

# Dealing with randomisation bias in a social experiment exploiting the randomisation itself: The case of ERA

4<sup>th</sup> Joint IZA/IFAU Conference on Labor Market Policy Evaluation

October 25-26, 2012

**Barbara Sianesi**



# The Employment Retention and Advancement (ERA) program

## **ERA treatment**

*Offer* of a package of time-limited support once in work

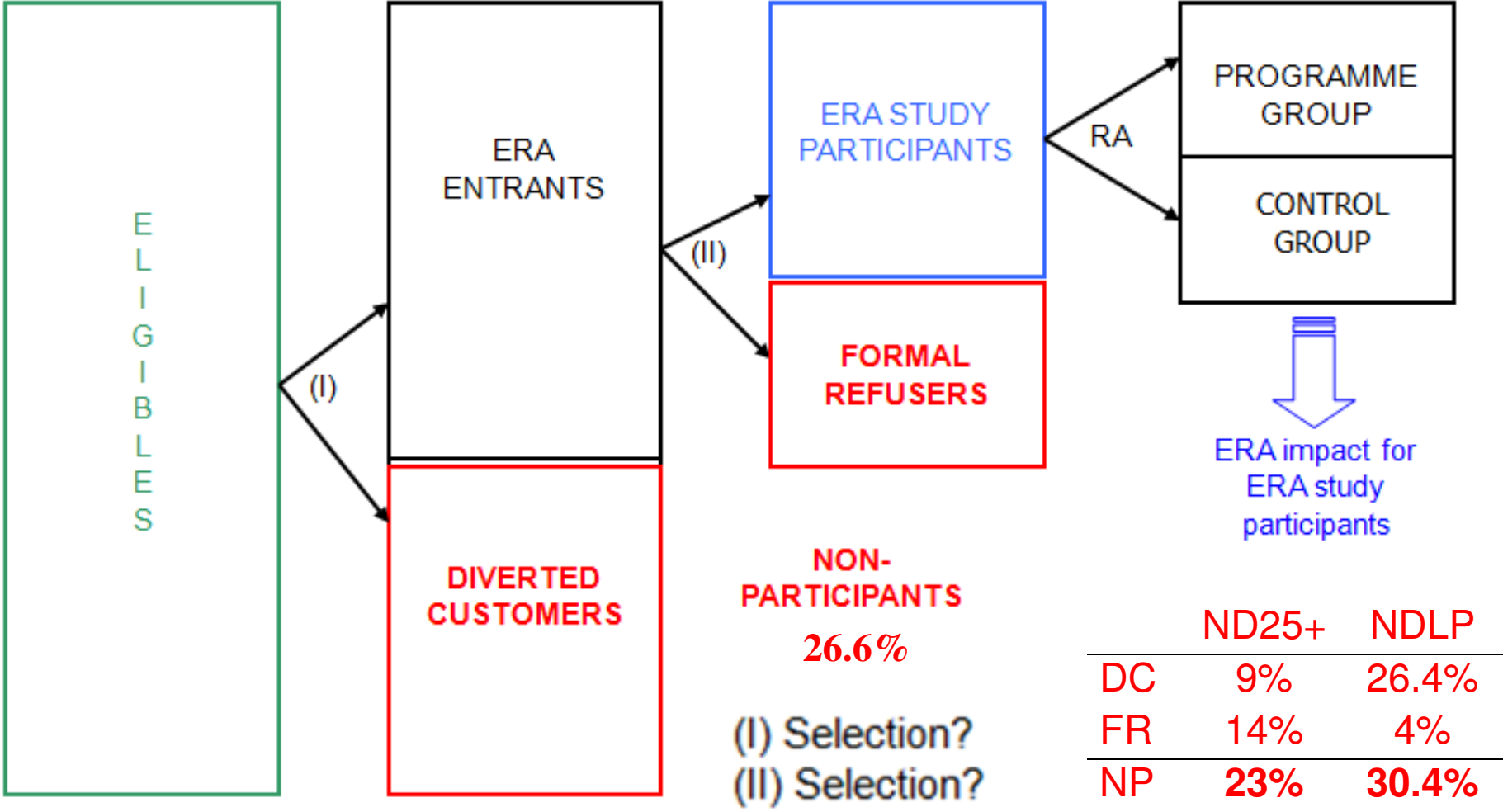
## **Eligibles**

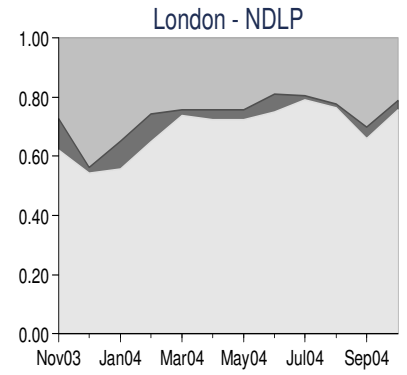
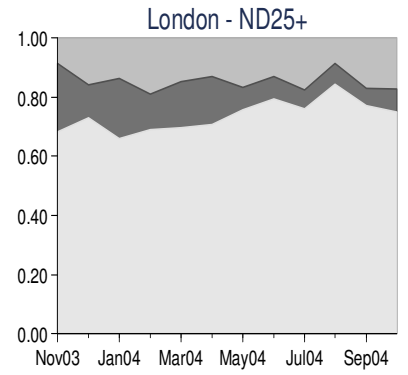
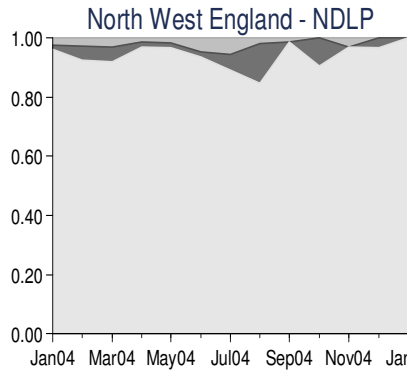
