

유럽조사연구학회 (ESRA) 컨퍼런스 참가 결과 보고

2013. 8.

I 개요

1. 회의 개요

○ 참가회의

- 명 칭 : 제5회 유럽조사연구학회 컨퍼런스
- 기 간 : 2013.7.15(월) ~ 7.19(금)
- 개최지 : 슬로베니아, 류블라나

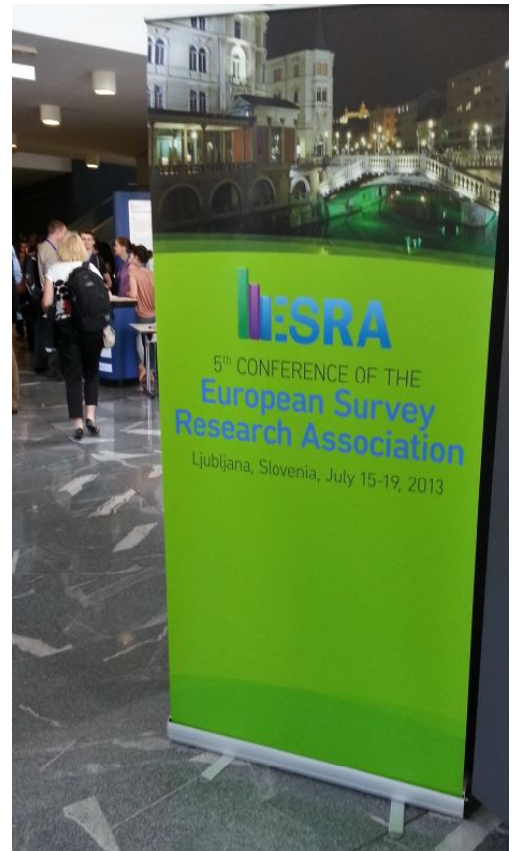
○ 참가규모

- 160여개의 세션에서 조사방법론 관련 다양한 주제 발표
- 유럽, 미국 등 세계 각국에서 조사방법론 관련 전문가 600여명이 참석

.이중 아시아권 참석자는 10명가량(일본, 방글라데시, 인도 등)

○ 주요 내용

- 본 컨퍼런스는 표본추출, 무응답, 가중치 등 전통적인 조사 방법 영역은 물론이고, 조사표설계, 조사원오차, 파라데이터와 현장조사, 데이터 아카이브 등 조사방법론과 관련한 전 영역에 걸쳐 많은 논문 발표



워크숍 주요 주제

- Sampling for cross-national surveys
- Mode effects in mixed-mode surveys
- Do pretesting methods identify 'real' problems and help us develop 'better' questions?
- Assessing the quality of survey data
- Using paradata to improve survey data quality
- Not able to participate: a neglected cause of nonresponse
- Innovations in measurement instrument construction for web-based

surveys

- Weighting: Approach and Sources
- Web data collection for probability-based general population survey
- The trouble with Logit and Probit

2. 출장 개요

- 출 장 자 : 표본과 심경보 주무관, 동향분석실 박영실 주무관
 - 출장기간 : 2013. 7. 14 (일) ~ 7.19(금) (*참가기간: 15~17일)
 - 참가목적
 - 국가 간 비교 조사를 위한 표본설계 및 자료의 분석, 패널조사의 가중치 및 문제점 등에 관한 연구동향을 파악하고자 함
 - 통계청에 조사표실험실이 구축됨에 따라 향후 효율적인 운영을 위해 조사표 설계 연구동향을 파악하고, 우리나라 통계청의 연구경험을 발표.공유하고자 함
- * 발표내용: Comparing Questionnaire Pretesting Methods for the 2015 Census of Korea

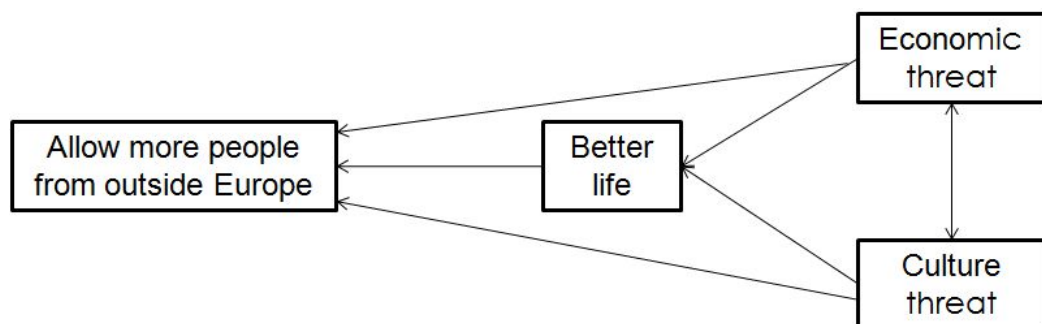
II 주요 내용

1. Key note 발표



- 주제: Correction for measurement error in survey research using SQP
- 발표자: Willem E. Saris (스페인 Pompeu Fabra 대학 교수, Research and Expertise Center for Methodology 소장)
- 주요 내용
 - 조사 자료는 측정오차(measurement error)를 가지고 있고, 그 사실을 모두 알고 있지만 실제 조사 연구 시 오류를 수정하는 경우가 별로 없음
 - 그러나, Saris는 측정오차가 수정될 경우 자료의 품질이 향상될 수 있음을 경험적인 근거를 통해 제시하고 있음
 - .질문문항의 품질을 예측하는 시스템인 SQP(Survey Quality Prediction system)를 이용한 연구 결과를 발표하였음
 - .해당 시스템은 <http://sqp.upf.edu/> 에서 이용가능

- 측정오차가 매우 작은가?
 - .Saris와 동료들의 연구결과에 따르면 오차 분산은 비교적 큰 편임을 알 수 있음
 - .그 오차로 인해 관찰변수 사이의 상관계수 혹은 회귀계수가 잠재변수 사이의 상관계수와 회귀계수와 매우 다를 수 있음
 - .국가 간 오차분산 사이의 차이는 국가 간 비교를 어렵게 함
- 측정오차를 수정하기가 어려운가?
 - .측정오차의 수정은 최근 매우 수월해졌음 : 질문문항별 품질 값과 관찰값의 상관계수를 이용하여 수정할 수 있다!
 - .또한, 측정오차의 수정은 반드시 필요함
- 질문문항과 복합점수(composite scores) 품질의 추정이 어려운가?
 - .질문문항과 복합점수 품질을 추정하는데 있어서 다양한 절차가 있을 수 있으나 연구자가 구현한 SQP 시스템을 통해 품질에 관한 많은 정보 구득 가능함
 - .사용자들은 자신의 질문을 이용하여 질문에 대해 코딩함으로써 품질을 예측할 수 있음
 - .단일 질문의 품질이 알려져 있다면, 복합점수의 품질 또한 추출 가능함
- 사례 : 「비유럽인의 이민에 대한 의견 조사」
 - .이민에 대한 찬성/반대가 “경제적 위협”, “문화적 위협”, “삶의 질”이라는 세 가지 변수에 영향을 받는다고 가정하고 질문지를 작성



.설문 조사 후 SQP 시스템에 구축된 4가지 변수에 대한 예측 값(확신(reliability)계수, 타당성(validity)계수, 방법효과(method effect), 품질(quilty)계수)을 이용하여 각 변수 간 측정오차를 계산하고 이를 수정하기 전과 후의 상관계수 도출
 .그 결과, 수정 전과 수정 후 많은 차이를 나타냄

	Without correction On allow immigration	with correction for errors on allow immigration
By		
Better life	-.265*	-.609*
Economic threat	-.133*	.001
Cultural threat	-.154*	-.140*
Total explained (R ²)	.254	.547
	On better life	on better life
By		
Economic threat	-.310*	-.007
Cultural Threat	.500*	.938*
Total explained (R ²)	.564	.868

- 결론

.측정오차를 수정하지 않고 자료를 분석할 이유가 더 이상 없으며, 심도 있는 연구를 위해서는 측정오차는 반드시 수정되어야 함

.그렇지 않다면 도출된 연구결과에 대한 신뢰성을 담보할 수 없음

○ 총평

- 조사통계자료의 신뢰성이 점점 낮아지고 있는 시대적 배경에 걸맞는 흥미로운 주제라고 생각하며, 조사 자료의 연구 및 분석을 위해 측정오차의 수정은 선택이 아닌 필수라는 것을 강하게 주장하였음

.오차가 표본오차와 비표본오차로 구분되며, 이러한 비표본오

차에 여러 가지 유형이 있고 그 유형들을 최소화하기 위해 현장에서의 보이지 않는 (?) 노력이 필요하다는 것이 기존의 주장이었다면, 비표본 오차를 구체적으로 측정해서 이를 자료분석과 정에서 반영해야 한다는 것은 진일보한 논의라고 생각

2. 참석 세션 내용 [표본추출→가중치→조사표설계→기타 순]

표 집

Estimation and Imputation Under Informative Sampling and Nonresponse

□ Dealing with non-ignorable nonresponse in the German job vacancy survey

Dr Hans Kiesl(Regensburg University of Applied Sciences)

○ 연구문제

- 독일에서는 가용일자리의 총수를 추정하기 위해 조사를 실시
- 8페이지의 분량의 우편조사로 인하여 응답률이 20%정도로 나타났으며, 결국 행정자료인 등록 가용일자리 수를 이용하여 설계가 중치를 보정해야 하는 것으로 나타났음

○ 실험

- simple GREG
 - 등록된 가용일자리 수를 과소추정함
- 6개 중 1, 2번의 보정 추정식을 실험함
 - 두 개의 다른 GREG 추정식 (비교를 위한 등록된 가용일자리 수를 이용하는 방법)

$$w_k = d_k (1 + c_k q x_k)$$

- (1) $c_k = 1$ for all units
- (2) $c_k = 1 / (1 + \text{regvac}_k)$

▪ Reason: GREG minimizes

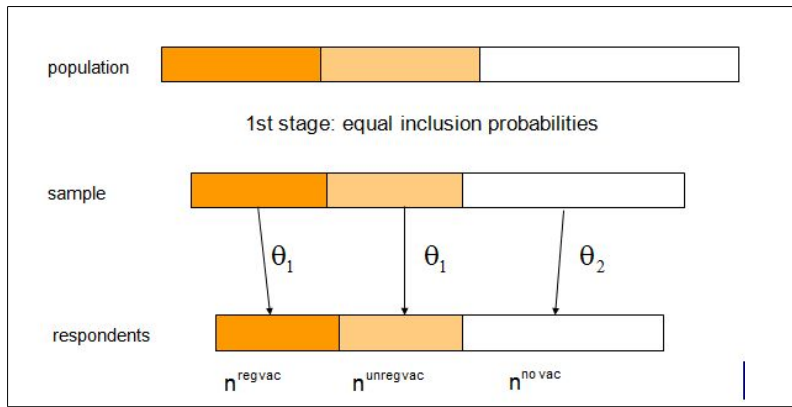
$$\text{dist}(w, d) = \sum_{k=1}^n \frac{(w_k - d_k)^2}{c_k d_k}$$

