

2016년 국제 조사방법론 심포지엄 참가를 위한

캐나다 국외 출장 결과 보고서

2016. 4.



통계청

I 출장 개요

1. 출장 개요

□ 출장목적

- 2016년 국제 조사방법론 심포지움 참가
 - 조사환경의 개선을 위해 선진 조사과정자료 수집 및 활용 현황, 조사방법 등 통계 선진국의 사례를 벤치마킹
 - 또한, 세계 조사방법 동향, 통계기법 등 최신 이슈와 시사점을 파악

□ 출장자

- 조사관리국 조사기획과 나수빈 주무관

□ 출장기간 및 출장지

- 2016. 3. 21. ~ 3. 26. (4박 6일)
- 캐나다 오타와

□ 출장일정

날짜	업무수행내용	담당자
3.21(월)	○ 출국(인천→시카고→오타와)	
3.22(화)~24(목)	○ 2016 국제방법론 심포지움 참가	Statistics Canada
3.25(금)~26(토)	○ 귀국(오타와→토론토→인천)	

□ 심포지엄 일정

Day 1 - 2016.03.22(화)

08:45 - 9:00	Opening Remarks	
09:00-10:00 Plenary Session	Session 1: Keynote Address	
10:00-10:30	Morning Break	
10:30-12:00 Concurrent Sessions	Session 2A: Big Data in Official Statistics	Session 2B: Applications Related to Growth in Statistical Information
12:00-13:30	Lunch	
13:00 -15:00 Concurrent Session	Session 3A: Total Survey Error	Session 3B: Alternative Data Sources to Replace or Complement Survey Data
15:00-15:30	Afternoon Break	
15:30-17:00 Concurrent Sessions	Session 4A: Open Data	Session 4B: Quality of Administrative Data

Day 2 - 2016.03.23(수)

08:45-09:45 Plenary Session	Session 5: Waksberg Award Winner Address	
09:45-10:00 Plenary Session	Speed Advertisement for Posters and Software Demonstration	
10:00-10:30	Poster Session, Software Demonstration and Morning Break	
10:30-12:00 Concurrent Sessions	Session 6A: New Advancements in Record Linkage	Session 6B: Confidentiality
12:00-13:30	Lunch	
13:00 -15:00 Concurrent Session	Session 7A : Non-traditional Methods for Analysis of Survey Data	Session 7B : Applications of Record Linkage and Statistical Matching
15:00-15:30	Poster Session, Software Demonstration and Afternoon Break	
15:30-17:00 Concurrent Sessions	Session 8A: Paradata	Session 8B: Use of Administrative Data

Day 3 - 2016.03.24(목)

08:45-10:15 Concurrent Sessions	Session 9A: Scanner Data	Session 9B: Health Data
10:15-10:45	Morning Break	
10:45-11:45 Plenary Session	Session 10: Data Science for Dynamic Data Systems: Implications for Official Statistics	
11:45-12:00 Plenary Session	Session 11: Closing Remarks	

2. 심포지엄 개요

□ 명칭

- 2016 국제조사방법론 심포지엄(International Methodology Symposium)

□ 목적

- 국제조사방법론 심포지엄은 캐나다 통계청 주관으로 매년 개최되어 각 국 통계청 실무자 및 대학교 교수진이 모여 조사방법론에 관한 세계연구동향, 최신 통계기법 등을 발표하고 연구 결과 교류

□ 일시 및 장소

- 2016. 03. 22. ~ 03. 24. (3일간)
- 캐나다 가티노

□ 주요 내용

- 이번 심포지엄은 ‘통계정보의 성장: 도전과 장점’이라는 주제로 3일 간의 심포지엄으로 이루어졌으며, 그 중 심포지엄에 참여
 - 심포지엄은 빅데이터, 총 오차, 개방데이터, 행정자료의 품질, 행정자료의 이용, 파라데이터, 레코드 연계의 새로운 진보, 비전통적인 데이터 분석 방법, 스캐너 데이터 등 다양한 내용을 발표 및 논의

II 심포지엄의 주요 내용

◆ Session 8A : 조사과정자료(Paradata) 中

<주요 국가 통계에서의 파라데이터 활용 : 도전과 장점>

(Brady T. West, University of Michigan-Ann Arbor,

Frauke Kreuter, University of Maryland-College Park)

□ 요약 : 미국의 NSFG의 조사과정자료 품질 및 활용, 미국과 독일의 주요 패널조사에서의 조사과정자료 활용 연구

① 미국의 NSFG(National Survey of Family Growth)

○ 개념 : 미국의 대표적인 출산 관련 조사로 결혼, 성행위, 출산 등의 정책을 위한 중요한 자료를 제공하는 조사

- 표본설계(모집단) : 15세~44세(2015년 9월까지),

15세~49세(2015년 9월부터 현재)