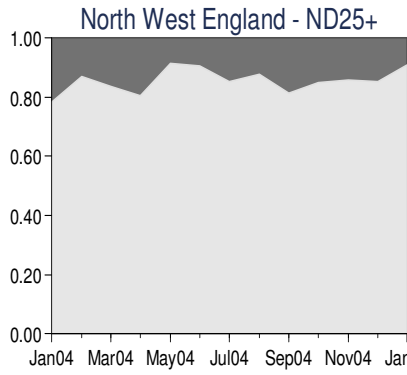
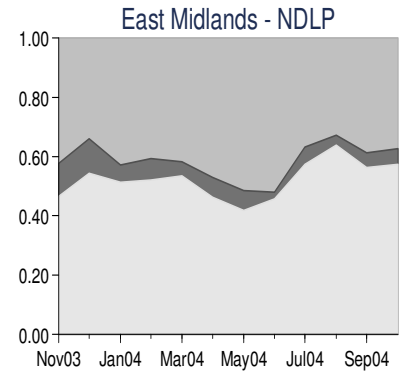
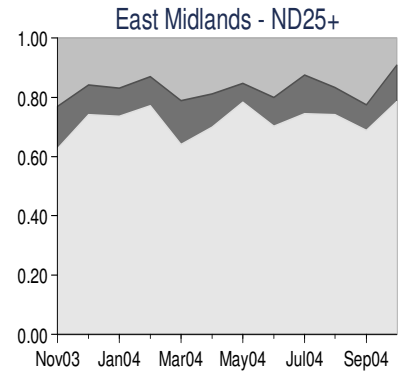
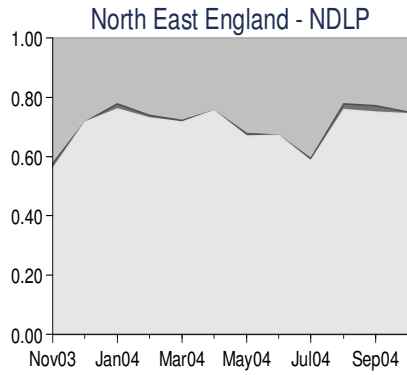
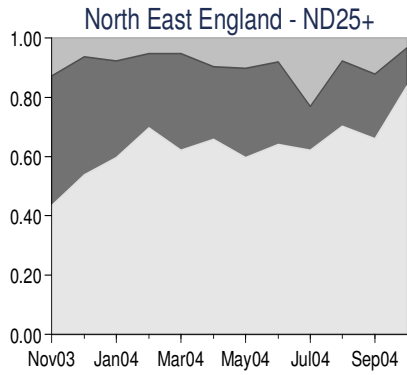
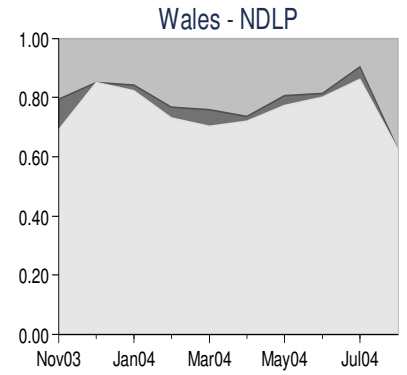
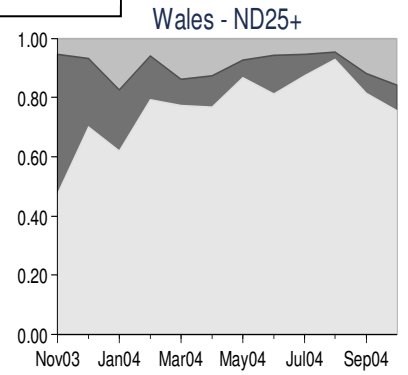
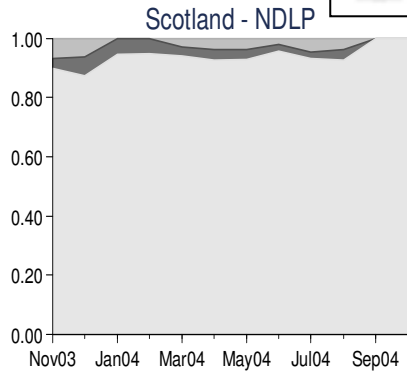
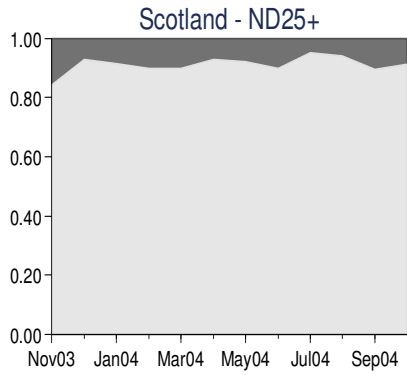
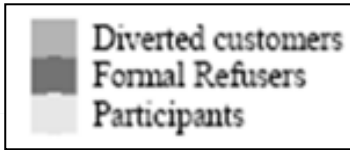
- 1) LT unemployed mandated for ND25+
- 2) Unemployed volunteering for NDLP
- 3) [LPs on WTC working PT who volunteer for ERA]