- 만약 무응답 모형이 맞다면 GREG는 편향됨
- 6개 중 3, 4번의 보정 추정식을 실험함
- 두 개의 다른 일반 보정 추정식 (총수를 아는 등록된 가용일자리수와 무응답변수를 위한 “가용일자리가 있느냐/없느냐” 문항을 이용하는 방법)

$$w_k = d_k f(q \cdot h_k)$$

- (1) $f(q \cdot h_k) = 1 + q \cdot h_k$
- (2) $f(q \cdot h_k) = \exp(q \cdot h_k)$ (raking)

- 만약 무응답 모형이 맞다면, 이 추정식은 편향이 적을 것이다. (비록 분산은 더 커지더라도)
- 6개 중 5, 6번의 보정 추정식을 실험함
- 두 개의 다른 “2단 GREG 추정식”
- 총수를 아는 등록된 가용일자리수를 이용하는 GREG를 사용 (등록된 가용일자리를 가진 사업체를 위한 보정가중치만을 믿는 경우)
- 등록된 가용일자리가 있는 사업체의 응답 성향을 추정
- 등록되지 않은 가용일자리를 가진 사업체 또한 동일한 응답성향을 가정
- 이를 통해, 가용일자리가 없는 사업체의 응답성향을 추정
- 무응답 조정 가중치를 계산하고 다른 GREG 추정식에 해당 가중치를 사용



○ 결론

- 일반 보정 추정식은 큰 분산을 가지고 편향됨
- GREG는 적은 분산을 가지지만 큰 편향을 가짐
- 2단 GREG가 가장 합리적인 것으로 판단됨

□ **Two-phase sampling experiment for propensity score estimation in voluntary samples**

Dr Jae-kwang Kim(Iowa State University) et al.

○ 연구문제

- 자발적 표본(Voluntary Samples)은 추출확률을 알 수 없고, 선택편향이 존재하므로 인구학적 요인에 의해서만 가중치를 조정 (Calibration)하는 전통적인 방법은 편향을 야기함
- 따라서, 자발적 표본의 편향의 요인인 흥미(Interest)를 측정하여 전통적인 가중치 조정방법을 보완하여야 함(Propensity score weighting)

○ 연구사례

- 2012 Iowa Caucus Survey에서 확률표본으로 뽑은 표본집단의 전화응답률이 1차는 15%, 2차는 1차응답자의 75%를 기록하였음
- 이 경우 응답자의 가중치를 어떻게 계산할 것인가?

○ 제안된 모델

○ 기본 설정

U: 목표 유한 모집단

A₁: 표본설계를 알수 없는 1차 자원자 표본

A₂: 표본설계를 알수 없는 2차 자원자 표본

○ 설계

x: 1차와 2차표본에서 동시에 발견되는 변수

y₁: 1차 표본집단의 관심변수

y₂: 2차 표본집단의 관심변수

δ_{1i}, δ_{2i}: 1, 2차표본설계를 위한 표본추출 인디케이터

x변수에 대하여 추가 정보가 있다고 가정(예: \overline{X}_N 값을 알고 있다)

흥미(Interest)에 대한 매개변수: $\theta_1 = E(Y_1)$, $\theta_2 = E(Y_2)$

○ Two-phase approach

집단의 포함관계: $A_2 \subset A_1 \subset U$ (y_{1i} 는 A_1 과 A_2 모두에 존재)

1차 표본추출 원리를 다음으로 설정

$$\pi_{1i}(\phi) = \Pr(\delta_{1i} = 1 | x_i, y_{1i}) = \frac{\exp(\phi_0 + \phi_1 x_i + \phi_2 y_{1i})}{1 + \exp(\phi_0 + \phi_1 x_i + \phi_2 y_{1i})}$$

2차 표본추출 원리는 다음으로 설정

$$\pi_{2i}(\phi^*) = \Pr(\delta_{2i} = 1 | x_i, y_{2i}, \delta_{1i} = 1) = \frac{\exp(\phi_0^* + \phi_1 x_i + \phi_2 y_{2i})}{1 + \exp(\phi_0^* + \phi_1 x_i + \phi_2 y_{2i})}$$

1, 2차 표본추출을 위한 Propensity model은 동일함

N은 알고 있다고 가정

○ Proposed estimator

- Proposed estimator for θ_1 :

$$\hat{\theta}_1 = \frac{1}{N} \sum_{i \in A_1} \frac{y_{1i}}{\pi_{1i}(\hat{\Phi})}, \tag{1}$$

where $\hat{\Phi}$ is a consistent estimator of the true parameter

$\Phi = (\phi_0^*, \phi_0, \phi_1, \phi_2)$.

- Proposed estimator for θ_2 :

$$\hat{\theta}_2 = \frac{1}{N} \sum_{i \in A_2} \frac{y_{2i}}{\pi_{1i}(\hat{\Phi})\pi_{2i}(\hat{\Phi})}. \tag{2}$$

- How to estimate Φ ?

○ Estimating Φ

- Idea: Calibration using $(1, x_i, y_{1i})$ (instrumental variables).

$$U_1(\Phi) := \sum_{i \in A_1} \left\{ \frac{\delta_{2i}}{\pi_{2i}(\phi_0^*, \phi_1, \phi_2)} - 1 \right\} (1, x_i, y_{1i})' = (0, 0, 0)'$$

and

$$U_2(\Phi) := \sum_{i \in A_1} \frac{1}{\pi_{1i}(\phi_0, \phi_1, \phi_2)} - N = 0,$$

where $\Phi = (\phi_0^*, \phi_1, \phi_2, \phi_0)$.

- Note: (x_i, y_{1i}) observed in both A_1 and A_2 & used in computing $\hat{\phi}_2$ in

$$\pi_{2i} = \frac{\exp(\phi_0^* + \phi_1 x_i + \phi_2 y_{2i})}{1 + \exp(\phi_0^* + \phi_1 x_i + \phi_2 y_{2i})}.$$

○ 2012 Iowa Caucus Survey에 적용

- Propensity Model

$$\pi_{1i}(\phi) = \frac{\exp(\phi_0 + \phi_1 X_i + \phi_2 Y_{1i})}{1 + \exp(\phi_0 + \phi_1 X_i + \phi_2 Y_{1i})}$$

and

$$\pi_{2i}(\phi^*) = \frac{\exp(\phi_0^* + \phi_1 X_i + \phi_2 Y_{2i})}{1 + \exp(\phi_0^* + \phi_1 X_i + \phi_2 Y_{2i})},$$

where $X=(\text{Party, Age})$ and $Y=\text{"First choice" of the presidential candidate for Republic party.}$

- Results for propensity model parameter estimation

Table : Point Estimation and Standard Error for Φ

Coefficient	Age	Party	Romney	Perry	Paul	Others
Est	0.588	0.782	0.991	0.454	0.866	1.307
S.E.	0.266	0.251	0.454	0.663	0.841	0.985
t.value	2.211	3.116	2.183	0.685	1.030	1.327

- Results for estimation for θ =proportion of first choice candidate

Table : Point Estimation (s.e.) for 2012 Iowa Caucus Survey Results

Survey	Method	Romney	Perry	Paul	Others
Nov.	Naive	0.340	0.108	0.130	0.422
	Ignorable	0.316	0.103	0.146	0.435
	Proposed	0.303 (0.062)	0.106 (0.039)	0.093 (0.067)	0.499 (0.046)
Dec.	Naive	0.281	0.140	0.131	0.448
	Ignorable	0.270	0.144	0.148	0.437
	Proposed	0.244 (0.043)	0.134 (0.026)	0.112 (0.046)	0.509 (0.036)

- Actual outcome (Jan 3, 2012): $\theta_0 = (24.5\%, 10.3\%, 21.4\%, 43.7\%)$

○ 결론

- capture-recapture sampling과 같은 자발적 표본에 propensity score 추정법을 제시
- 원 참가자로부터 다음 표본이 추출된다면, 동일한 propensity model을 사용하여 추정할 수 있음
- 동일한 모집단에 대하여 독립적인 표본을 추출할 때 역시 해당 추정법을 사용가능함
- 회귀 추정을 통해 모집단 및 표본의 추가 정보를 통합 이용할 수 있음
- Web survey에 적용이 유망할 것으로 보임

가중치

Weighting issues in Panel surveys

□ From PISA to LSAY: Weighting the Australian Longitudinal Surveys of Australian Youth

Mr Patrick Lim(National Centre for Vocational Education Research)

○ 주제 : PISA*를 이용하여 LSAY** 조사 가중치 작성

* PISA : The Program of International Student Assessment, OECD 회원국에서 3년 주기로 의무교육이 끝나는 15세 학생을 대상으로 실시하는 학력평가

** LSAY : Longitudinal Surveys of Australian Youth, 호주 젊은이(15~약25세)들을 대상으로 실시하는 패널조사

○ 연구내용

- LSAY는 15세 학생때부터 약 25세 청년까지의 그룹을 패널로 선정하여 매년 면접조사를 실시하여 학습능력, 학교생활, 고등 교육과정, 노동시장활동, 학습과 경력에 대한 태도 및 의지 등에 관한 정보를 수집함
- PISA는 이단집락추출법을 이용하여 PSU인 학교를 층화확률비례추출법으로 각 코호트(cohort)당 350여개의 학교를 추출, 학교당 FSU인 15세 학생을 50명씩, 대략 14,000여명의 학생을 추출

- LSAY의 1차 웨이브는 PISA에 자체설문을 더하여 질문지 기반의 조사를 실시, 2차 웨이브부터는 CATI를 이용하여 자료수집
- 무응답과 표본마모에 따라 패널탈락이 발생 → 해결을 위해 가중치 조정 실시
- 가중치 조정
 - 1차 웨이브의 PISA 가중치(학교가중치, 개인가중치)를 설계가중치로 설정
 - 각 웨이브마다 응답률과 표본마모에 따라 가중치 조정 실시

<ul style="list-style-type: none"> - 총화변수와 무응답 분석을 통한 결정변수를 이용하여 가중치 조정 실시 - 표본에 대한 모든 정보는 PISA의 조사결과를 이용함 - 다음의 Logistic regression을 이용하여 조정가중치 도출 	
$\text{logit}(y_i) = \alpha_i + \beta X_i + \varepsilon_i$ $\hat{p}_i = \frac{e^{\hat{\alpha} + \hat{\beta}X}}{1 + e^{\hat{\alpha} + \hat{\beta}X}}$ $wt_i = PISA_wt_i \times \frac{1}{\hat{p}_i}$	<p>Where:</p> <ul style="list-style-type: none"> • y is an indicator variable 1/0 representing response and non-response from wave 1 • β is the vector of regression coefficients (including both stratification and variables contributing to attrition) • X is the design matrix for regression coefficients and • ε represents the random error component • The same model is run for each new survey wave.

