

# Wage & Employment Dynamics

## THE WED PROJECT

### ASHE-Census Linkage

*Very preliminary:  
please do not cite or share without permission*



Presentation prepared by: John Forth (City University), Van Phan (UWE), and Carl Singleton (University of Reading)

*"This work was produced using statistical data from ONS. The use of the ONS statistical data in this work does not imply the endorsement of the ONS in relation to the interpretation or analysis of the statistical data. This work uses research datasets which may not exactly reproduce National Statistics aggregates."*

# Motivation



- Personal information
  - Age
  - Gender
  - Home postcode
- Employment information
  - Wage
  - Working hours
  - etc
- Employer information
  - Work postcode
  - etc

- Personal information
  - Age
  - Gender
  - Education Qualification
  - Country of Birth
  - Religion
  - Language ability
  - Ethnicity
- Family information

- Value of ASHE-Census linkage
  - Returns to human capital
  - Wage gaps by gender, ethnicity, disability
  - Wage progression among migrants
  - Relevance of partner status

# Background

- ASHE:
  - 1% sample of PAYE jobs in GB and NI
  - Sample drawn from PAYE register in January 2011 (top-up in April 2011)
  - Survey reference date: 13th April 2011
- Census:
  - It's a census... but we only have access to E&W
  - Census Day: 27th March 2011
- Linkable on common fields:
  - Primary fields: First name, middle name, surname; DOB; Gender; Home postcode
  - Secondary fields: Occupation (SOC); Industry (SIC); Work postcode
- ONS have done a first-run at the linkage
- WED team contributing with feedback, QA and documentation

# Linkage process

- ONS seek matches for all individuals with E&W resident addresses in ASHE 2010, 2011 or 2012
- Stage 1: Deterministic match on 46 combinations of name, gender, DOB and home postcode, for example:
  - Name + gender + DOB + postcode / Name + DOB + postcode (no gender)
  - Full name / surname only
  - Literal name / soundex
  - Full postcode / postcode district only
- If unmatched -> Stage 2: Probabilistic match, including SIC and SOC
- If unmatched -> Stage 3: Determ. match on: gender, DOB, home pcode, work pcode and SIC/SOC (not done yet)
- Each matched record has a 'score' based on the completeness of the matchkeys

# Linkage outcomes: conceptual

- False negatives should not arise: all persons in ASHE should appear in the Census (new arrivals in the UK after 27th March are unlikely to have NINO by ASHE reference date)
- False positives are possible, if the linking process links to the wrong person (e.g. through errors in the match keys)
  - ASHE data missing individual's name in 23% of cases
  - ASHE 2010 coded to SOC(2000): converted to SOC(2010) for comparability with ASHE 2011, 2012 and Census 2011, but crosswalk not perfect
  - ONS undertake clerical checks on match accuracy
  - Common fields beyond the match keys will help us to flag up 'risky cases' worthy of further investigation during QA (e.g. Census record indicates that the person is not in work in w/c 27th March 2011)
  - May lead to improvements in the linkage algorithm

