điểm nhân khẩu học hay thông tin liên quan khác thông qua người dẫn đường hay người đã được nghiên cứu ở gần hay có mối liên hệ tốt với người không tiếp cận được. Thông qua đó xác định việc suy rộng kết quả từ chọn mẫu có đảm bảo đại diện cho tổng thể điều tra hay không.

1.5. Khó khăn chọn mẫu do nội dung nghiên cứu

Thứ nhất, các nội dung nghiên cứu thông qua xác định các chỉ báo, biến số, thang đo thường khá nhiều tùy theo các đề tài nghiên cứu. Về nguyên tắc mỗi giá trị đo lường sẽ có một lượng mẫu phù hợp và nhà nghiên cứu phải tính toán cỡ mẫu tương ứng với độ chính xác mong muốn, sau đó cộng hết các cỡ mẫu đó để có một lượng mẫu tổng thể. Với cách làm này có thể dẫn đến lượng mẫu cần thiết có thể lớn hơn cả tổng thể điều tra (vì mỗi cuộc điều tra có thể có từ vài trăm đến hàng ngàn giá trị đo lường). Ngược lại, nhà nghiên cứu có thể lấy cỡ mẫu tổng thể điều tra theo giá trị đo lường nào có giá trị lớn nhất (với điều kiện các giá trị đo lường phải được

chứng minh có quan hệ độc lập lẫn nhau). Nếu mục đích nghiên cứu cần phân tích tương quan giữa nhiều loại biến số theo một chiều hai chiều hay nhiều chiều khi đó cỡ mẫu tổng thể điều tra cũng cần đảm bảo lượng mẫu tối thiểu để phân tích thống kê theo một giá trị đo lường là 30 hay 50 mẫu (luật số lớn). Xu hướng ngày nay nghiên cứu xã hội học đang chia nhỏ nhiều nhóm nghiên cứu theo nhiều chiều cạnh khác nhau do vậy khi tính cỡ mẫu cho mỗi cuộc điều tra cần quan tâm giải quyết vấn đề này.

Thứ hai, khi nội dung vấn đề nghiên cứu gắn với đặc điểm của từng đơn vị nghiên cứu có thể xuất hiện nhiều hay rất ít xuất hiện trong tổng thể điều tra. Về nguyên tắc khi vấn đề nghiên cứu xuất hiện nhiều thì lượng mẫu không cần nhiều nhưng khi rất ít xuất hiện như 1/1000 hay 1/1000000□ sẽ dẫn đến lượng mẫu được lựa chọn phải rất lớn để đảm bảo số lượng tối thiểu cần thiết xuất hiện trong mẫu được lựa chọn (không được nhỏ hơn 30).

Với các khó khăn nêu trên cho thấy, việc chọn mẫu ngẫu nhiên luôn là những thách thức liên tục và yêu cầu sự cân nhắc kỹ lưỡng để đảm bảo tính đại diện và chính xác của mẫu trong quá trình nghiên cứu. Các nhà nghiên cứu khoa học nói chung và xã hội học nói riêng hiện nay đang đối mặt và cố gắng giải quyết những thách thức này thông qua sự sáng tạo trong phương pháp nghiên cứu và sử dụng các công nghệ mới.

2. Các cách khắc phục khó khăn trong chọn mẫu ngẫu nhiên

2.1. Xu hướng mở rộng và cụ thể hóa cách chọn mẫu phi ngẫu nhiên áp dụng cho nghiên cứu định lượng

Xu hướng ứng dụng phương pháp chọn mẫu theo tỷ lệ (trong nghiên cứu định tính) được áp dụng cho nghiên cứu định lượng trên phạm vi rộng. Trên thực tế, có một số nghiên cứu mặc dù biết cách chọn mẫu định mức (tỷ lệ) là phi xác suất, kết quả nghiên cứu không có khả năng suy rộng cho tổng thể, nhưng để so sánh giữa các nhóm đối chứng có đặc điểm đồng nhất giữa các đơn vị nghiên cứu nhà nghiên cứu đã tăng số lượng mẫu nghiên cứu dựa trên chia nhiều nhóm có đặc điểm đồng nhất theo nhiều biến số. Theo Nguyễn Đình Thọ⁽²⁾ số lượng mẫu có thể lên 200 người được chọn theo 18 nhóm dựa trên 3 biến số: giới tính (thang đo 2 bậc); độ tuổi (thang đo 3 bậc); thu nhập (thang đo 3 bậc). Mỗi nhóm sẽ chọn từ 6 đến 20 người tùy theo tỷ lệ đặt ra cho các nhóm. Theo Phạm Văn Quyết, Nguyễn Quý Thanh⁽³⁾ dựa trên 2 biến số giới tính (thang đo 2 bậc); độ tuổi (thang đo 3 bậc) có thể chia 6 nhóm với mỗi nhóm từ 100 đến 300 người từ đó nhà nghiên cứu chọn