○ 연구결과

- 무응답조정가중치 적용 결과 총화변수 및 결정적 변수에서 모집단의 형태에 가까운 코호트를 형성할 수 있게 되었음

○ 총평

- 패널조사에서의 이슈는 패널 지속에 따른 “표본 마모”로 그에 대한 대책이 중요한데, 호주의 LSAY는 패널 가중치 조정의 좋은 사례라 판단됨
- 이에 더하여, 우리나라도 호주의 LSAY와 같은 조사를 실시한다면 앞으로 더욱 열악해질 청년취업 및 고등교육 현장에 대한 정책 결정에 있어 기초자료로 사용될 수 있을 것이라 생각됨
- 현재 교육과정연구원에서 PISA 실시 중

□ A Generation stand-alone weights for a sub-population with time-varying characteristics: a SOEP case study

Dr Mathis Schroder(DIW, Berlin), Mr Rainer Siegers(DIW, Berlin)

○ 연구배경

- 독일 가족정책 결정의 수단으로 사용하기 위해 FiD(Family in Deutschland) 조사를 기획하였으며, 본조사는 세 개의 표본집단을 가짐
- 2007년부터 2010년 사이에 태어난 아이가 있는 가족을 목표모집단으로 하는 ①코호트, 다음의 요건 중 하나이상을 충족하는 가족을 목표모집단으로 하는 ②2010년, ③2011년 표본

- 편부 또는 편모가족
- 아이가 둘 이상인 가족
- 저소득 가족(2010년도만)

- SOEP* 가중치 작성법을 적용하여 SOEP-FiD 통합 가중치 작성
- * 독일에서 실시하고 있는 패널조사인 Socio-Economic Panel의 약자

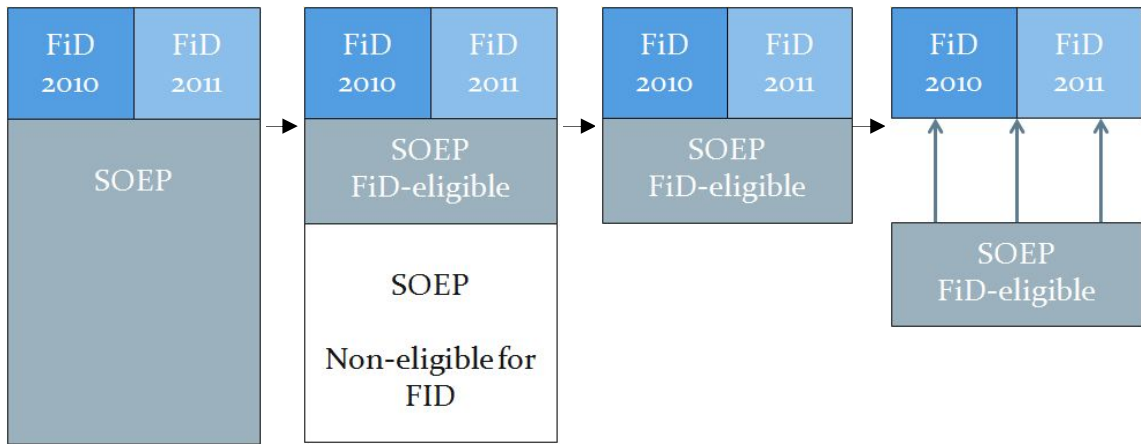
○ 연구내용

- Fid Wave2를 위한 Cross-Sectional 가중치 작성
- 목적: 관계부처의 요구에 따라 특정 단면의 FiD자료만을 분석할 수 있는 자립적 가중치 작성이 필요
- 문제점: SOEP의 기본 가중치작성법을 적용할 경우 2011년도 표본집단을 2010년 표본집단과 통합하는 것은 불가능

- 2010과 2011의 모집단 성격이 다름
- Mikrozensus*를 통해 보정을 실시할수 없음(SOEP 지침상 Mikrozensus를 통해 보정을 하며, 본 자료는 2010년과 2011년의 인구주택에 대한 동적정보(dynamic information)가 없음)
- * Mikrozensus : 지역적인 정보를 조사하는 자료로, 가구의 크기, 가구주의 주거형태 등의 자료를 제공
- 2011년도에 추출된 표본에 대한 2010년 정보를 알수 없음

○ 연구결과

- SOEP-FID 통합 설계 가중치를 가져온 다음 SOEP조사에서 FiD 조사에 적당한 가구를 선별, 그 가구의 가중치 합을 FiD조사 가구에 재분배



· SOEP조사에서 FiD 조사에 적당한 가구를 선별하는 방법

- 일시적으로 탈락하였거나 새로 뽑힌 표본을 이용
- 2010년 조사에 참여하지 않은 가구의 2010년 조사 적격성을 추정
- 안정적인 코호트를 이용
- 로지스틱모형에 의해 추정되고 예측확률이 주어진 상황에 배정된 선발 속성을 이용

○ 총평

- 가중치 재분배를 위한 가구 선별법과 재분배 방법 부분은 연구자의 추가 정보가 필요할 것으로 생각됨

조사표 설계1	Do pretesting methods identify 'real' problems and help us develop 'better' questions?
----------------	--

□ **A mixed methods approach to investigate how cognitive interviewing findings can improve the validity of survey estimates?**

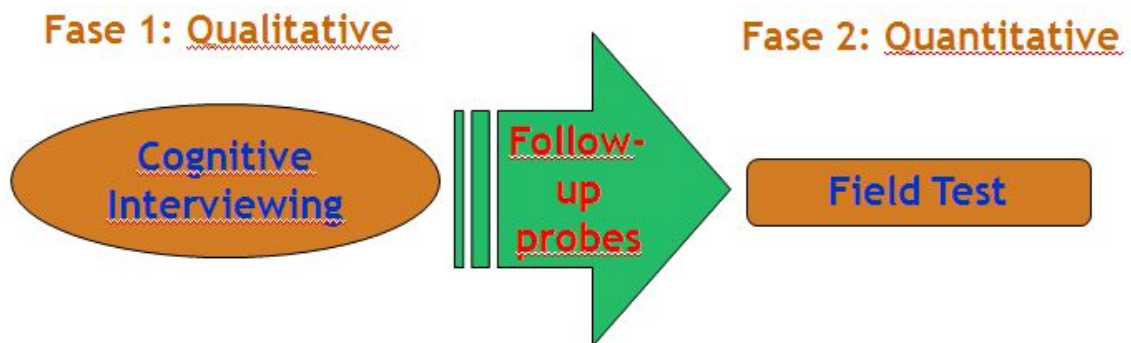
Dr Jose-Luis Padilla (University of Granada) et al.

○ 연구질문

- 인지면접 결과가 조사표 내 ‘진짜 문제’를 얼마나 발견할 수 있는가?
- 인지면접 결과가 서베이 추정치를 향상시키는가?
- 인지면접에 대한 두 가지 시각: Search for problems(Willis, 2005) vs. Interpretative(Miller, 2011)

○ 연구대상 및 연구방법

- (연구대상) 워싱턴 그룹과 UNESCAP 주최로 센서스와 비교가 가능한 장애인 관련 설문 모듈을 개발 중이며, 그 중 '고통(pain)'을 측정하는 두 가지 (빈도 및 심각성) 항목 검토
 - (연구목적) 대상 질문이 좀 더 타당성 있는 데이터를 제공하고 응답자간 비교가능성을 제시할 수 있는 근거를 수집하기 위함
 - (연구방법) 1단계 인지면접과 2단계 현장조사 실시
- 현장조사의 경우 2010 US HIS 본 조사 후 추가적인 프로빙 질문을 통해 고통을 측정: 인지면접 분석 결과 발견된 고통의 속성을 측정하기 위한 질문으로 구성



○ 분석결과

- 인지면접: 20명의 참가자는 고통의 속성(attributes)을 다르게 응답하는 것으로 나타났음
- 현장조사: 본 조사의 고통 항목을 종속변수로, 추적질문의 속성 항목을 독립변수로 하여 로지스틱 회귀분석 실시⇒ 이 속성들이 빈도와 심각성을 설명하는지 모델을 구성하고 설명력을 비교
- 그 결과, 추적질문에서 제시한 일부 항목은 고통에 유의미한 영향을 미치지 않는 것으로 진단

○ 시사점

- 양적인 분석과 함께 제시된 인지면접 결과는 서베이 추정치 해석과정에서 타당한 근거를 제공해 줄 수 있음

- 인지면접 결과가 현장조사를 통한 양적인 분석과 연결될 수 있음을 보여주는 추가적인 연구 필요

□ Comparing Questionnaire Pretesting Methods for the 2015 Population and Housing Census of Korea

Youngshil Park(Statistics Korea)

○ 연구질문

- 세 가지 조사표 평가 방법*이 질문의 문제점을 식별하는데 동일하게 유용한가?

* 인지면접 vs 조사원보고 vs 반분조사

○ 연구대상 및 연구방법

- (대상항목) 혼인상태, 가구구분 등 센서스 핵심항목 10개
- (연구방법) 3단계에 걸친 조사표 평가: 인지면접 ⇒ 반분조사 ⇒ 조사원 보고
- (인지면접) 총 29명에 대해 2라운드에 걸쳐 표준화된 지침서를 통한 인지면접 실시 후 새로운 버전 조사표 제안
- (반분조사) 기존 버전(2,331가구)과 새로운 버전(2,322)에 동일한 조사절차를 거쳐 자료를 수집
- (조사원보고) 조사완료 직후 질문읽기 난이도, 응답이해 난이도, 개념이해 불일치도에 대해 조사원이 각각 기존버전과 새로운 버전에 대해 평가하도록 함

○ 분석결과 (예시: 가구구분 항목)

- (인지면접) 기존 버전의 가족(family) 용어가 조사상의 정의와 응답자가 이해하는 것과 차이가 있는 것으로 나타났으며 이 차이를 줄이기 위해 새로운 버전에서 친인척을 포함한 가족이라고 수정하였음
- (반분조사) 무응답과 응답분포에서는 차이가 없는 가운데 새로운 버전에서 응답오류가 통계적으로 유의미하게 감소한 것으

로 나타났음

- (조사원보고) 이 항목의 응답 난이도를 조사원이 평가한 결과 오히려 새로운 버전을 조사한 조사원이 기존 버전에 비해 응답 난이도를 약간 높게 평가(1.00 vs 1.07)하였으나 그 차이는 유의미한 수준은 아님

Control (2010 Census)	Test (based on Cognitive Interviewing)
Which among 1 ~ 5 best describes this household?	Which among 1 ~ 5 best describes this household?
1. One-family household	1. One-family household including kinship
2. Household consisting of a family and a non-family member(s)	2. Household consisting of a family including kinship and no blood ties member(s)
3. One person household	3. One person household
4. Household of 5 or less persons who have no blood ties	4. Household of 5 or less persons who have no blood ties
5. Household of 6 or more persons who have no blood ties	5. Household of 6 or more persons who have no blood ties

○ 결론

- 잠정적으로, 인지면접과 반분조사에서는 ‘응답자 관점’에서 유사한 결과를 보여주는 것으로 결론
.인지면접은 질문이해와 관련하여 숨겨진 문제를 발견하는데 적절한 방법이며,
.반분조사는 그 실험실 결과를 타당화할 수 있는 방법임
- 조사원 보고는 기존버전과 새로운버전의 항목을 비교하는데 적절하지는 않으나, 조사표 내에서 서로 다른 항목간의 난이도 등을 평가하는데 유용할 것으로 보임

[부록: 발표자료 첨부]

□ Horses for Courses: Why different QT methods uncover different findings and implications for selecting methods

Ms Jo Dardenne(NatCen Social Research)

○ 연구질문

- 서로 다른 조사표평가(Questionnaire Test, QT) 방법이 서로 다른 문제를 진단하는가?
- 서로 다른 QT 방법은 단일의 연구 내에서 서로 다른 결과를 도출하는가? 그렇다면 왜 이런 결과가 발생하는가?
- QT 방법을 선택하는 데 있어서 중요한 점은 무엇인가?

