

2. TỔNG QUAN CUỘC TỔNG ĐIỀU TRA DÂN SỐ NĂM 2009: MẪU ĐIỀU TRA, CẤU TRÚC SỐ LIỆU VÀ CÁC BIẾN SỐ CHÍNH

Do thiếu một hệ thống đăng ký hộ tịch cung cấp số liệu sinh hàng năm theo giới tính, TĐTDS là nguồn số liệu đáng tin cậy hơn cả cho việc phân tích sự khác biệt TSGTKS. Số liệu của TĐTDS có những ưu điểm sau: tuổi và giới là các biến số tin cậy thu thập được qua cuộc TĐTDS; nhiều biến số kinh tế-xã hội cũng được thu thập đồng thời và có thể sử dụng cho phân tích xu hướng và mối tương quan kinh tế-xã hội; số liệu TĐTDS là toàn diện và có tới đơn vị hành chính nhỏ. Nghiên cứu này dựa trên số liệu mẫu của TĐTDS năm 2009 và sử dụng các biến số cá nhân và hộ gia đình. Mặc dù số liệu toàn diện TĐTDS sẽ cho phép phân tích chi tiết hơn trong thời gian tới, mẫu 15% của TĐTDS là cần thiết cho các phân tích sâu về tình trạng mất cân bằng giới tính khi sinh ở Việt Nam. Bộ số liệu này có nhiều ưu điểm, thứ nhất, kích thước mẫu lớn cho phép thực hiện nhiều phân tích khác nhau (xem Phụ lục 1). Hơn nữa, số liệu mẫu cho phép tiếp cận với các biến số không có trong số liệu toàn bộ của TĐTDS. Số liệu này cũng tạo cơ hội duy nhất cho việc phân tích mối tương quan giữa TSGTKS và các đặc tính ít được nghiên cứu của người mẹ và hộ gia đình.

Mẫu điều tra là một phần của TĐTDS, nhưng số liệu được thu thập dựa trên phiếu điều tra dân số đầy đủ hơn (phiếu dài), trong khi thông tin của phần dân số còn lại được thu thập dựa trên phiếu điều tra ngắn hơn. Bộ số liệu này bao gồm tất cả các thông tin thu thập được trong cuộc TĐTDS ngày 1/4/2009 (ngày tham chiếu). Mẫu chiếm 15% tổng dân số cả nước, ước lượng khoảng 85.789.573 người. Cụ thể là số liệu bao gồm 3.692.042 hộ gia đình và 14.177.590 người. Số liệu này được tổ chức thành hai bộ (bộ số liệu cá nhân và bộ số liệu hộ gia đình) được kết nối với nhau thông qua một bộ mã duy nhất xây dựng từ mã tỉnh, mã huyện và mã hộ gia đình.

Mẫu được lựa chọn đảm bảo tính đại diện về mặt thống kê cho các đơn vị hành chính cấp tỉnh và huyện, cũng như các khu vực nông thôn-thành thị trong cả nước. Vì vậy, một số khu vực được lấy mẫu vượt và do vậy tổng số quan sát lớn hơn 15% (16,5% tổng số nhân khẩu của cả nước). Số liệu mẫu được gia quyền thích hợp. Toàn bộ các kết quả trong báo cáo này đều được gia quyền dựa trên quyền số cá nhân và hộ gia đình do TCTK cung cấp. Cần lưu ý rằng mẫu chỉ bao gồm các hộ trên cơ sở gia đình chuẩn, không bao gồm các “hộ tập thể” như trường học, khách sạn, doanh trại quân đội, cảnh sát và trại giam. Bảng 2 trình bày các biến số theo đặc tính kinh tế-xã hội, nhân khẩu học và địa lý của dân cư và hộ gia đình trong TĐTDS.