được 1000 người để nghiên cứu. Trên cơ sở các nhóm có đặc điểm về giới tính và độ tuổi (cũng như thu nhập trong ví dụ của Nguyễn Đình Thọ) nhà nghiên cứu giao cho các điều tra viên tự tìm người để phỏng vấn nếu họ có đặc điểm (ví dụ nam giới, dưới 18 tuổi) với một số lượng nhất định dựa trên số lượng mẫu đã tính và số điều tra viên được huy động tham gia nghiên cứu. Trong nghiên cứu xã hội học nhiều khách thể nghiên cứu khó xác định trong thực tế như: gái mại dâm, người đồng tính, người chuyển giới, người sử dụng ma túy... Nhà nghiên cứu có thể đưa ra nguyên tắc chọn mẫu với số lượng lớn theo mô hình quả bóng tuyết (phương pháp chọn mẫu tích lũy nhanh, tăng nhanh, phát triển mầm, theo chuỗi⁽⁴⁾). Khi đó số lượng mẫu nghiên cứu có thể tăng gần 200 người qua nhiều lớp nghiên cứu. Thậm chí, đã có một số nhà nghiên cứu đưa ra những ý tưởng để biến loại nghiên cứu này tiến gần đến chọn mẫu ngẫu nhiên theo xác suất và đưa ra các công thức tính toán cỡ mẫu với độ chính xác nhất định với tên gọi là chọn mẫu □ Dây chuyền có kiểm soát □ (Respondent-driven sampling- RDS) và phương pháp chọn mẫu theo thời gian - địa điểm (Time Location Sampling-TLS)⁽⁵⁾. Hai phương pháp chọn mẫu này chỉ áp dụng cho các khách thể khó xác định (chỉ có thể tiếp cận được thông qua mạng lưới xã hội họ có quen biết nhau), trong phạm vi hẹp với thời gian nghiên cứu nhất định. Về cơ bản các tác giả của các phương pháp này vẫn khẳng định đây là phương pháp chọn mẫu phi xác suất (nghiên cứu trường hợp trong phạm vi hẹp) và trong khu vực hẹp họ tìm mọi cách để nghiên cứu toàn bộ các đơn vị nghiên cứu thuộc khách thể (như vậy đây là nghiên cứu tổng thể do vậy các tính toán chỉ tiêu thống kê có thể áp dụng được). Tuy nhiên, hai phương pháp này hiện vẫn chưa được nhiều nhà khoa học khẳng định tính chính xác về lý luận và thực tiễn.

Chọn mẫu tự phát cũng là xu hướng đang được nhiều cuộc điều tra sử dụng, đặc biệt khi Internet và các mạng xã hội ngày càng phát triển và người dân phần lớn sử dụng điện thoại thông minh và thường xuyên tiếp cận với thông tin qua mạng xã hội. Theo Bộ Thông tin và Truyền thông hiện nay trên 84% người trưởng thành Việt Nam sử dụng điện thoại thông minh và phần đầu năm 2025 đạt 100%⁽⁶⁾. Do vậy, khi đưa ra các câu hỏi trên Internet sẽ có rất nhiều người tham gia trả lời và nhà nghiên cứu có thể thu được thông tin từ rất nhiều người trong thời gian ngắn, xử lý kết quả cũng có thể có được ngay sau khi người trả lời lựa chọn phương án phù hợp và từ đó tiết kiệm rất nhiều chi phí cho nghiên

cứu. Tuy nhiên, phương pháp này cũng có nhược điểm lớn như chỉ hỏi được rất ít câu hỏi và rất nhiều người không tham gia nghiên cứu làm sai lệch kết quả nghiên cứu khi suy rộng cho tổng thể điều tra.