○ 연구대상 및 연구방법

- 메타 분석: 3개 이상의 QT 방법을 활용한 7개의 연구
- 사용된 QT 방법
 - .질적: 포커스 그룹, 인지면접, 조사원피드백, 전문가리뷰
 - .양적: 시험조사자료분석, 이차분석

○ 분석결과

- 조사표 설계 단계에 따라서 목적과 그 방법이 달라짐
- 질적 평가 방법
 - .질적 방법 내에서 분석결과의 오버랩 : 단어의 단순화, 불필요한 말의 제거 등
 - .다른 방법에 비해 인지면접에서 가장 많은 워딩 문제가 진단되었으며, 포커스그룹에서 워딩의 문제점은 잘 발견되지 않음
 - .그러나, 인지면접에서의 willingness가 시험조사에서의 willingness와 항상 일치하는 것은 아님 (샘플링, 맥락 혹은 조사원 효과?)
- 양적 평가 방법
 - .양적 평가 방법은 응답자의 willingness를 평가하는데 좀 더 유용한 방법

.이슈는 진단할 수 있으나 왜 그 이슈가 발생하는지에 대해서는 명확한 근거를 제시할 수 없을 수도 있음

- 조사원 피드백 방법은 불필요한가?

.조사원 피드백과 인지면접 결과와 높은 중복 진단율: 예컨대 응답자가 명확화를 요구하는 부분, 명확화를 요구하는 이유, 응답자가 사용하는 용어, 무응답 이유 등이 유사하게 문제점으로 도출되었음

○ 결론

- 조사표 평가 방법의 조합은 연구 목적과 조사표 설계 진행단계에 따라서 달라져야 하되,
- 각 방법의 장단점에 대한 직접적 비교는 현재 이 연구에서는 어려우며,
- 향후 동일한 질문에 대해 서로 다른 방법을 활용한 실험을 통한 비교 연구가 필요함

□ 총평

- 이전에 참가했던 조사방법론 컨퍼런스에 비해 최근 조사표 평가 방법 영역의 연구들은 어느 한 방법에 의한 조사표 평가 보다는 다양한 방법을 복합적으로 적용하여 결과를 도출하고 그 타당성을 입증하는 방향의 연구가 증가하고 있음
 - 특히, 질적인 방법과 함께 양적인 분석을 종합하여 평가해야 한다는 주장이 증가하고 있음
- 이와 관련하여, 출장자가 발표한 연구는 단일의 조사표에 대해서 다양한 방법을 적용하였으며 그 방법들을 비교하고 있다는 점에서 흥미롭다는 평가를 받았음

□ **Translatability Assessment of Draft Questionnaire Items in Multilingual Surveys?**

Mr Steve Dept(cApStAn Linguistic Quality Control)

- 조사 항목 번역에 대한 Harkness의 ‘전체론적 관점’
 - (다문화 조사 cross-cultural survey) 다양한 언어에 적용되기에 앞서 조사항목을 정성스럽게 만드는 것이 얼마나 중요한가에 대한 역설
- 번역 가능성(translatability)에 대한 평가
 - item writer의 중요성 : 원본(source) 조사표 만들기 ⇒ 다른 언어로의 번역(translation) 및 적용(adaptation)에 적합하게 만들기
 - 실험설계 : 숙련된 언어학자에게 새로운 항목에 대한 초안 버전을 제시한 후 평가하도록 함 (평가 시 코드를 활용하여 문제유형을 표준화하였음) ⇒ 문제를 수정한 대안적인 워딩 제안 ⇒ 씨니어 언어학자의 분석 및 피드백 ⇒ 번역가능성 보고서 작성
 - 문제유형 분류 코드
 - ① Straightforward(문제없음) ⑦ Double-barrelled
 - ② Known difficulty ⑧ Agreement issue
 - ③ Potential ambiguous ⑨ Consistency
 - ④ Unnecessarily complex ⑩ Redundancy
 - ⑤ Requires review ⑪ Possible addition
 - ⑥ Potential cultural issues ⑫ Logical problem
- 번역가능성 보고서 작성 예시
 - 코드 3번(Potential ambiguous)

English	Translatability Evaluation	LINGUIST'S COMMENT	SUGGESTIONS (including suggestions for alternative wording)
I can solve my problems on digital devices myself.	POTENTIALLY AMBIGUOUS	The sentence seems straightforward, but in fact it is unclear whether the respondent knows how to use digital devices to solve his/her problems, or whether the respondent knows how to solve problems related to the digital devices s/he is using.	When I have problems with digital devices, I can solve them myself. OR I can use digital devices to solve my problems by myself.

- 코드 7번(Double-barrelled)

English	Translatability Evaluation	LINGUIST'S COMMENT	SUGGESTIONS (including suggestions for alternative wording)
My child's school provides an inviting and accepting atmosphere for parents to get involved.	DOUBLE-BARRELLED	"inviting and accepting" are overlapping adjectives that point to 'tolerance' (interpreted differently in multicultural versus monocultural environments), whereas 'inviting' by itself is overarching.	My child's school provides an inviting atmosphere for parents to get involved.

○ 결론

- 언어학자의 전문성과 조사항목 개발자와의 융합 필요
- 소스(source) 언어로 쓰여진 초벌 항목과 그 항목의 실제 버전 사이의 적합한 번역과 적용 사이의 갭을 연결하는 노력 필요
- 이 방법은 비교적 낮은 비용을 통해서도 이해도 높은 다큐멘테이션 가능

□ Can we survey migrants by questionnaires developed for the population of the respective source countries?

Mr Patrick Brzoska(University of Bielefeld) et al.

○ 연구배경

- EU 국가 간 이민자의 증가와 이들에 대한 자기보고식 건강 정보 필요성이 급증하나 이민자에 대한 연구는 다음과 같은 한계로 어려움 : 이민자 정의.접촉.샘플링의 어려움
- 간과되었던 점이, 많은 이민자가 이민국 언어에 능숙하지 못하다는 점이며, 이에 따라서 모국어 버전의 조사표가 필요하다는 것이 이슈로 부각되고 있음
- 일반적으로 source population을 위한 조사표가 이민자를 조사

- 하는 데에도 그대로 적용되는 경향이 있음: 이는 두 집단이 동일하며 그 도구는 서로 상호활용가능하다는 가정
- 그러나, 여기에서는 source population과 이민자 사이의 잠재적인 차이는 고려되지 않았다는 점이 문제 발생 소지

○ 연구질문

- 조사표가 이민자와 그들의 source population 사이에 동일하게 작동하는가?

○ 연구내용 및 방법

- 터키에 사는 터키인과 독일에 사는 터키인에 대한 건강조사 항목 중 ‘질병’ 항목에 대한 두 집단 간 비교
- 집단 간 비교를 위해 다음과 같은 측면에서 동위성이 전제되어야 함

개념적 동위성	의미상 동위성	운영상 동위성	측정상 동위성
↓	↓	↓	↓
그룹 간 유사 개념	그룹 간 번역된 언어 의미	그룹 간 같은 연구 방법 (조사표 포맷, 준거 기간, 측정 유형, 조사방법 등)	그룹 간 관찰 변수 간 유사 통계치 (요인평균, 분산 등)

- .개념적 동위성 : 터키에 사는 터키인 24명과 독일에 사는 터키 이민자 20명 통한 질적 연구
- .의미상 동위성 : 터키와 독일에 사는 전문가 8명, 독일에 사는 터키 이민자 58명 통한 질적 연구(인지면접 등)
- .운영상 동위성 : 문헌 연구를 통해서 확보
- .측정상 동위성 : 터키에 사는 터키인 302명 독일에 사는 터키 이민자 299명으로부터 양적 자료 수집 후 확인적 요인분석 실시

○ 연구결과

- 원래 조사표의 구조적 타당성이 이민자 집단에서 감소하는 것

으로 나타났음: 현 연구에서 터키에 사는 터키인과 터키 이민자 사이의 질병에 대한 개념을 등가적으로 비교하는 것은 제한적임을 의미함

- 그룹 간 비교 타당성을 위해서는 모든 조건에서의 등위성이 전제되어야 하며, 그 등위성을 확보하기 위해 추가적인 테스트가 더 필요

□ 총평

- 유럽 연합 국가 간 비교 조사가 많이 이루어지는 상황적인 맥락상 번역에 관한 연구발표가 다수 이루어졌다는 점이 특징적임
- 우리나라의 경우에도 이민자가 증가하고 있으며, 특정 조사의 경우 그 조사표의 원본이 외국어인 점을 감안한다면 조사표에서의 번역의 문제 또한 향후 중요한 이슈로 제기되어야 할 부분으로 판단됨

조사표 설계3

Use of Eyetracking in Survey Research

□ Does left still feel right? The optional position of answer boxes in Web surveys-revisited?

Dr Timo Lenzner (GESIS-IEIBNIZ Institute for the Social Science) et al.

○ 연구질문

- 웹 서베이에서 폐쇄형 질문에 대한 응답박스의 위치는 오른쪽과 왼쪽 어느 곳에 배치해야 할까?

Wählen Sie drei Eigenschaften aus dieser Liste aus, die ein Kind Ihrer Meinung nach vor allem haben sollte. Dass es...

- ... gute Umgangsformen hat
- ... wirklich etwas leisten will
- ... ehrlich ist
- ... sauber und ordentlich ist
- ... gesunden Menschenverstand und eine gute Urteilsfähigkeit besitzt
- ... Selbstbeherrschung besitzt
- ... sich als Junge wie ein Junge und als Mädchen wie ein Mädchen benimmt

Wählen Sie drei Eigenschaften aus dieser Liste aus, die ein Kind Ihrer Meinung nach vor allem haben sollte. Dass es...