## **Tested**

Large-scale ( $N=16,000$ ), multi-site (6 districts) RA social experiment  
Intake: Nov 2003 – Jan 2005 (pilots end Oct 2007)

# Non-participation in the ERA study





# Issues raised by non-participation

- Policymaker interested in impact of offering ERA for all those eligible to receive the offer. (ERA as an integral component of the New Deal)
- But... ERA tested only on a potentially selective subset of the eligibles

## 2 ways to view non-participation

1. **Impact of offering ERA eligibility on the *eligibles*** (in the 6 districts)  
→ Assess the scope for **randomisation bias** (Heckman, 1992 and Heckman *et al.*, 1999) in the experimental estimate for the parameter of interest
2. **Impact of offering ERA eligibility on the *study participants*** (in the 6 districts)  
→ Has non-participation affected the extent of **external validity** of the experimental results, and hence their representativeness and policy relevance?

## **ERA study offers rare chance to look at this issue!**

- offer (ITT)
- whole population (ATE)
- admin data

## **Research questions**

- Impact on all eligibles
  - Impact on the non-participants
- Impact on all eligibles *versus* experimental impact on the participants

## Methodology & “What’s new”

Moving beyond an experiment’s limitations by climbing on its shoulders

- Impact estimates under **selection-on-observables** (CIA)
  - matching and re-weighting estimators
- When follow-up data for non-participants available
  - exploit experiment to assess whether CIA is corroborated (plus help with choice of how to summarise  $X$ )
  - in cases where not, use information from experiment to adjust the non-experimental estimates
- When follow-up data for non-participants not available:
  - deal with additional issue of survey/item non-response
  - exploit experiment for subgroups for whom
    - admin earnings outcomes are available
    - equivalent survey and admin earnings outcomes are available

# Sample and data

	ND25		NDLP	
Eligibles	7,796	100.0%	7,261	100.0%
– Study non-participants	1,790	23.0%	2,209	30.4%
– Study participants	6,006	77.0%	5,052	69.6%
– with survey outcome	1,840	30.6%	1,745	34.5%
– without survey outcome	4,166	69.4%	3,307	65.5%