- 'continuous design' 표본을 설계하여 매년 4개의 국가 주택지 샘플을 발표하고 연구하고 있음

- 대면조사 방식, 1명 조사원이 1가구에 대해 조사하며 60~80분정도 소요됨

- ACASI방식(audio computer-assisted self-administered interviewing) : 민감한 질문에 대해 사용

○ NSFG의 조사과정자료

- 조사원의 관찰 : 지역 수준(ex)안전문제), 가구 수준(ex) 어린자녀 유무), 이전 조사 관찰(예)ACASI 사용 여부) 등
- 전화 수집 데이터 : 전화 시도 횟수, 응답 거부 등
- 사례별 결과 : 응답자, 최종 응답거부, 비표본 등
- Keystroke 정보 : 응답자의 응답변경, 각 조사문항별 응답 시간, 조사원의 조사협조 요청 등

② 미국의 MEPS(Medical Expenditure Panel Survey)

○ 미국의 의료 지출 패널조사

- National Health Interview Survey(NHIS)의 표본을 추출 하여 대면조사, 2년동안 5회 순환

○ 조사과정자료 : 전화 기록 자료(전화 시도 횟수, 응답 거부 등 증거), 사례별 결과(응답자, 최종 응답거부, 비표본 등)

③ 독일의 PASS(Labour Market and Social Security)

○ 연간 혼합 모드 가구 조사 : 두 개의 랜덤샘플 사용(복지 연금 수령인, 주택조사의 가구)

○ 조사과정자료 : 전화 기록 자료, 사례별 결과, 조사원 관찰

□ 조사과정자료 활용

① 무응답 조정

- 가중치 조정을 한 응답성향 모델을 포함한 모든 샘플에 대해 조사원 관찰로 수집
- 결과와 응답성향과 관련된 관찰을 통해 무응답 편향을 감소할 수 있음
 - PASS : 수입 납세계층(저,중,고)과 복지 연금 수령

※ 도전 1 : 관측 품질

- 만약 관측에 오차가 있었다면?
 - (성생활 : 정확도 78% / 어린자녀 : 정확도 72%
 - 연금수급자 : 정확도 78% / 납세계층 : 정확도 55%)
- 정확도는 조사원에 따라 대체로 달라진다. 왜?
- 오류발생이 쉬운 관측은 무응답 조정의 효과를 방해함

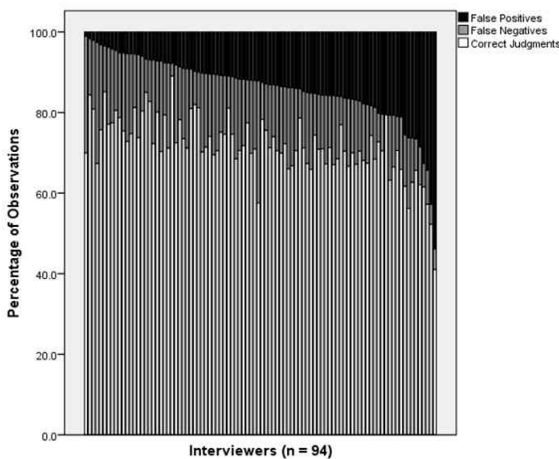


Figure 1: Variance in Observation Accuracy Among 94 NSFG Interviewers.

UB:	Unemployment benefit: Self-reported	
	On UB	Not on UB
Interviewer Observed	n=1866	n=1347
On UB (n=1906)	72.8%	27.2%
Not on UB (n=1234)	21.9%	78.1%
Missing (n=73)	43.8%	56.2%

Income:	Income: Self-reported		
	Low	Medium	High
Interviewer Observed	n=1961	n=684	n=568
Low (n=1511)	82.3%	13.7%	4.0%
Medium (n=1362)	45.2%	29.2%	25.6%
High (n=267)	19.1%	24.7%	56.2%
Missing (n=73)	69.9%	17.8%	12.3%

○ 이에 대한 해결 방법은?