- ... gute Umgangsformen hat
- ... wirklich etwas leisten will
- ... ehrlich ist
- ... sauber und ordentlich ist
- ... gesunden Menschenverstand und eine gute Urteilsfähigkeit besitzt
- ... Selbstbeherrschung besitzt
- ... sich als Junge wie ein Junge und als Mädchen wie ein Mädchen benimmt

○ 문헌 연구에서는 오른쪽에 배치하는 것이 논리적으로 좀더 자연스럽다는 것(Jenkins & Dillman, 1995)과 왼쪽에 배치하는 것이 응답범주와 응답박스간의 거리를 감소시킨다(Couper, 2008)는 두 가지 의견이 있으나, 대부분 응답자들은 그 위치를 신경쓰지 않으며 두 레이아웃 중 명확한 선호는 없다는 결론

○ 시선추적실험을 통해 응답자가 응답선택과정에서 소요되는 인지적인 노력이 어느 정도 소요되는지를 평가

○ 실험결과1

- 대부분의 응답자가 응답에 앞서 여러번 응답범주를 읽으며, 응답보조도구로써 마우스 포인터를 사용하지 않음
- 응답박스가 왼쪽에 위치한 경우, response latencies가 짧으며, 응답범주에 머무는 시간이 짧고, 응답범주와 응답박스간 gaze switch가 더 적음
- 한계: 표본이 교육수준이 높은 젊은이로 한정되었으며, 응답범주와 응답박스간의 거리를 조정한 제3의 버전의 경우 결과가 달라질 수 있음

○ 실험결과2

모형1	모형2	모형3
<p>Was glauben Sie, wie wird Ihre <i>eigene</i> wirtschaftliche Lage in einem Jahr sein?</p> <p><input type="radio"/> Wesentlich besser</p> <p><input type="radio"/> Etwas besser</p> <p><input type="radio"/> Unverändert</p> <p><input type="radio"/> Etwas schlechter</p> <p><input type="radio"/> Wesentlich schlechter</p>	<p>Was glauben Sie, wie wird Ihre <i>eigene</i> wirtschaftliche Lage in einem Jahr sein?</p> <p>Wesentlich besser <input type="radio"/></p> <p>Etwas besser <input type="radio"/></p> <p>Unverändert <input type="radio"/></p> <p>Etwas schlechter <input type="radio"/></p> <p>Wesentlich schlechter <input type="radio"/></p>	<p>Was glauben Sie, wie wird Ihre <i>eigene</i> wirtschaftliche Lage in einem Jahr sein?</p> <p>Wesentlich besser <input type="radio"/></p> <p>Etwas besser <input type="radio"/></p> <p>Unverändert <input type="radio"/></p> <p>Etwas schlechter <input type="radio"/></p> <p>Wesentlich schlechter <input type="radio"/></p>

- 대부분의 응답자가 응답범주를 개략적으로 스캐닝하기에 앞서 응답범주를 선택하고, 응답을 하며, 응답보조도구로 마우스포인터를 사용하지 않음
- 응답박스가 왼쪽에 위치한 경우, response latencies가 짧으며, 응답범주에 머무는 시간이 짧고, 응답범주와 응답박스간 gaze switch가 더 적음

○ 결론

- 실험1의 결과가 반복되었으며, 최종적으로 인지적인 능력이 가장 적게 들어간 경우는 응답박스가 왼쪽에 있는 경우이며, 오른쪽에 있는 경우가 가장 큰 것으로 나타났음

□ **Combining eye tracking and cognitive interviewing to pretest survey questions?**

Cornelia Neuert and Timo Lenzner (GESIS-IEIBNIZ Institute for the Social Science)

○ 연구질문

- 시선추적실험이 인지면접을 보완해주는 효과적인 도구인가?

○ 분석결과

- 식별된 문제 질문 수 (시선추적 25, 인지면접 20)

Eye tracking identifies a problematic question	Cognitive interview identifies a problematic question	Problematic questions
Yes	Yes	18
Yes	No	7
No	Yes	2

- 식별된 문제 유형 및 수

Number of unique problems identified	CI	EYE
Overall	164	166
Type of problems		
Comprehension	84,1 % (138)	83,7 % (139)
Retrieval	1,8 % (3)	1,2 % (2)
Judgment	2,4 % (4)	3,0 % (5)
Response Selection	11,6 % (19)	11,5 % (19)
Questionnaire Navigation	0,0 % (--)	0,6 % (1)

○ 결론

- 인지면접과 시선추적실험 모두 이해문제를 진단하는데 효과적인 것으로 나타났으며,
- 공통으로 발견하는 문제 수 못지않게 두 방법이 서로 다른 문

제를 진단해내는 것으로 나타났는데 특히 인지면접보다 시선추적실험에서 진단해 낸 문제수가 더 많았음

- 이를 통해 시선추적실험이 응답자 행동에 대한 통찰력을 제공해주며, 또한, 의식되지 않는 문제 유형을 탐지할 수 있는데 매우 유용한 것으로 판단됨

□ 총평

- 조사표 평가 영역 중 시선추적연구는 다른 항목에 비해 많은 비중있게 다루어지지 않는 않았으나, 시선추적실험이 질적인 조사표 평가 방법을 보완해주는 것으로 많은 사람들의 관심을 받았음

세션명

Mode Effect and the Cross-Cultural Survey

□ Mode Differences in Socially Desirable Responses to Sensitive Questions

Michael W. Traugott (University of Michigan) et al.

- 서베이 응답과정에서 방법효과(mode effect) 관련 이슈: 다음과 같은 효과가 혼재되어 있음
 - 선택효과(selection effect): 누가 응답하느냐?
 - 방법효과(mode effect): 방법에 따라서 달라지는 응답패턴
 - 측정효과(measurement effect): 자료수집방법에 따른 질문형태에 의한 효과
- 연구설계
 - 대상 항목인 출생지에 대한 태도를 묻고 ⇒ 민감한 혹은 덜 민감한 질문으로 확장하여 조사표 설계
 - .출생지 질문은 “오바마가 미국에서 태어났다고 생각하는지”를 묻는 기본형에 조사기관에 따라서 질문형태 및 질문방법, 응답범주를 약간씩 변형
 - 다양한 기관에서 조사: CNN, Gallup 등 8개 조직

- 자료수집방법은 RDD 전화조사, 웹조사, IVR
- Pulse Opinion Research의 웹조사와 IVR 패널 표본특성 차이

○ 분석결과

- (가설1) 표본특성과 방법이 민감한 질문에 대한 바람직한 응답에 영향을 미칠 것이다 ⇒ (결과) 오바마가 ‘확실히’ 미국에서 태어났다고 생각한다는 응답이 조사기관에 따라서 42.9% ~ 63.0% 까지 차이를 보였음
 - (가설2) 민감한 질문에 대한 일반 응답(요약치) 수준은 자료수집 방법에 따라 다양할 것이다 ⇒ (결과) 모든 질문에 100% 응답한 비율은 기관에 따라서 93.9% ~ 46.8%로 차이가 큼
 - (가설3) 구체적 질문에 대한 응답 수준은 자료수집 방법에 따라 다양할 것이다 ⇒ (결과) 질문 유형에 따른 무응답 수준이 차이가 있음
- .기관에 따라서 각각의 질문에 대한 무응답 비율: 민감하지 않은 질문은 0%~0.2%, 다소 민감한 질문은 1.6% ~ 22.9%, 민감한 질문은 5.3% ~ 28.6% (이러한 비율은 인종에 따라서 다양함)

○ 결론

- 방법효과를 질문 형태와 응답범주 효과와 분리하는 것이 쉽지 않음
- 응답자와 표본 특성(selection effect)은 응답 패턴에 영향을 미칠 수 있음
- 좀 더 민감한 질문이 덜 민감한 질문에 비해 무응답이 더 높으며 이 효과는 인종 특성에 따라서 다양함
- 이러한 민감한 질문이 그 주제에 따라서 다양한지에 대한 추가적인 연구가 필요함

□ An Exploratory Study of the Cross-cultural Equivalence of Wellbeing of Measures in the Gallup World Poll

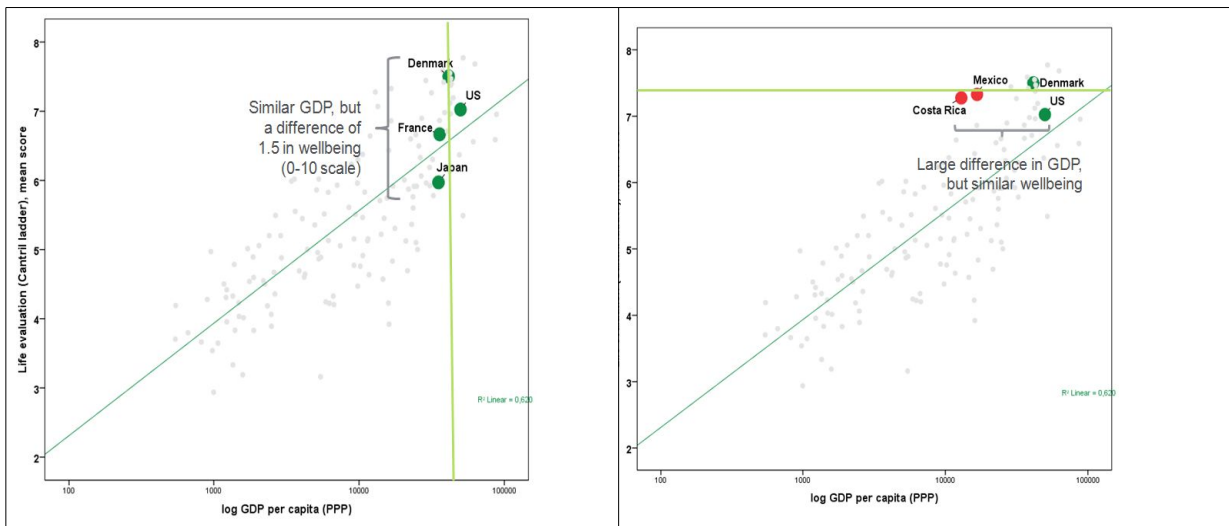
Dr Femke De Keulenaer, Rober Manchin(Gallup Europe)

○ 연구배경

- GDP와 국민의 행복도는 비례하는가?
- 대체로 비례하지만 몇몇 국가에서는 예측보다 더 높은(혹은 더 낮은) 행복도가 나타나는데 그 이유는 무엇일까?