## Outcomes

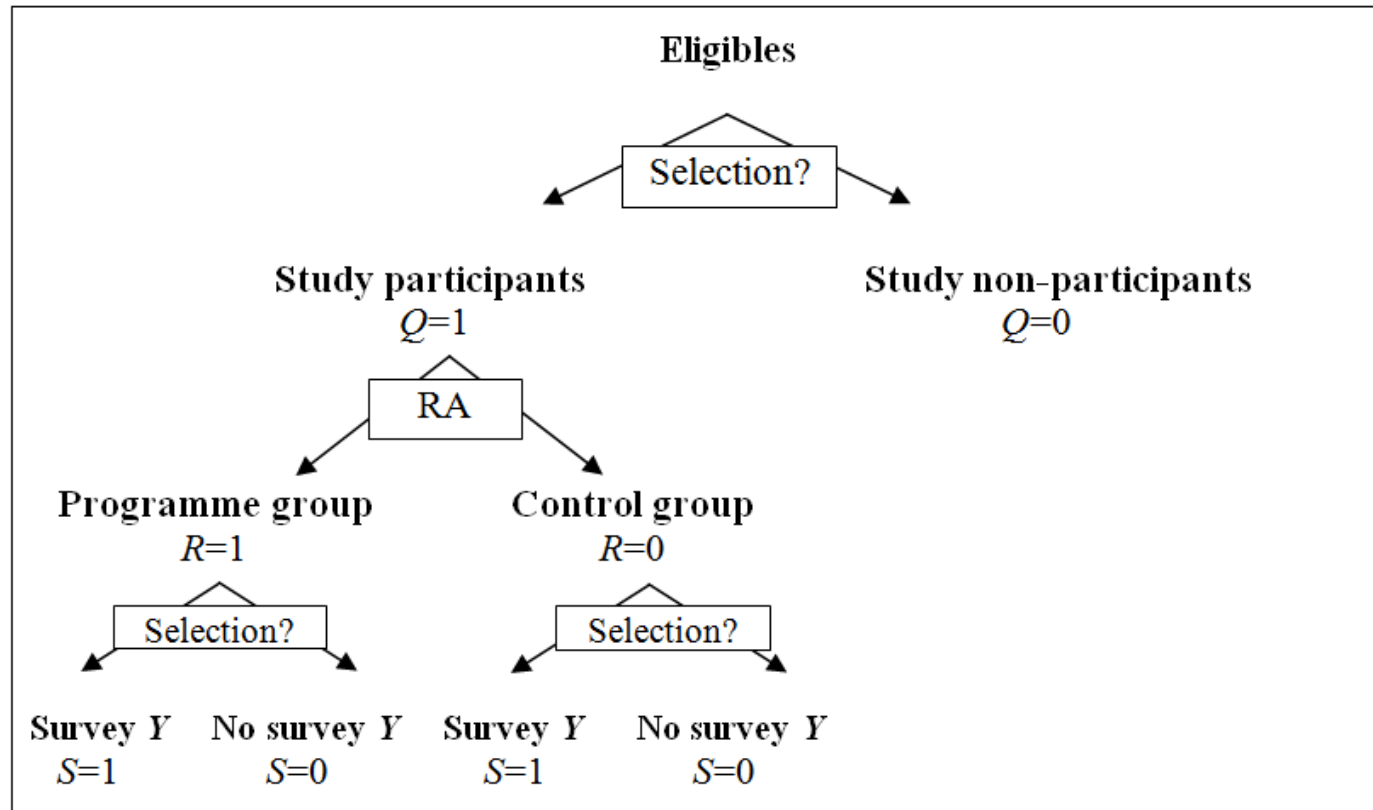
- 12-month follow-up
- employment (ever employed and days) – admin data
- annual earnings – survey data
- fiscal year 2004/05 earnings – admin data



## Control variables

ERA district	
Inflow month	District-specific month from random assignment start when the individual started the ND25 Gateway or volunteered for NDLP
Local conditions	Total New Deal caseload at office, share of lone parents in New Deal caseload at office, quintiles of the index of multiple deprivation, local unemployment rate
Demographics	Gender, age, ethnic minority, disability, partner (ND25+), number of children (NDLP), age of youngest child (NDLP)
Current spell	Not on benefits at inflow (NDLP), employed at inflow (indicator of very recent/current employment), time to show up (defined as the time between becoming mandatory for ND25+ and starting the Gateway or between being told about NDLP and volunteering for it), early entrant into ND25+ programme (Spent <540 days on JSA before entering ND25+)
Labour market history (3 years pre-inflow)	<p>Past participation in basic skills, past participation in voluntary programmes (number of previous spells on: NDLP, New Deal for Musicians, New Deal Innovation Fund, New Deal Disabled People, WBLA or Outreach), past participation in ND25+</p> <p>Active benefit history, inactive benefit history, employment history:</p> <ol style="list-style-type: none"> <li>(1) parsimonious summary</li> <li>(2) monthly employment dummies</li> <li>(3) dummies for sequences of employment/benefits/neither states; 90% of sample</li> <li>(4) dummies for ever employed in 12m window at any time in the past</li> </ol>