- 하나의 아이디어 : 관찰 경향의 형태에 대한 중요한 예측을 조사원들에게 제공하는 것
- 다른 조사원이 관측하는 방법에 대한 전략과 단서를 이해하고, 가장 효과적인 전략에 대한 표준화된 트레이닝을 실시

② 조사원 평가

- 여러 타입의 조사과정자료로 적합성, 접촉(일별), 협조도(일별) 성향 모델을 알 수 있음 - 기댓값 측정에 이용
- 조사원 행동으로 기댓값 결과와 편차의 평균 등을 비교하여 평가할 수 있음
- Keystorke 정보를 이용하여 조사원의 정확한 응답을 유도하거나 너무 빠르게 이동하는 특수한 경향을 확인
- 실제 조사시 문제 상황에 개입하여 조사 성과를 향상시킴
- 요인1 : 너무 빠름(모든 질문에 제로 베이스인 상태)

요인2 : 오차가 큼

요인3 : '잘 모르겠다'와 '응답거절'이 많음

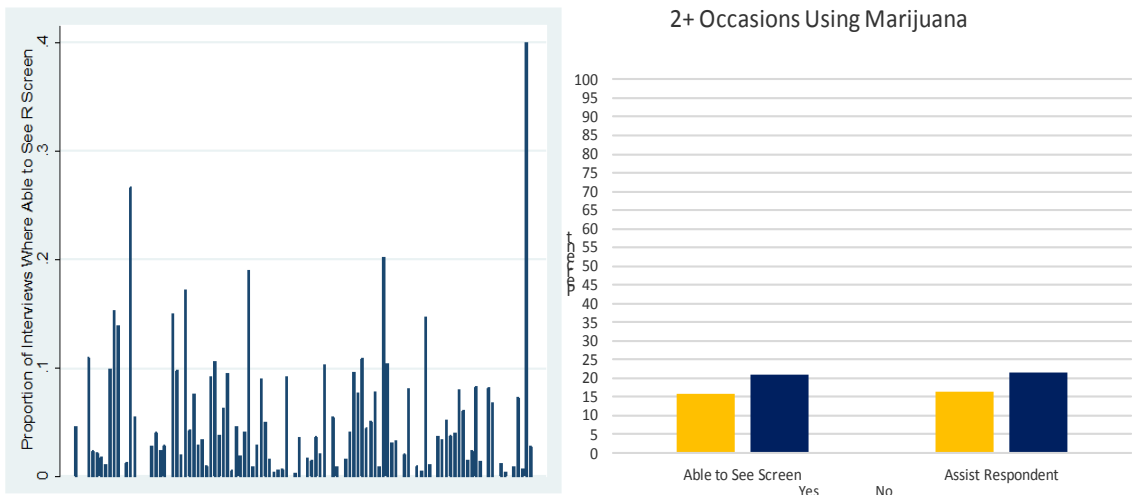
Average of Zscore		Column Labels						
		<input type="checkbox"/> C8 <input type="checkbox"/> Y2 <input type="checkbox"/> Q01 <input type="checkbox"/> Q02						
Row Labels		W08	W10	W12	W06	W08	W10	W12
<input type="checkbox"/> Margolis, Suzanne								
factor1		-0.36	-0.52	-0.54	-0.79	-0.83	-0.74	-0.82
factor2		3.48	3.41	3.25	-0.83	-0.36	0.26	0.16
factor3		0.53	0.74	1.47	0.08	0.50	1.20	1.44

※ 도전 2 : 모델 규격

- 조사대상 성향모델이 정확히 규격되었는지 알 수 있을까?
- 오류발생이 쉬운 조사원의 관찰은 오류 재발생 확률이 높음

③ 응답 품질

- 과거 조사자료를 관측하여 나쁜 품질의 데이터를 제공하는 응답자를 확인
- 민감한 질문에 대한 ACASI응답에 영향을 주는 조사원의 행동
 - 조사원이 컴퓨터 스크린에 얼마나 충분히 가깝게 앉아있는지도 영향을 미침
- ACASI 보고서에 따르면 조사원이 스크린을 볼 수 있다는 사실을 말하는 등의 민감한 행동도 응답 품질에 영향을 줌