Bảng 2: Các biến số chính trong Tổng điều tra Dân số

	Số lượng biến số
Thông tin cá nhân	
Thông tin cư trú: nông thôn/thành thị, vùng, tỉnh	3
Đặc điểm nhân khẩu học: Giới tính, tuổi, quan hệ với chủ hộ, tình trạng hôn nhân	4
Tình trạng di cư	2
Đặc điểm về giáo dục: tình trạng biết đọc biết viết, trình độ học vấn, số năm đi học	4
Dân tộc và tôn giáo	2
Việc làm: tình trạng việc làm, lĩnh vực và hoạt động kinh tế (đối với cá nhân trên 15 tuổi)	11
Số con đã từng sinh ra, số con hiện còn sống, giới tính và tháng và năm sinh của lần sinh gần nhất, bao gồm đa sinh (đối với phụ nữ tuổi 15-49)	4
Thông tin hộ gia đình	
Tình trạng và sở hữu nhà ở	4
Thông tin về nhà ở: sàn nhà, số phòng, vật liệu xây dựng...	5
Tiện nghi trong hộ: loại nhiên liệu (năng lượng) thắp sáng, loại nhiên liệu (năng lượng) để nấu ăn, nguồn nước ăn uống chính, hố xí	4
Tài sản hộ gia đình: tivi, điện thoại, máy vi tính, xe máy ...	8

Tổng cộng có khoảng 50 biến số có thể dùng cho phân tích, tuy nhiên số lượng các biến có thể sử dụng cho phân tích mất cân bằng giới tính khi sinh ít hơn. Các biến số chính sử dụng cho phân tích tỷ số giới tính bao gồm tuổi, giới tính, lần sinh gần nhất. Chi tiết về các biến số này như sau:

- Tuổi được tính theo năm dựa trên thông tin về năm và tháng sinh dương lịch. Phân tích chi tiết về tháp tuổi và giới tính cho thấy có sai số do làm tròn tuổi trong số liệu. Tuy nhiên, hiện tượng sai số này chỉ diễn ra trong nhóm dân số già, vì vậy không có ảnh hưởng tới kết quả phân tích của nghiên cứu này vốn chỉ tập trung phân tích nhóm dân số trẻ em.
- Câu hỏi “lần sinh cuối” có nghĩa là “lần sinh gần nhất”, bao gồm cả những lần sinh diễn ra đã lâu trong quá khứ đối với những phụ nữ cao tuổi. Câu hỏi này được hỏi cho những phụ nữ trong độ tuổi 15-49. Số liệu thu được có thể bị chệch do việc báo cáo thiếu những trẻ em đã tử vong, nhất là đối với những lần sinh đã lâu trong quá khứ.
- TĐTDS 2009 cho phép câu hỏi về lần sinh cuối ghi chép chi tiết những lần đa sinh như sinh đôi, sinh ba. Kết quả là số sinh nhiều hơn số phụ nữ tương ứng. Số sinh vì vậy đã được gia quyền lại (sử dụng quyền số của người mẹ chia cho số sinh được báo cáo).

Tuy nhiên, có thể có sai số do báo cáo thiếu. Điều này có thể do nhiều nguyên nhân bao gồm báo cáo thiếu số sinh các lần sinh thứ 3 trở lên do các quy định về kế hoạch hóa gia đình ở địa phương, và đặc biệt thường báo cáo thiếu trẻ em sinh ra còn sống nhưng chết sau đó. Tuy

nhiên không có dấu hiệu rõ ràng về chênh lệch giới tính khi báo cáo thiếu số sinh ở Việt Nam (như tình trạng báo cáo thiếu số sinh gái ở Trung Quốc).

Hai khái niệm quan trọng trong báo cáo này là TSGTKS và tỷ số giới tính trẻ em (TSGTTE) được mô tả sơ bộ dưới đây (Phụ lục 2 và 3).

- TSGTKS được tính bằng số trẻ em trai được sinh ra trên 100 trẻ em gái. Mức sinh học bình thường trên thế giới là 105. Cũng cần lưu ý rằng TSGTKS khá nhạy cảm với quy mô mẫu (xem Phụ lục 1). TSGTKS được tính toán cho số sinh xảy ra trong vòng 12 tháng trước thời điểm điều tra (từ 1/4/2008 đến 31/3/2009). Sử dụng giai đoạn dài hơn có thể có kết quả chênh lệch đáng kể (xem Phụ lục 2).
- Khi ước lượng của TSGTKS không tin cậy hoặc thiếu, có thể dựa vào TSGT của trẻ em được tính toán cho nhóm dân số dưới 5 tuổi.