○ 연구주제

- “문화적 편향” 문제 발생: 일인당 유사 GDP 국가의 “삶의 질” 차이와 비슷한 행복도를 가진 국가의 GDP 차이를 보여줌



- 문화의 역할: 국가 간 차이는 다음에 기인함

① 측정되지 않은 생활환경(예: 사회적 맥락)

② 생활환경을 평가하는 방식의 차이

→ 문화의 영향력: 사람들이 삶에 대해 실제로 어떻게 느끼는지에 영향

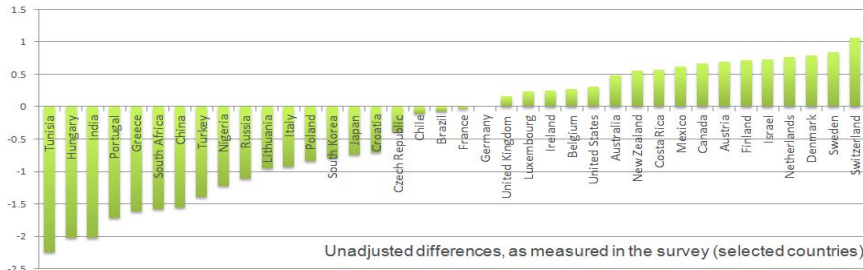
③ 측정되지 않은 국가 환경(예: 도시 사회)

④ 응답 편향(예: 과한 응답, 대변효과 등)

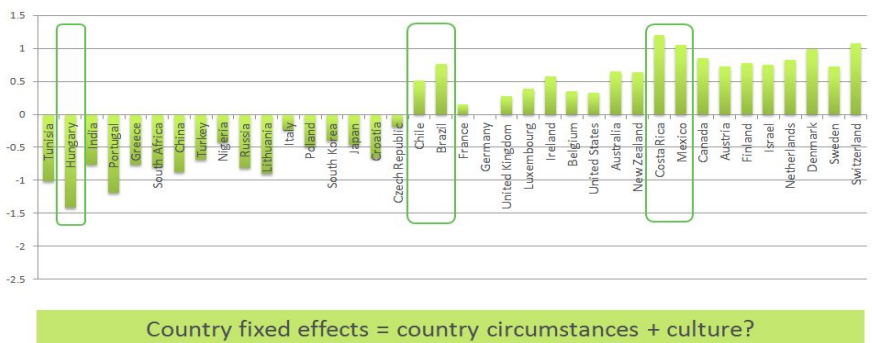
→ 문화적 편향: 자신이 느끼는 것이 아닌 국민이 말하는 것에 영향

- 국가 간 웰빙의 차이와 국가 고정 효과 모형 비교

- 독일의 점수(6.71)를 기준으로 한 실제 조사 결과(조정 없음)

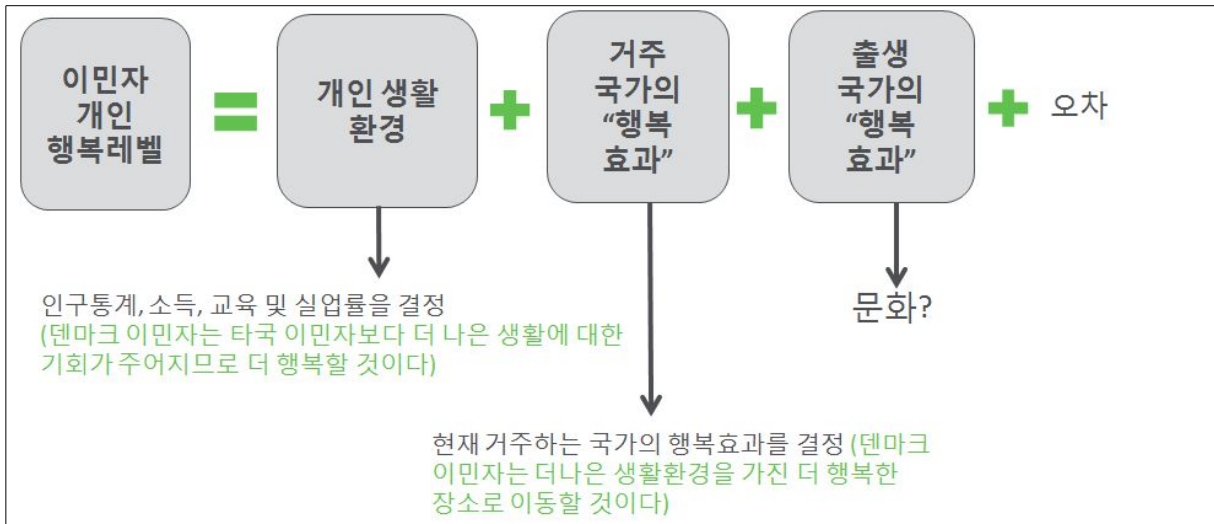


- 고정 효과 모형에 따른 국가 간 삶의 질 비교(인구통계, 교육, 소득, 실업률 지표를 사용하여 회귀추정)



○ 분석사례

- 이민자를 통하여 환경과 문화를 분리하여 분석



- GWP(Gallup World Poll) 2009 to 2012의 이민자 26,239명(140개국 이상에서 140개국 이상의 나라로 이민을 한)의 자료를 토대로 분석

- 최초의 국가(거주/출생) 행복 영향은 국가 고정 영향 모델의

- 계수를 사용하여 양을 정하고 두 개의 분리된 변수로 개별 이민자 자료에 적용
- 이민자 삶의 질 예측: 개인 삶의 질에 대한 18.5%의 총분산은 다음으로 설명됨
 - .개인 생활 환경에서 14.1%
 - .국가 환경에서 3.9%
 - .출생 국가(문화를 포함하는)에서 0.5%

Life evaluation (Cantril ladder, 0-10 scale)

	Step 1	Step 2	Step 3
<i>Intercept</i>	4.992	5.084	5.236
Men	.068	.081	.050
<25 year-olds	.672	.663	.705
25-44 year-olds	.236	.225	.286
65+ year-olds	.113	.103	.125
Married/cohabiting	.285	.246	.271
Child under 15 at home	.171	.117	.117
Ln household income	.467	.410	.395
Secondary education	.404	.348	.342
Tertiary education	.772	.691	.689
Unemployed	-.533	-.558	-.580
Fixed effect coefficient of RESIDENCE country		.467	.437
Fixed effect coefficient of BIRTH country			.229
Observations	26,239	26,239	26,239
Adjusted R-square	.141	.180	.185

- 자료이용의 한계
 - .현 자료에서 이민자라고 응답한 사람의 검증이 어려움
 - .편향되지 않은 자연적 실험이라고 볼 수 없음
 - .실제로 국가와 문화를 동일하게 놓을 수는 없음
 - .문화적 편향과 문화의 영향을 분리할 수 없음

□ Comparing subjective well-being across countries: Measurement equivalence and the question of multiple dimensions

Dr Joakim Kulin(Department of Sociology, Umea University) et al.

○ 연구문제

- 감성적 웰빙(EWB-Emotional Well-Being) 패턴은 삶의 만족 (LS-Life Satisfaction) 패턴과 차이가 있을까?
- 일반적으로 사용되는 1문항 행복도 혹은 삶의 만족도 조사가 국가 간 주관적 웰빙(SWB-Subjective Well-Being)을 균등하게

평가할 수 있는가?

○ 연구내용

- 자료

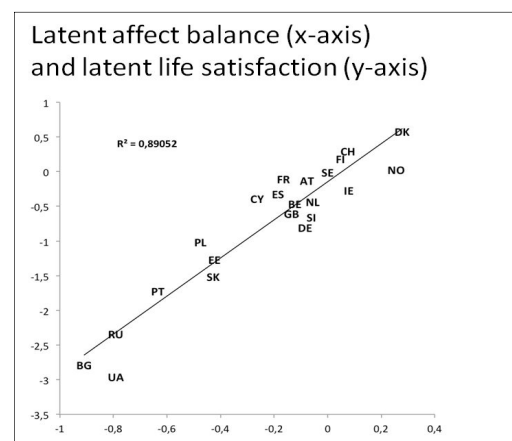
- 유럽 사회 조사 2006(European Social Survey)
 - 23개 국가, 국가 당 약 2,000명의 응답자
 - 측정 결과
 - Life satisfaction: 3 항목
 - . “모든 것을 고려했을 때, 현재 당신은 얼마나 당신의 삶에 만족하시나요?”
 - . “당신은 당신이 지속해온 삶의 방식에 얼마나 만족하십니까?”
 - . “나의 일생은 내가 원했던 것에 가깝다”
 - Emotional well-being: 6 항목
 - 당신이 가질만한 감정의 목록을 읽어드리겠습니다... ...부디 지난 일주일간 당신이 느꼈던 감정을 말씀해 주세요.
 - . 긍정적 감정: “행복한”, “활기찬”, “편안한”
 - . 부정적 감정: “슬픈”, “화나는”, “무기력한”