2.2. Một số cách khắc phục trong xây dựng khung chọn mẫu nghiên cứu

Ngay từ những năm đầu của thế kỷ XX nhiều điều tra nghiên cứu dư luận xã hội đã sử dụng niên giám điện thoại để lựa chọn mẫu nghiên cứu⁽⁷⁾. Tuy nhiên, kết quả nghiên cứu qua niên giám điện thoại tại Mỹ năm 1936 đã không đúng như kết quả bầu cử thực tế, trong đó lý do chủ yếu là những người có điện thoại lúc đó là những người có mức sống trên trung bình. Phỏng vấn qua điện thoại được sử dụng rộng rãi ở nơi có mức độ sở hữu cao điện thoại. Ở một số nơi thuộc các nước phát triển, sau chiến tranh thế giới lần thứ hai hầu hết các hộ gia đình đều sử dụng điện thoại, phương pháp này được sử dụng nhiều hơn. Một trong những ưu điểm của phương pháp này được lý giải: do hệ thống an ninh trong các tòa nhà chung cư và các hạn chế khác đối với việc tiếp cận với mọi người trong nhà ở buộc các nhà nghiên cứu phải phỏng vấn thông qua điện thoại. Về mặt tích cực, phỏng vấn qua điện thoại mất ít thời gian hơn so với phỏng vấn trực tiếp thông thường. Người phỏng vấn ở tại một nơi và được cung cấp thiết bị đặc biệt để thực hiện các cuộc gọi điện thoại và ghi lại câu trả lời của người trả lời. Với các phần mềm máy tính hiện có đã tạo điều kiện cho việc nhập dữ liệu ngay lập tức để có kết quả ngay khi mẫu phỏng vấn được hoàn thành. Người phỏng vấn khi đó không phải điền vào các bảng hỏi, không mất thời gian và không phải trả tiền cho việc đi lại hoặc tìm kiếm người trả lời. Thông thường nhiều cuộc phỏng vấn có thể được thực hiện trong cùng một thời gian. Kiểm tra lại được thực hiện bởi người giám sát điều tra lại một tỷ lệ nhất định của tất cả những người được hỏi (thường là 5%).

Phỏng vấn qua điện thoại có một số nhược điểm lớn. Ngay cả ở Hoa Kỳ giàu có, khoảng 5% hộ gia đình không có điện thoại. Ở hầu hết các quốc gia, tỷ lệ không có điện thoại dĩ nhiên là cao hơn. Phương pháp này không phù hợp với các cuộc khảo sát về dân số nói chung ở các quốc gia nơi sử dụng điện thoại rất thấp. Tuy nhiên, các cuộc khảo sát sử dụng điện thoại có thể phù hợp trong đó các nhóm người cụ thể là mục tiêu nghiên cứu gắn với có điện thoại. Phỏng vấn qua điện thoại có thể là một cách thích hợp để đánh giá việc sử dụng truyền hình vệ tinh ở nhiều nước châu Phi và châu Á.

Một nhược điểm khác của phỏng vấn qua điện thoại là nó không phải là một phương tiện phù hợp để có được

nhiều thông tin chi tiết. Những người được hỏi đôi khi phần nộ và từ chối một cuộc gọi điện thoại bất ngờ từ một người hoàn toàn xa lạ, đặc biệt nếu nó trở thành một cuộc gọi dài. Hơn thế nữa, nếu một người tình cờ được thưởng thức các chương trình phát thanh hoặc truyền hình yêu thích, người ta có thể đặc biệt bức bối vì bị gián đoạn để trả lời các câu hỏi về nó.

Cách khắc phục danh sách chọn mẫu hiện nay có thể được sử dụng ngày càng rộng rãi là chọn theo khu vực dựa vào bản đồ của Google map, các danh sách chọn mẫu thường được áp dụng dựa trên cách chọn ngẫu nhiên phân chùm (cụm) bao gồm danh sách các chùm với cộng đồng dân cư lớn như xã/ phường; (hay qua nhiều giai đoạn bao gồm danh sách quận/ huyện hay tỉnh/ thành phố). Dựa trên các chùm đã được lựa chọn nhà nghiên cứu cần lập danh sách các đơn vị nghiên cứu (thường là cá nhân hay hộ gia đình) hay danh sách là các nóc nhà được đánh dấu qua bản đồ Google map (cần loại trừ các nóc nhà không phải nơi ở như cơ quan, trường học, trung tâm thương mại...), tại các đô thị, nơi ở có nhiều loại hình nhà ở như chung cư cao tầng, thấp tầng và nhà liền kề khi đó nhà nghiên cứu cần thiết kế danh sách chọn mẫu sao cho đảm bảo xác suất lựa chọn các đơn vị nghiên cứu ngang bằng nhau⁽⁸⁾. Trong một số trường hợp nghiên cứu việc xây dựng danh sách các chùm đối với một số khách thể nghiên cứu tại từng khu vực theo chùm có sự biến động khá lớn về khách thể nghiên cứu theo thời gian như nhà hàng, khách sạn, trường học, câu lạc bộ, công viên, nơi vui chơi công cộng... việc xây dựng danh sách chọn mẫu còn phải đảm bảo theo thời gian là giờ, ngày trong tuần, mùa...