# Methodology



Average effect on participants

$$ATE_1 \equiv E(Y_1 - Y_0 \mid Q=1)$$

Average effect on non-participants

$$ATE_0 \equiv E(Y_1 - Y_0 \mid Q=0)$$

Average effect on *all* eligibles

$$ATE \equiv E(Y_1 - Y_0) = (1-p) \cdot ATE_1 + p \cdot ATE_0 \quad p \equiv \Pr\{Q=0\}$$

## Follow-up data

$$ATE = (1-p) \cdot \{E(Y | R=1) - E(Y | R=0)\} + p \cdot \{E(Y_1 | Q=0) - E(Y | Q=0)\}$$

Akin to getting the **ATNT** using matching methods

### Assume (CIA-1)

$$(CIA-1) \quad E(Y_1 | Q=0, X) = E(Y_1 | Q=1, X) \quad \text{and} \quad (CS)$$

### Implementation

Match to each non-participant one or more similar programme group member(s) based on  $p(x) \equiv P(Q=0 | X) = P(Q=0 | Q=0 \vee Q=1, X)$  or  $P(Q=0 | Q=0 \vee R=1, X)$ .

### Test (CIA-0)

$$(CIA-0) \quad E(Y_0 | Q=0, X) = E(Y_0 | Q=1, X) \quad \text{i.e.} \quad E(Y | Q=0, X) = E(Y | R=0, X)$$

By OLS or matching

## **Robustness analysis / adjustment**

If (CIA-0) not met:  $E(Y_0 | Q=0, X) = \theta_0 E(Y_0 | Q=1, X)$   $\theta_0 \neq 1$

Relax identifying (CIA-1) by allowing participants and non-participants with the same  $X$  to differ in terms of some unobservable translating into a proportional difference of  $\theta$

(CIA-1)  $E(Y_1 | Q=0, X) = \theta E(Y_1 | Q=1, X)$  in particular, set  $\theta = \theta_0$

## ***No follow-up data***

$$ATE = (1-p) \cdot ATE_1 + p \cdot E(Y_1 - Y_0 | Q=0)$$

Akin to **attrition**

→ reweigh  $Y$  of the participants on the basis of the  $X$  of the eligibles to make them representative – in terms of  $X$  – of the full eligible population

### **Motivation: Survey earnings**

- all we had originally
- clean definition of components, incl. all part-time and self-employed
- same horizon (and all post-treatment)

### **Assume (CIA-1 & CIA-0)**

$$(CIA) \quad E(Y_1 - Y_0 | Q=1, X) = E(Y_1 - Y_0 | Q=0, X) \quad \text{hence} = E(Y_1 - Y_0 | X)$$

### **Implementation**

The empirical counterpart can be derived in several ways:

## 1) Reweighting

Directly weigh the outcomes of the (responding) participants so as to reflect the distribution of  $X$  in the eligible population.

### ***Ignoring survey/item non-response***

$$A\hat{T}E = \left[ \frac{(1-p)p_R}{\#(R=1)} \sum_{i \in \{R=1\}} \frac{y_i}{(1-p(x_i))p_R(x_i)} \right] - \left[ \frac{(1-p)(1-p_R)}{\#(R=0)} \sum_{i \in \{R=0\}} \frac{y_i}{(1-p(x_i))(1-p_R(x_i))} \right]$$

where  $p_R \equiv P(R=1 \mid Q=1)$  and  $p_R(x) \equiv P(R=1 \mid Q=1, x)$

### ***Allowing for survey/item non-response (selective non-response based on X)***

$$A\hat{T}E = \left[ \frac{1}{\#(R=1, S=1)} \sum_{i \in \{R=1, S=1\}} \frac{(1-p)p_{RS1}}{(1-p(x_i))p_{RS1}(x_i)} y_i \right] - \left[ \frac{1}{\#(R=0, S=1)} \sum_{i \in \{R=0, S=1\}} \frac{(1-p)p_{RS0}}{(1-p(x_i))p_{RS0}(x_i)} y_i \right]$$

where  $p_{RS1} \equiv P(R=1, S=1 \mid Q=1)$ ,  $p_{RS0} \equiv P(R=0, S=1 \mid Q=1)$  and  $p_{RS1}(x)$  and  $p_{RS0}(x)$  are the corresponding probabilities conditional on  $x$