## PLEASE NOTE:

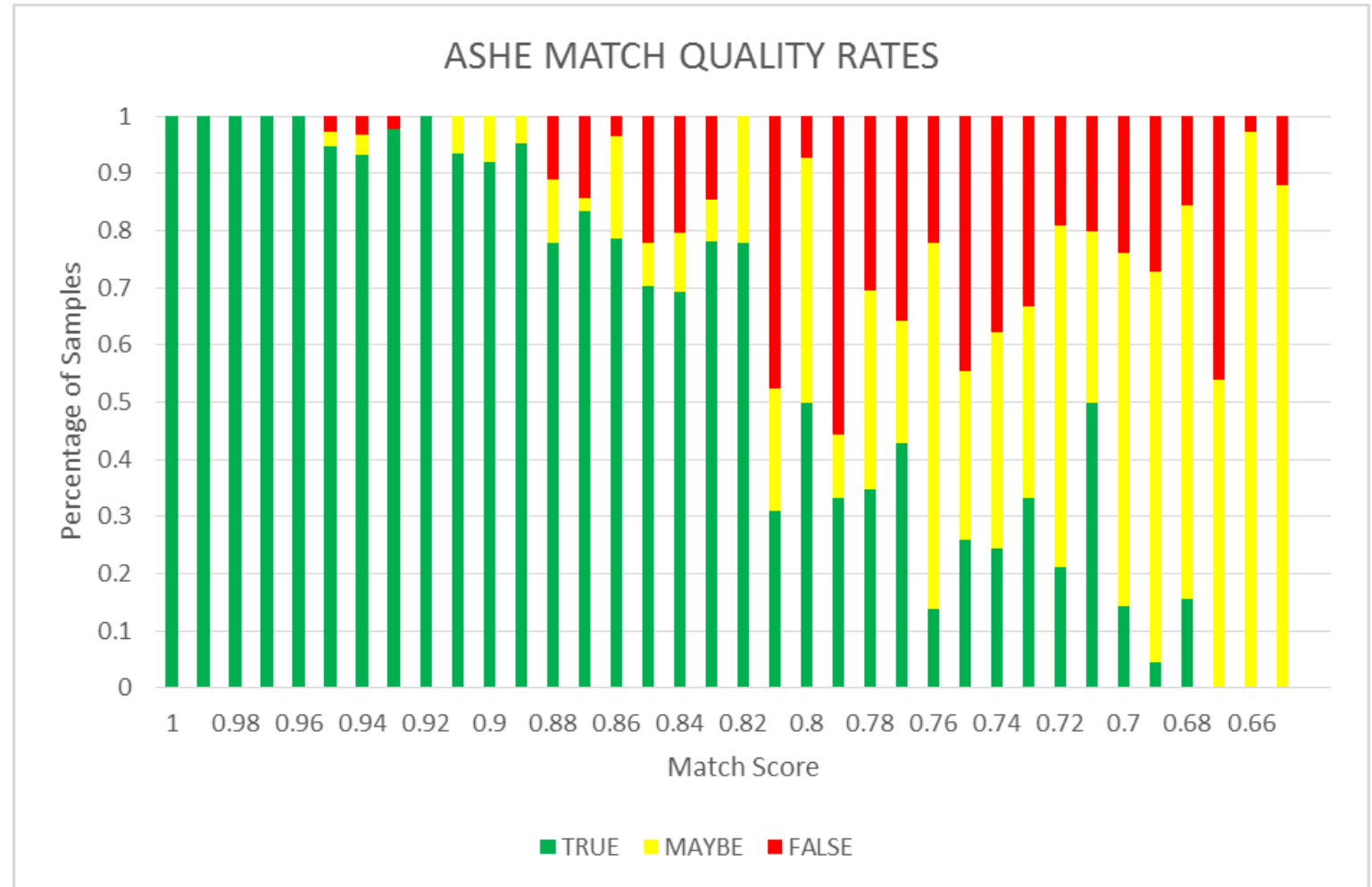
- ONS have undertaken a **first-run**: the outcomes are **preliminary** and are likely to change