- 분석방법

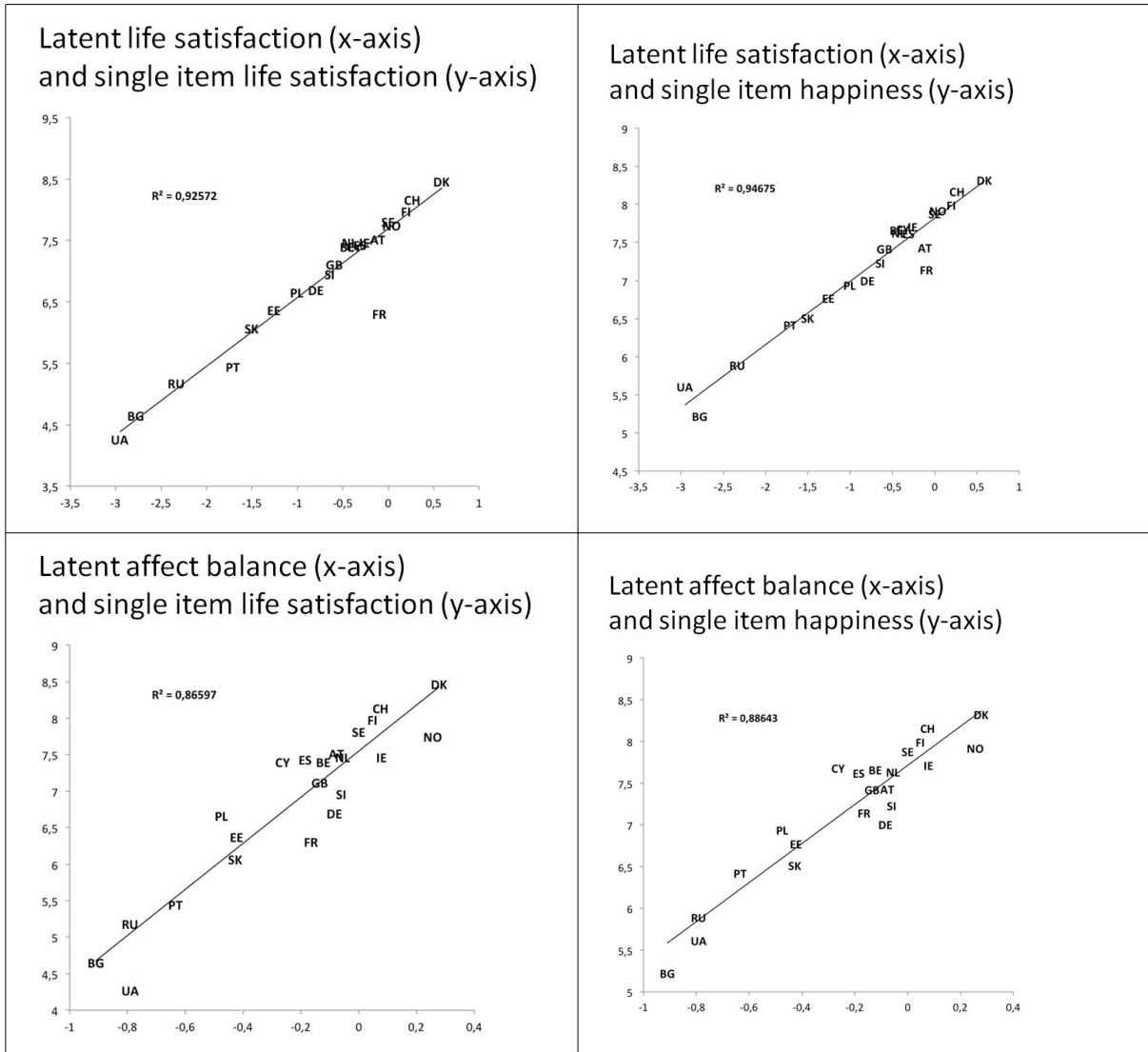
- 다 집단 확정 지표 분석(Multi-group confirmatory factor analysis)
 - 신빙성과 대조능력을 검증
 - 측정오차를 다룸
 - 집단 간 측정의 불변성을 시험
 - 집단(국가들)의 잠재 평균(latent mean) 추정
 - 집단(국가)간의 잠재 평균의 종대한 차이점 검증
 - 결과: scalar invariance(CFI>0.9)

- 분석결과1

- .여러 순위산정에서 LS와 EWB 결과에 대한 잠재적 평균(예: 두 개의 다른 개념을 측정)
- .EWB와 LS가 높은 상관을 보이지만, 동일하다고 볼 수 없음 (그림 →)



.LS의 잠재적 평균은 두 개의 단일문항 측정치와 강한 상관관계
 .EWB와 행복 문항과의 관계가 EWB와 LS 문항과의 관계보다
 강하지는 않은 것으로 판단됨 (그림↓ 참조)



- 분석결과2

.행복관련 문항은 감동요소를 제대로 잡아내지 못하는 것으로
 판단됨(실험전의 예측)

.LS관련 단일문항은 국가단위의 LS를 측정할 수 있는 것으로
 판단됨

.EWB는 제안된 잠재 구성법 혹은 다른 단일 문항 방법으로
 측정하는 것이 바람직해 보임

Affect balance calculation

Country	Positive Affect	Negative Affect	Affect balance
AT	0,076	0,154	-0,078
BE	0,072	0,194	-0,122
BG	-0,412	0,497	-0,909
CH	0,198	0,144	0,054
CY	-0,095	0,168	-0,263
DE	-0,063	0,021	-0,084
DK	0,176	-0,102	0,278
EE	-0,099	0,323	-0,422
ES	-0,033	0,152	-0,185
FI	-0,036	-0,084	0,048
FR	0,069	0,234	-0,165
GB	0,011	0,146	-0,135
IE	0,145	0,066	0,079
NL	0,107	0,161	-0,054
NO	0,147	-0,109	0,256
PL	-0,158	0,316	-0,474
PT	-0,208	0,426	-0,634
RU	-0,251	0,54	-0,791
SE	0	0	0
SI	0,082	0,143	-0,061
SK	-0,044	0,383	-0,427
UA	-0,163	0,626	-0,789

Latent and and global measures

Country	Affect balance	Life satisfaction	Happiness Single item	Life satisfaction Single item
AT	-0,078	-0,117	7,44	7,53
BE	-0,122	-0,45	7,67	7,41
BG	-0,909	-2,778	5,23	4,66
CH	0,054	0,195	8,07	8,03
CY	-0,263	-0,378	7,69	7,41
DE	-0,084	-0,796	7,01	6,71
DK	0,278	0,59	8,33	8,48
EE	-0,422	-1,257	6,78	6,38
ES	-0,185	-0,311	7,63	7,44
FI	0,048	0,197	8	7,99
FR	-0,165	-0,0954	7,15	6,32
GB	-0,135	-0,592	7,43	7,13
IE	0,079	-0,26	7,72	7,48
NL	-0,054	-0,422	7,64	7,48
NO	0,256	0,038	7,93	7,76
PL	-0,474	-1,003	6,95	6,67
PT	-0,634	-1,709	6,43	5,46
RU	-0,791	-2,332	5,9	5,19
SE	0	0	7,89	7,82
SI	-0,061	-0,646	7,24	6,97
SK	-0,427	-1,502	6,52	6,08
UA	-0,789	-2,95	5,62	4,28

Ⅲ 시 사 점

- 유럽조사연구학회는 유럽 뿐 아니라 미국 등에서 많은 연구자가 참여하는 학회로, 전 세계 각 국가의 조사방법론 관련 주요 이슈 및 연구 경험을 한 자리에서 공유할 수 있다는 점에서 의미있음
 - 동아시아 국가 참석률은 여전히 낮은 편인데, 조사방법 연구 관심이 매우 저조한 수준임을 간접적으로 보여주고 있음
- 선진 통계국의 면모를 구축하기 위해 향후 통계학회 뿐 아니라 유럽 및 미국 등의 조사연구학회에 대해 우리 통계청의 관심이 필요하다는 생각임
 - 조사연구학회는 통계 전공자 뿐 아니라 사회과학 전공자들도 쉽게 접근할 수 있는 연구 주제를 다양하게 다루고 있다는 점
 - 그리고 이러한 주제가 통계청 업무에서 주요한 부분(비표본 오차로써의 조사표, 조사원, 응답자 등) 중 하나라는 점에서,
 - 향후, 우리나라 통계청에서도 세계통계학회(ISI) 미팅 뿐 아니라 조사연구학회에 대해서도 많은 관심을 기울일 필요가 있음
- 이번 컨퍼런스에서 본 출장자들은 가중치 및 조사표 설계에 중점을 두고 참여하였음
 - 많은 세션이 동시에 열려 모든 관심 세션에 참석할 수 없다는 한계가 있었음
- 표본 가중치와 관련한 세션에 참석한 결과 다음의 동향을 파악할 수 있었음
 - 첫째, 무응답률이 매우 높은 유럽의 선진 사례 및 연구를 통하여 앞으로 다가올 미래 조사환경을 대비할 수 있는 기회마련

- 둘째, 패널조사를 이용하여 본래 목표 이외에 추가적인 통계 요구에 대응한 사례를 통하여 패널자료의 분석에 대한 새로운 시각을 갖게 되었음
 - 셋째, 국가 간 비교에 있어서 그동안 짐작만 하였던 부분을 통계적 분석을 통해 검증한 사례를 통하여, 미개척 분야라 할 수 있는 국가 간 비교통계의 활성화 가능성을 확인하였음
- 조사표 설계 및 평가 관련해서는 다음과 같은 동향을 파악할 수 있었음
- 첫째, 어느 하나의 방법론을 적용해서 조사표를 평가하던 것에서 벗어나 여러 가지 평가 방법을 동시에 적용하여 조사표 타당성을 높이는 방향으로 연구 동향이 전환되고 있으며,
 - 둘째, 인지면접 등과 같은 질적인 방법 뿐만 아니라 조사자료의 분석, 시선추적시스템 등을 활용한 객관적인 자료를 보완적으로 사용하여 조사표를 평가하고자 하는 노력들이 있음
 - 셋째, 조사표 평가 방법과 함께 번역의 문제가 조사표 설계와 관련된 또 다른 주요 이슈로 제기됨: 우리나라에도 이민자가 증가하고, 영어를 원본으로 하는 조사표 혹은 국가 간 비교를 위해 영어조사표를 제공하고 있는 점 등을 고려한다면 이 부분에 대한 관심을 기울일 필요가 있을 것으로 보임
- 이러한 동향 파악을 통해 향후 통계개발원 핵심 추진업무인 조사표 설계 분야의 연구과제 수행 시 이를 적용하고자 함
- 인구주택총조사 및 경제활동인구조사, 사회조사 등 조사표 개선 업무에 활용하고자 함
 - 특히 인총 조사표의 경우 금번 출장에서 발표한 내용을 더욱 발전시키기 위해 응답자 보고 및 시선추적연구 등 다양한 기법을 추가적으로 적용해 볼 계획임

IV 부록 : 발표자료

<p>Comparing Questionnaire Pretesting Methods for the 2015 Census of Korea</p> <p>Youngshil Park Statistical Research Institute, Statistics Korea European Survey Research Association, July 15-19, 2013</p>	<h3>Overview of Paper</h3> <ul style="list-style-type: none"> ▪ Introduction <ul style="list-style-type: none"> - Population and Housing Census of Korea - 2015 Long Form Questionnaire Redesign Plan ▪ Pretesting Methods <ul style="list-style-type: none"> - Cognitive Interviewing - Split Ballot Survey - Interviewer Debriefing ▪ Illustrative Analysis Results ▪ Conclusion <p>2</p>				
<h3>Introduction</h3> <ul style="list-style-type: none"> ▪ Population and Housing Census of Korea <ul style="list-style-type: none"> - Conducted by every 5 year based on household - Collected by paper & internet since 2010 - Two kinds of questionnaires <table border="1"> <tr> <td>~ 2010</td> <td>Short form: 90% Long form: 10%</td> <td>Short form: register based Long form: 20%</td> <td>2015~</td> </tr> </table> <p>3</p>	~ 2010	Short form: 90% Long form: 10%	Short form: register based Long form: 20%	2015~	<h3>Introduction</h3> <ul style="list-style-type: none"> ▪ The 2015 Long Form Questionnaire Redesign <ul style="list-style-type: none"> - Statistical Research Institute has been conducting pretesting for the redesign - Content test <ul style="list-style-type: none"> • 2012: core items, 2013: new items • Main methods: expert review, cognitive interviewing, split ballot survey, and interviewer & respondent debriefing - Design test <ul style="list-style-type: none"> • 2013: population section, 2014: internet survey • Main methods: usability test and field test <p>4</p>
~ 2010	Short form: 90% Long form: 10%	Short form: register based Long form: 20%	2015~		
<h3>Introduction</h3> <ul style="list-style-type: none"> ▪ Research Questions <ul style="list-style-type: none"> - Are these three pretesting methods* equally useful in identifying problems in the questionnaire? <p>* cognitive interviewing, split ballot survey, interviewer debriefing</p> <p>5</p>	<h3>Content Test</h3> <ul style="list-style-type: none"> ▪ Target Items <ul style="list-style-type: none"> - 10 items including marital status, residence one year ago, household type, total floor space and so on ▪ Procedures <ul style="list-style-type: none"> - Cognitive interviewing: 2012 Mar. ~ July - Proposed alternative questionnaire based on the results - Split ballot survey: 2012 Nov. - Interviewer debriefing: 2012 Nov. <p>6</p>				