## 2) Matching

Construct weights to realign  $X$  via matching

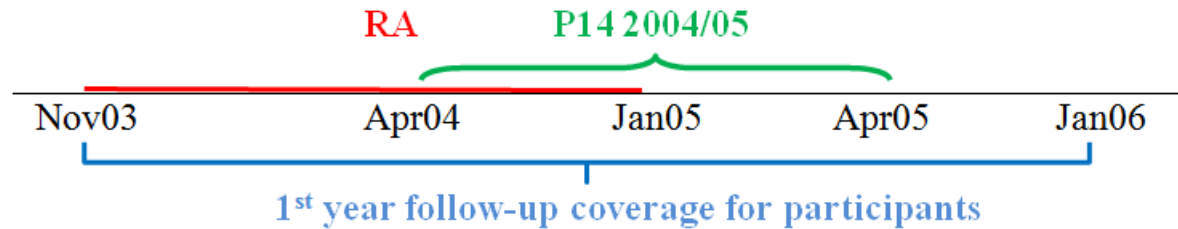
- exact specifications of pscore and response probabilities not needed
- can assess actual comparability

Can do it in 2 ways:

- A) separately recover  $ATE_0$  and then combine it with experimental  $ATE_1$  to get the  $ATE$  or
- B) recover  $ATE$  directly

Again, can do A) and B) both ignoring and allowing for survey and item non-response  
(Detailed matching protocols in Appendix)

## Sensitivity analysis



Exploit experiment for:

- (I) Subgroup for whom FY 2004/05 admin earnings *outcomes* are available  
→ **post April group** (35% of ND25+ and 41% of NDLP eligibles)
  - Is (CIA-0) test passed for admin earnings? (Plus guidance on how to construct  $X$ )
  
- (II) Subgroup for whom 'equivalent' 1<sup>st</sup>-year survey and FY 2004/05 admin earnings outcomes are available (i.e. cover same horizon)  
→ **March/May group** (25% of ND25+ and NDLP eligibles)
  - Is (CIA-0) test passed for admin earnings?
  - experimental impacts
    - admin earnings for March-May respondents vs admin earnings for full March-May group
    - admin earnings for March-May respondents vs survey earnings for March-May respondents
    - admin earnings for full March-May group vs survey earnings for March-May respondents
  - Compare *ATE* for full March-May group in terms of admin earnings to *ATE* on March-May respondents in terms of survey earnings – taking account of non-response or not



# Findings – ND25+

## Days employed (admin)

	$ATE_1$	OLS	CIA-0 test Matching	$\theta_0$
Raw	4.0		-9.4***	0.834
All other X's plus				
summary	4.6*	-7.9***	-9.7***	0.829
monthly employment	4.8**	-7.6***	-9.4***	0.835
ever employment	5.0**	-7.6***	-9.4***	0.835
sequence	4.8**	-7.9***	-8.8***	0.843
summary + monthly employment	4.8**	-7.7***	-9.2***	0.837
summary + ever employed	5.0**	-7.7***	-9.3***	0.837
summary + sequence	4.8**	-8.0***	-8.8***	0.843

## Days employed (admin)

	<i>p</i>	<i>Formal</i>	<i>Diverted</i>	$ATE_1$	<b>Raw</b>	$\theta$ raw	<b>OLS</b>	<b>Matching</b>	$\theta_0$
Scotland	8.7	8.7	0.0	8.6	-17.3	0.690	-8.2	-8.3	0.828
NE Eng	34.9	26.1	8.8	-10.3	-33.9***	0.565	-23.6***	-27.8***	0.616
NW Eng	14.6	14.6	0.0	7.5	-7.0	0.864	-1.8	-3.0	0.937
Wales	20.7	11.1	9.6	-13.6	-12.0	0.816	-16.3	-7.9	0.864
E Midlands	27.5	10.7	16.8	8.0	-4.3	0.934	-5.7	-7.7	0.885
London	25.8	11.1	14.8	8.9**	-3.6	0.915	-3.7	-2.8	0.932
but NE Eng	21.0	11.5	9.5	6.5***	-5.7*	0.894	-4.7	-5.3	0.901

	<i>p</i>	$ATE_1$	$ATE_0$	$ATE$	$ATE_1 \neq ATE$
<b>All but NE England</b>	<b>0.210</b>	<b>6.5**</b>	<b>9.7***</b>	<b>7.2***</b>	<b>not sign</b>
All districts			10.1***	5.9***	*
All districts, $\theta$ -adjusted	0.230	4.6*	5.3*	4.8*	not sig
NE England			8.1	-3.9	**
NE England, $\theta$ -adjusted	0.349	-10.3	-12.7**	-11.2*	not sig