# First-run: clerical checks

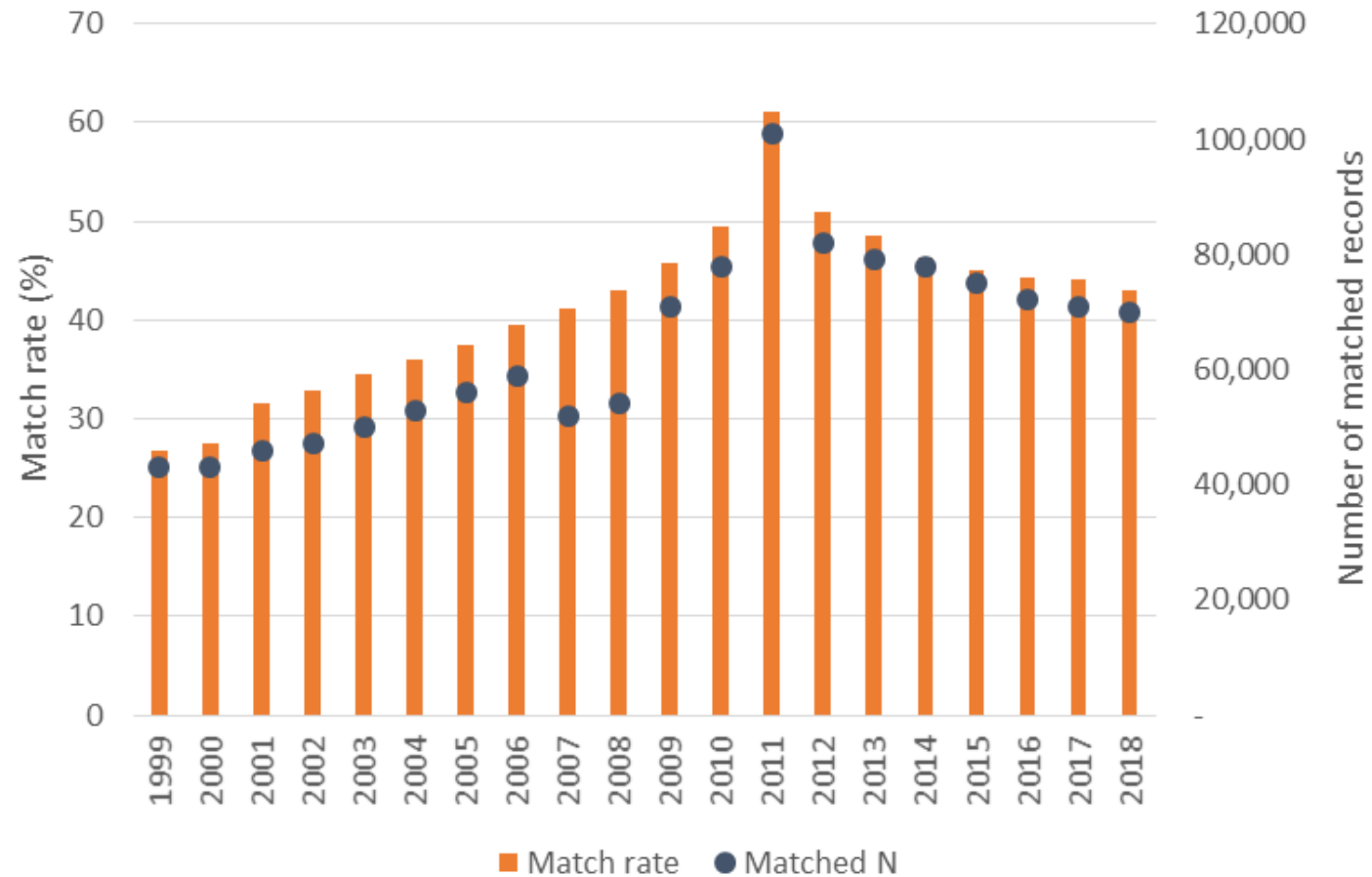
- Checks on 895 matched records with match score  $\geq 0.65$

In this sample of 895:

- 90% of matches with a score  $\geq 0.82$  are found to be 'true'
- 82% of all 'true' matches have a score  $\geq 0.82$



# First-run: linkage outcomes



- Matched records = those with match score  $\geq 0.82$
- Data includes match score
- 79% of all matched records have a match score of 1.0



# First-run: who gets matched from ASHE in 2011?

	Odds ratio of being matched
Female (vs. Male)	0.765*** (0.010)
Age (+1 year)	0.975*** (0.001)
Age <sup>2</sup>	0.729*** (0.003)
Firm tenure (+1 year)	1.033*** (0.001)
Firm tenure <sup>2</sup>	0.843*** (0.006)
Gross hourly pay, excl. overtime (hexo)	1.000 (0.000)
hexo <sup>2</sup>	1.000 (0.000)
Basic paid weekly hours (bhr) (+1 hour)	0.997*** (0.001)
bhr <sup>2</sup>	1.001 (0.003)
London (E12000007) (vs. North East)	0.611*** -0.018
Southeast (E12000008) (vs. North East)	0.889*** -0.026
Observations/Jobs	155,768
Pseudo-R <sup>2</sup>	0.063