2.3. Một số giải pháp đảm bảo tăng tỷ lệ người tham gia nghiên cứu

Tăng cường giao tiếp và tương tác với cộng đồng có thể làm gia tăng sự tin cậy, tin tưởng của người dân vào các nghiên cứu xã hội học, họ hy vọng có thể truyền tải được ý kiến, quan điểm của mình đến các nhà quản lý hay nhà khoa học. Các hoạt động truyền thông, thông báo về cuộc điều tra qua báo chí, mạng xã hội, loa truyền thanh, người dẫn đường, người phỏng vấn... thường làm cho người dân thấy được nghĩa vụ và trách nhiệm của mình tham gia nghiên cứu, từ đó làm gia tăng tỷ lệ người tham gia nghiên cứu.

Trong quá trình chọn mẫu thường gắn liền với các hoạt động thu thập thông tin, giao tiếp với nhiều người tại cộng đồng việc đảm bảo an toàn và thuận lợi cũng là các giải pháp để tăng tỷ lệ người tham gia nghiên cứu. Người tham gia nghiên cứu thường rất e ngại nếu các ý

kiến, quan điểm của họ không được giữ bí mật, kể cả những người cùng tham gia thu thập thông tin cũng không được kể lại các thông tin đã được người trả lời cung cấp. Các thông tin trích dẫn cần đảm bảo tính khuyết danh và chỉ dùng cho mục đích nghiên cứu khoa học. Trong quá trình tiến hành chọn mẫu và thu thập thông tin nhóm nghiên cứu luôn chú ý đến sự thuận lợi dành người tham gia nghiên cứu đảm bảo chỉ gặp gỡ trao đổi thông tin khi người cung cấp thông tin có thời gian rảnh rỗi, khi đã hẹn gặp mà không bỏ trí thời gian gặp được cần xin lỗi và trao đổi để hẹn vào thời gian khác khi được họ đồng ý. Trình tự ưu tiên trong quá trình nghiên cứu luôn đảm bảo tính thuận lợi cho người trả lời.

Mỗi phương pháp chọn mẫu ngẫu nhiên thường có những ưu điểm và nhược điểm, nhà nghiên cứu cần nắm vững các kiến thức cơ bản và phối hợp sử dụng nhiều phương pháp, kỹ thuật khác nhau nhằm hạn chế nhược điểm của từng phương pháp thông qua sự phối hợp để phát huy các ưu điểm của các phương pháp khác. Các sự phối hợp đó cả thể bao gồm cả chọn mẫu phi xác suất với chọn theo xác suất (như đã nêu ở trên), các phương pháp chọn mẫu áp dụng cho tổng thể điều tra hạn hẹp (chọn ngẫu nhiên hoàn lại, không hoàn lại, chọn ngẫu nhiên hệ thống) với các phương pháp chọn mẫu ngẫu nhiên có thể áp dụng cho tổng lớn như chọn mẫu phân chùm, phân tổ hay nhiều giai đoạn.

Thông qua việc phân tích những khó khăn thường gặp trong quá trình thiết kế quy trình chọn mẫu ngẫu nhiên của nghiên cứu xã hội học định lượng, đồng thời đưa ra một số cách khắc phục dựa trên kinh nghiệm tập hợp được của các nhà nghiên cứu thực nghiệm trong nhiều lĩnh vực trên thế giới về chọn mẫu có thể giúp các nhà nghiên cứu xã hội học Việt Nam vận dụng trong chọn mẫu nghiên cứu định lượng./.

(1), (4), (8) Lưu Hồng Minh, Phạm Hương Trà (2003), *Giáo trình kỹ thuật thu thập thông tin và chọn mẫu trong nghiên cứu xã hội học*, Nxb. Đại học Quốc gia Hà Nội, tr.153,154,144, 211-223.

(2) Nguyễn Đình Thọ (2014), *Giáo trình phương pháp nghiên cứu khoa học trong kinh doanh*, Nxb. Tài chính (ấn bản lần thứ hai), Tp Hồ Chí Minh, tr.231-233.

(3) Phạm Văn Quyết, Nguyễn Quý Thanh (2001), *Phương pháp nghiên cứu xã hội học*, Nxb. Đại học Quốc gia Hà Nội, tr.225-226.

(5) Hoàng Văn Minh, Trường Đại học Y tế Công cộng (2020), *Phương pháp chọn mẫu và tính toán cỡ mẫu trong nghiên cứu khoa học sức khỏe*, Hà Nội, tr.69-77

(6) *Hướng đến 100% người dùng điện thoại thông minh vào cuối năm 2024*, Tạp chí Tuyên giáo điện tử ngày 23/2/2024.

(7) Nguyễn Thị Thanh Hương (2019), *Nghiên cứu công chúng truyền thông đại chúng (Tổng quan các tài liệu nước ngoài)*, Sách tham khảo: *Công chúng truyền thông đại chúng đầu thế kỷ XXI*, Khoa Xã hội học và Phát triển, Học viện Báo chí và Tuyên truyền, Hà Nội.