## Ever employed (admin)

	<i>p</i>	<i>Formal</i>	<i>Diverted</i>	$ATE_1$	<b>Raw</b>	$\theta$ raw	<b>OLS</b>	<b>Matching</b>	$\theta_0$
All	23.0	13.6	9.4	0.017	-0.062***	0.808	-0.044***	-0.056***	0.825
Scotland	8.7	8.7	0.0	0.047	-0.096*	0.726	-0.039	-0.041	0.861
NE Eng	34.9	<b>26.1</b>	<b>8.8</b>	-0.036	-0.191***	0.541	-0.149***	-0.172***	0.571
NW Eng	14.6	14.6	0.0	0.033	-0.024	0.915	0.010	0.010	1.038
Wales	20.7	11.1	9.6	-0.035	-0.027	0.923	-0.017	-0.004	0.987
E Midls	27.5	10.7	16.8	0.031	-0.073**	0.817	-0.060**	-0.071**	0.819
London	25.8	11.1	14.8	0.022	-0.017	0.929	-0.009	-0.010	0.958
no NE,EM	18.8	11.8	7.0	0.023*	-0.041**	0.858	-0.007	-0.011	0.956

	<i>p</i>	$ATE_1$	$ATE_0$	$ATE$	$ATE_1 \neq ATE$
<b>All but NE Eng, E Midls</b>	<b>0.188</b>	<b>0.023*</b>	<b>0.026*</b>	<b>0.024**</b>	<b>not sign</b>
All districts			0.056***	0.026**	***
All districts, $\theta$ -adjusted	0.230	0.017	0.003	0.014	not sig
NE England			0.092**	0.009	***
NE England, $\theta$ -adjusted	0.349	-0.036	-0.048**	-0.040	not sig
E Midlands			0.083***	0.045**	*
E Midlands, $\theta$ -adjusted	0.275	0.031	0.010	0.025	not sig

## Annual earnings (survey)

		$\Delta_{S=1,X} \neq ATE$	
$\Delta_{S=1,X}$		445.4**	
	allowing for non-response, weighting	579.6**	not sig
<i>ATE</i>	allowing for non-response, separate CS	551.2***	not sig
	ignoring non-response, separate CS	469.4*	not sig
	ignoring non-response, joint CS	468.1*	not sig

### (CIA-0) test in terms of 2004/05 earnings (admin)

	<i>X</i> + history	Raw	$\theta$ raw	OLS	Matching	$\theta_0$
Post-April group	monthly employment	-147	0.937	-240	-208	0.910
March-May group	summary + monthly empl.	-465*	0.776	-275	-109	0.938

### Full March-May group:

	<i>p</i>	$ATE_1$	$ATE_0$	$ATE_a$	$ATE_1 \neq ATE_a$
<b>(A) 2004/05 earnings (admin)</b>	<b>0.248</b>	<b>183.9</b>	<b>531.7**</b>	<b>270.2</b>	<b>not sig</b>

### **(B) annual earnings (survey)**

		$\Delta_{S=1,X} \neq ATE_b$	$ATE_a \neq ATE_b$
$\Delta_{S=1,X}$		273.1	
	allowing for non-response, weighting	819.6	not sig
$ATE_b$	allowing for non-response, separate CS	700.4**	not sig
	ignoring non-response, separate CS	365.6	not sig
	ignoring non-response, joint CS	377.1	not sig

# Findings – NDLP

## Days employed (admin)

	<i>p</i>	<i>Formal</i>	<i>Diverted</i>	<i>ATE</i> <sub>1</sub>	<b>Raw</b>	<i>θ</i> raw	<b>OLS</b>	<b>Matching</b>	<i>θ</i> <sub>0</sub>
All	30.4	4.0	26.4	-2.2	3.8	1.003	-10.4***	-11.2**	0.914
Scotland	5.3	2.8	2.5	9.6	-75.0***	0.478	-71.1***	-64.2**	0.490
NE Eng	29.2	1.0	28.2	0.0	2.7	1.023	-14.7	-18.8*	0.864
NW Eng	6.2	3.7	2.5	21.1**	38.4*	1.336	31.6*	27.6	1.224
Wales	23.6	3.6	20.1	-16.6	20.3	1.141	-4.9	-7.6	0.955
E Midlands	47.1	5.9	41.2	-15.5**	4.9	1.044	-11.1*	-10.7	0.916
London	31.0	4.9	26.1	-3.5	12.9	1.127	-3.4	-6.4	0.947
NW,W,L	23.4	4.3	19.1	-1.9	13.3*	1.117	-2.9	-9.2	0.931