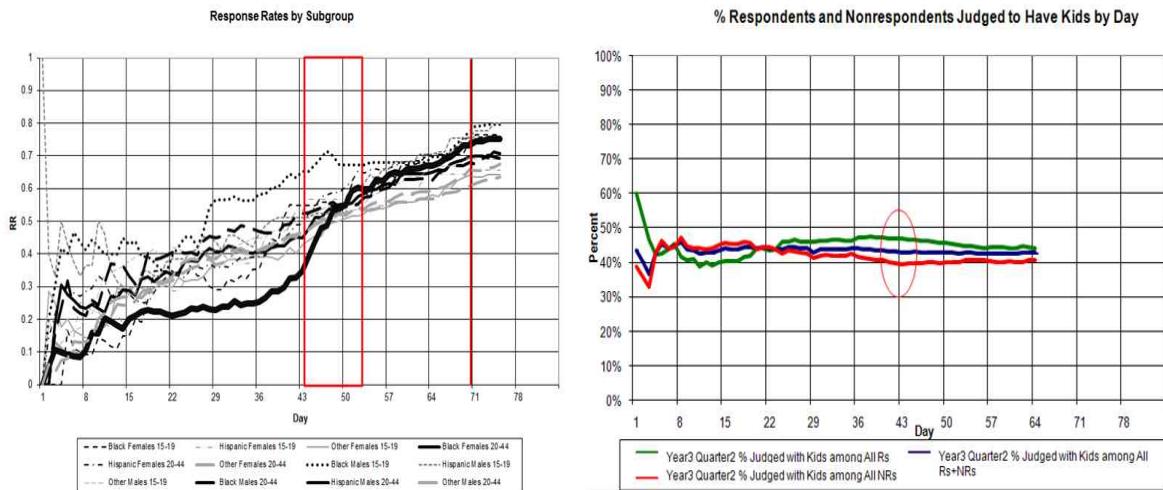


※ 도전 3 : 과거 조사의 관찰

- 과거 조사에 대한 조사원의 관찰이 데이터 품질에 영향을 주는 지표인가?
- 과거 연구에 따르면 응답자 특성이 영향을 미침.
 - 응답자 특성에 대한 편차에 대한 조사원의 지속적인 기록 등
- 위와 같은 조사원의 관측이 데이터 품질에 영향을 주는 신뢰할만한 지표인가?(Open Question)

④ 응답 조사 설계

- NSFG에서 수집한 조사과정자료는 현장의 생산성 및 효율성 모니터링을 위한 응답 조사 설계의 틀로 시험
- 조사과정자료가 특정 과정의 편향이나 비효율성을 초래할 때 조정이 필요
- 사회인구학의 부분군과 줄어드는 부분군에 대한 조사원 증가에 대한 응답률에 대한 모니터링

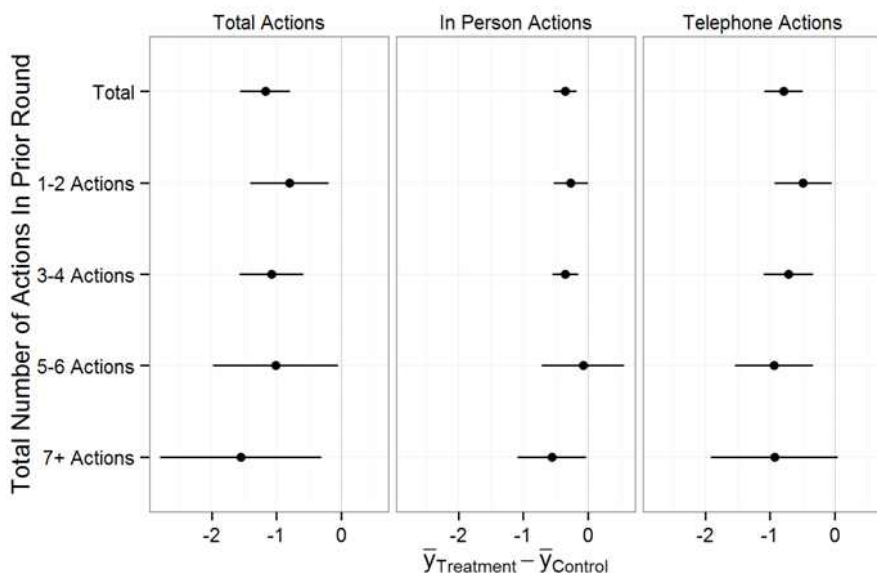


※ 도전 4 : 편향 지표

- 사회인구학 부분군의 응답률은 불응률에 대한 최선의 지표일까?
- 우리는 응답률과 관련된 부분군의 다양한 키 값 중 가장 연관이 깊은 요소를 실제로 모니터링하고 있는가?(ex)자녀의 유무)
- 조사과정자료는 불응률 지표에 대한 다양한 가능성을 소개함

⑤ 전화조사 효율성

- MEPS와 PASS의 조사과정자료는 과거자료에 기반하여 가장 최적의 시간에 접촉하는데 사용됨
- MEPS에서는 방문과 동시에 우편도 발송
- PASS에서는 성공적인 전화 시간대를 이용하여 최초 접촉 시간과 2차 접촉 시간을 감소하였음



※ 도전 5 : 전화조사 효율성

- 이전 전화 조사의 정보를 이용하여 응답자의 잠재된 편향성 정보를 얻음 -> 더 많은 조사에서 변화에 대한 예측이 필요함
- 기존의 접촉 전략 규정은 현장조사 과정과 맞지 않는 경우가 있음 : 다른 대안이 필요
- 최적화된 접촉시간 배정이 중요 요소

⇒ NSFG, MEPS, PASS의 3개 조사에 대해 최적의 데이터 생산성을 향상시키기 위해 조사과정자료를 다양하게 활용

⇒ 데이터에 대한 오차 편향 측정방법을 위한 조사과정자료 연구가 필요함