Cognitive Interviewing

- **Purpose of Cognitive Interviewing**
 - (1st round) To evaluate R's understanding of concepts and terms
 - (2nd round) To review the effect of revised items
- **Developing a Protocol**
 - Expected problems related to questions
 - Used the retrospective probing, vignette, and paraphrasing techniques

7

Cognitive Interviewing

- **Recruiting**
 - Advertisement
 - Statistics Korea's homepage
 - Free local newspaper → Added this method for the 2nd round to recruit more diverse group
 - Selection
 - (1st round) 15 of 49 (2nd round) 14 of 69
 - Tried to evenly split into major demographic characteristics
 - Age, sex, marital status, housing type etc.
 - Rs were paid by \$30 for their participation

8

Cognitive Interviewing

- **Distribution of R's Characteristics**

	Sex		Age				Education		Housing Owned	
	M	F	20'S	30'S	40'S	50'S+	UNIV.-	UNIV.+	Rented	Owned
1R (15)	4	11	3	3	6	3	5	10	11	4
2R (14)	7	7	2	2	4	6	8	6	6	8

- **Schedule**
 - (1st round) : April 12~17 (4 days), Daejeon city
 - (2nd round) : June 12~17 (4days), Daejeon and Seoul cities

9

Cognitive Interviewing

- **Conduct of Cognitive Interviewing**
 - Interviewed each respondent in the lab
 - Three researchers played a role of interviewers
 - Procedures
 - The Rs first completed a test questionnaire
 - Afterwards, the interviewer asked series of probing questions
 - tape-recorded

10

Cognitive Interviewing

- **Analysis of Cognitive Interviewing**
 - Reviewed and documented the individual interview
 - Compiled results across interviews
 - From a qualitative point of view
 - What problems were
 - Whether problems were similar across interviews
 - From a quantitative point of view
 - How often problems emerged

11

Split Ballot Survey

- **Design**
 - The two cities were selected for the target area
 - 83 enumeration districts (20%) were systematically sampled from the population of 415 Eds.

Compared two versions

- **Control version:** replicated the 2010 census question
- **Test version:** proposed an alternative questionnaire based on the cognitive interviewing results

12

Split Ballot Survey

- **Design (Con.)**
 - Half of the interviewers surveyed with a control version and the other did with a test version
 - Averagely 2 or 3 Eds were assigned to interviewers

	Control Group	Test Group
N. Of Eds.	40	43
N. Of HHs.	2331	2322
Response Rate	97.2%	96.8%

13

Split Ballot Survey

- **Data Collection**
 - Used a similar data collection methodology as the previous Census

Internet Survey
Oct.28 ~ Nov.4

→

Face to Face Survey
Nov.5 ~ Nov.19

- Initially, voluntary internet survey was used from sampled households
- Face to face interviewing was followed up with internet nonrespondents

14

Interviewer Debriefing

- Interviewer Debriefing Methods**
 - Distributed rating forms by asking interviewers to evaluate each question with 3 points scale just after finishing field test
 - Difficulty of Interviewer's Reading
 - Difficulty of R's answering
 - Inconsistency of understanding terms from Rs
 - These forms completed by 89% (33 of 37) of the interviewers who participated in the test
 - Control version : 18 interviewers
 - Test version : 15 interviewers

15

Illustrative Analysis Results

- Response Process and Error**

Response Process	Response Error	Example Question
Comprehension	Misunderstanding of concepts and terms	Household type
Retrieval	Difficult to remember	Residence one year ago
Judgment and estimation	Computation complexity	Total floor space
Reporting	Sensitivity	Marital status

16

Household Type

- Wording**

Control (2010 Census)	Test (based on Cognitive Interviewing)
Which among 1 ~ 5 best describes this household?	Which among 1 ~ 5 best describes this household?
1. One-family household 2. Household consisting of a family and a non-family member(s) 3. One person household 4. Household of 5 or less persons who have no blood ties 5. Household of 6 or more persons who have no blood ties	1. One-family household including kinship 2. Household consisting of a family including kinship and no blood ties member(s) 3. One person household 4. Household of 5 or less persons who have no blood ties 5. Household of 6 or more persons who have no blood ties

17

Household Type

- Cognitive Interviewing**
 - R's understanding of a family concept differ from a survey definition
 - Most Rs thought that only parent-children relationship was included in a family concept and it didn't cover relatives by blood and marriage ties
- Interviewer Debriefing** (1: lowest, 3 highest)

	Difficulty of I's Reading	Difficulty of R's Answering	Inconsistency of Understanding from Rs
Control	1.00	1.00	1.00
Test	1.00	1.07	1.13

18

Household Type

- Split Ballot Survey**

Distribution	Control		Test		x ²	p
	Freq.	%	Freq.	%		
1 (wording changed)	1809	82.75	1807	82.02	2.90	0.575
2 (wording changed)	2	0.09	4	0.18		
3	373	17.06	388	17.61		
4	2	0.09	2	0.09		
5	0	0.00	2	0.09		
Total	2186	100.00	2203	100.00		
Nonresponse	8	0.35	10	0.44	0.24	p=0.625
Edit failure	72	3.18	35	1.56	12.80	p<0.001

19

Residence One Year Ago

- Wording**

Control (2010 Census)	Test (based on Cognitive Interviewing)
Where did this person live one year ago(11.1. 2011.)?	Where did this person live one year ago(11.1.2011.)?
1. Wasn't born 2. In the current house 3. In the same Si/Gun/Gu but different house 4. In another Si/Gun/Gu 5. North Korea or foreign country (if 4 or 5)	1. Wasn't born 2. In the current house 3. In the same Up/Myun/Dong but different house 4. In another Up/Myun/Dong 5. North Korea or foreign country (if 4 or 5)
_____ Si/Do/Country _____ Si/Gun/Gu	_____ Si/Do/Country _____ Si/Gun/Gu _____ Up/Myun/Dong

20

Residence One Year Ago

- Cognitive Interviewing**
 - R's remember depending on retrieval ques
 - Some Rs said that they didn't remember 'Gu' but knew 'Dong' because 'Dong' was much more familiar to them despite the fact that Census division said asking detailed information occurred response burden
- Interviewer Debriefing** (1: lowest, 3 highest)

	Difficulty of I's Reading	Difficulty of R's Answering	Inconsistency of Understanding from Rs
Control	1.11	1.17	1.18
Test	1.07	1.21	1.11

21

Residence One Year Ago

- Split Ballot Survey**
 - Nonresponse of Open-ended questions

	Control		Test		x ²	p
	Freq.	%	Freq.	%		
Si/do	5	4.20	6	2.49	0.79	0.375
Si/Gun/Gu	6	5.04	6	2.49	1.61	0.204
Eup/Myeon/Dong (added)	-	-	13	5.39	-	-

22

Total Floor Space

- Wording

Control (2010 Census)	Test (based on Cognitive Interviewing)
What is the total floor space of this house? - Indicate the area that is used for residential purposes and round to the whole number ___ m ² (1pyeong is 3.3m ²)	What is the total floor space of this house? - In the case of collective housing except for apartment, please indicate the net area, not the rental area - Indicate the area that is used for residential purposes and round to the whole number ___ m ² (1pyeong is 3.3m ²)

23

Total Floor Space

- Cognitive Interviewing
 - Rs confused with coverage
 - In calculating total floor space for the collective housing, most Rs did not include residential area where other household resided
- Interviewer Debriefing (1: lowest, 3 highest)

	Difficulty of I's Reading	Difficulty of R's Answering	Inconsistency of Understanding from Rs
Control	1.18	1.47	1.25
Test	1.27	1.67	1.40

24

Total Floor Space

- Split Ballot Survey

	Control		Test		x ²	p
	Freq.	%	Freq.	%		
Nonresponse	7	0.35	44	2.40	0.24	0.625
Edit failure	12	0.61	0	0.00	-	<0.001

	Total		Apartment		Except Apartment	
	Control	Test	Control	Test	Control	Test
Average	82.52	100.99	72.39	92.29	113.40	112.95
M-W U	126895.5		448715		168230	
P	<0.001		<0.001		0.680	

25

Marital Status

- Wording

Control (2010 Census)	Test (based on Cognitive Interviewing)
What is the person's marital status? 1. Never married 2. Married 3. Widowed 4. Divorced	What is the person's marital status? 1. Never married 2. Married (including cohabitation) 3. Widowed 4. Divorced 5. Separated

26

Marital Status

- Cognitive Interviewing
 - In the vignette of 'separated', most Rs answered that they were married while separated was considered divorced depending on the situation in the Census
 - R's thought that categories of separated and cohabitated were not that sensitive to answer (7 of 11)
- Interviewer Debriefing (1: lowest, 3 highest)

	Difficulty of I's Reading	Difficulty of R's Answering	Inconsistency of Understanding from Rs
Control	1.83	1.89	1.53
Test	1.67	1.73	1.33

27

Marital Status

- Split Ballot Survey

Distribution	Control		Test		x ²	p
	Freq.	%	Freq.	%		
1	1263	25.19	1271	24.83	0.32 (after collapsing 5 toward 4)	0.317
2(wording changed)	3119	62.22	3211	62.74		
3	440	8.78	441	8.62		
4	191	3.81	166	3.24		
5 (category added)	-	-	29	0.57		
Total	5013	100.00	5118	100.00		
Nonresponse	225	4.28	170	3.21	8.45	0.004
Edit failure	15	0.29	10	0.19	1.05	0.307

28

Conclusion

- Summary
 - Q. Are these three pretesting methods equally useful in identifying problems in the questionnaire?
- Tentatively, cognitive interviewing and field test showed similar results from respondents' perspective
 - Cognitive Interviewing: a appropriate method for finding out hidden problems
 - Filed test: a good way to validate from lab results

29

Conclusion

- Interviewer's rating form: it is more effective when comparing to other items in same questionnaire than comparing the test version with the control version

Difficulty of I's Reading	Household Type	Residence one year ago	Total floor space	Marital Status
Control	1.00	1.11	1.18	1.83
Test	1.00	1.07	1.27	1.67

- Future Plan
 - Preparing for the second Census test in coming Sep.
 - Using a respondent debriefing method for large samples to overcome small number of voluntary subjects in cognitive interviewing

30