## Days employed (admin)

	$p$	$ATE_1$	$ATE_0$	$ATE$	$ATE_1 \neq ATE$
<b>NW Eng, Wales, London</b>	<b>0.234</b>	<b>-1.9</b>	<b>-3.8</b>	<b>0.5</b>	<b>not sig</b>
All districts			-2.1	-2.2	not sig
All districts, $\theta$ -adjusted	0.304	-2.2	-9.6**	-4.5	*
Scotland			72.1	12.9	not sig
Scotland, $\theta$ -adjusted	0.053	9.6	-3.7	8.9	not sig
NE England			5.7	1.7	not sig
NE England, $\theta$ -adjusted	0.292	0.0	-11.1	-3.2	not sig
East Midlands			-4.4	-10.3	not sig
East Midlands, $\theta$ -adjusted	0.471	-15.5**	-13.7	-14.7**	not sig

## Ever employed (admin)

	<i>p</i>	<i>Formal</i>	<i>Diverted</i>	$ATE_1$	<b>Raw</b>	$\theta$ raw	<b>OLS</b>	<b>Matching</b>	$\theta_0$
All	30.4	4.0	26.4	-0.006	0.004	1.009	-0.041***	-0.041**	0.928
Scotland	5.3	2.8	2.5	0.041	-0.130	0.786	-0.063	-0.056	0.895
NE Eng	29.2	1.0	28.2	-0.020	-0.003	0.994	-0.063**	-0.071*	0.880
NW Eng	6.2	3.7	2.5	0.063*	0.165**	1.319	0.130*	0.130	1.242
Wales	23.6	3.6	20.1	-0.044	0.031	1.049	-0.052	-0.038	0.946
E Midlands	47.1	5.9	41.2	-0.036	-0.001	0.998	-0.049**	-0.043	0.923
London	31.0	4.9	26.1	0.000	0.046	1.105	-0.026	-0.030	0.942
no NE,EM	23.4	4.3	19.1	0.009	0.033	1.066	-0.018	-0.030	0.947

	<i>p</i>	$ATE_1$	$ATE_0$	$ATE$	$ATE_1 \neq ATE$
<b>All but NE Eng, EMidls</b>	<b>0.234</b>	<b>0.011</b>	<b>0.007</b>	<b>0.010</b>	not sig
All districts			0.015	0.000	not sig
All districts, $\theta$ -adjusted	0.304	-0.006	-0.010	-0.007	not sig
NE England			0.033	-0.005	not sig
NE England, $\theta$ -adjusted	0.292	-0.020	-0.034	-0.024	not sig
East Midlands			0.020	-0.009	**
East Midlands, $\theta$ -adjusted	0.471	-0.036	-0.020	-0.028	not sig

## Annual earnings (survey)

		$\Delta_{S=1,X} \neq ATE$	
$\Delta_{S=1,X}$		788.1***	
	allowing for non-response, weighting	762.1***	not sig
<i>ATE</i>	allowing for non-response, separate CS	708.5***	not sig
	ignoring non-response, separate CS	763.3***	not sig
	ignoring non-response, joint CS	761.8***	not sig

### (CIA-0) test in terms of 2004/05 earnings (admin)

	<i>X</i> + history	Raw	$\theta$ raw	OLS	Matching	$\theta_0$
Post-April group	summary	210	1.087	-82	-69	0.976
March-May group	summary	323	1.132	-10	52	1.019

### Full March-May group:

	<i>p</i>	$ATE_1$	$ATE_0$	$ATE_a$	$ATE_1 \neq ATE_a$
(A) 2004/05 earnings (admin)	<b>0.320</b>	<b>375.9</b>	<b>621.8</b>	<b>454.7*</b>	<b>not sig</b>

### (B) annual earnings (survey)

		$\Delta_{S=1,X} \neq ATE_b$	$ATE_a \neq ATE_b$
$\Delta_{S=1,X}$		736.1	
	allowing for non-response, weighting	759.9	not sig
$ATE_b$	allowing for non-response, separate CS	566.0	not sig
	ignoring non-response, separate CS	863.2	not sig
	ignoring non-response, joint CS	864.2	not sig



# Conclusions

## Experiments

- can suffer from randomisation bias
- with admin outcome data offer ways to support non-experimental methods in addressing it

## ND25+

- For employment outcomes, experimental set-up qualified conclusions from non-experimental methods in a way which was consistent throughout:  
Non-experimental methods →  $ATE_1$  under-estimates  $ATE$   
Once 'corrected' →  $ATE_1$  representative of  $ATE$
- Earnings results (gain) appear reliable; again  $ATE_1$  representative of  $ATE$

## NDLP

- For employment outcomes, irrespective of correction: a zero (negative for E Midls durations)  $ATE_1$  representative of  $ATE$
  - Earnings results appear reliable; sizeable  $ATE_1$  representative of  $ATE$
- ERA experiment does not seem to have suffered from randomisation bias/loss of external validity in terms of year-1 results.