Note: Exponentiated coefficients of logistic regression model; Standard errors in parentheses; Robust SEs. {Squared terms centred & x100}; includes other control variables (1-digit occ, 1-letter industry, firm size, private sector status, etc)

\* p<0.10, \*\* p<0.05, \*\*\* p<0.01



# Raw (and a bit adjusted) log gender wage gaps in 2011 – Comparing ASHE and ASHE-CENSUS estimates, Gross hourly pay excl. overtime

Ln(hexo)	ASHE (1)	ASHE-Census (2)	ASHE (3)	ASHE-Census (4)
Female	-0.180*** (0.003)	-0.197*** (0.003)	-0.174*** (0.003)	-0.183*** (0.004)
Observations/Jobs	158,195	98,998	123,723	74,673
Firm (enterprise) fixed effects	No	No	Yes	Yes
R <sup>2</sup>	0.028	0.034	0.466	0.462

Note: Robust standard errors in parentheses, no other control variables, constant not shown.

\* p<0.10, \*\* p<0.05, \*\*\* p<0.01

# Adding in some ASHE covariates – the residual gender wage gap in England and Wales, 2011

Ln(hexo)	ASHE (1)	ASHE-Census (2)	ASHE (3)	ASHE-Census (4)
Female	<b>-0.136***</b> (0.002)	<b>-0.145***</b> (0.003)	<b>-0.115***</b> (0.002)	<b>-0.122***</b> (0.003)
Age (+1 year)	0.004*** (0.000)	0.005*** (0.000)	0.004*** (0.000)	0.004*** (0.000)
Age <sup>2</sup>	-0.045*** (0.001)	-0.052*** (0.001)	-0.033*** (0.001)	-0.036*** (0.001)
Firm tenure (+1 year)	0.011*** (0.000)	0.011*** (0.000)	0.009*** (0.000)	0.009*** (0.000)
Firm tenure <sup>2</sup>	-0.023*** (0.001)	-0.028*** (0.002)	-0.019*** (0.001)	-0.025*** (0.002)
Full-time (>30 hours)	-0.082*** (0.002)	-0.084*** (0.003)	-0.046*** (0.003)	-0.051*** (0.003)
Permanent contract	0.022*** (0.004)	0.019*** (0.005)	0.023*** (0.005)	0.025*** (0.006)
Observations/Jobs	154,258	96,626	121,408	73,397
SOC 1-digit	Y	Y	Y	Y
SIC 1-digit	Y	Y	N	N
Home region	Y	Y	Y	Y
Firm (enterprise) fixed effects	N	N	Y	Y
R <sup>2</sup>	0.557	0.556	0.697	0.697

Note: Robust standard errors in parentheses, \* p<0.10, \*\* p<0.05, \*\*\* p<0.01. {Squared terms centred & x100}. Other control variables not shown: home address index of multiple deprivation, private sector status of firm and number of employees in the firm (columns (1) & (2) only)

# Adding in some Census covariates – the residual gender wage gap in England and Wales, 2011

Ln(hexo)	All ASHE-Census matched jobs		Central Govt only	
	(1)	(2)	(3)	(4)
Female	<b>-0.133***</b> (0.003)	<b>-0.112***</b> (0.003)	<b>-0.133***</b> (0.008)	<b>-0.138***</b> (0.009)
GCSEs ( <i>vs no qualifications</i> )	0.068*** (0.004)	0.058*** (0.004)	0.029* (0.015)	0.046*** (0.016)
Apprenticeship	0.098*** (0.008)	0.080*** (0.009)	-0.007 (0.034)	0.023 (0.037)
A-level	0.130*** (0.005)	0.109*** (0.005)	0.080*** (0.016)	0.100*** (0.016)
Degree	0.304*** (0.005)	0.270*** (0.005)	0.256*** (0.016)	0.258*** (0.017)
Vocational/other qualifications	0.072*** (0.006)	0.069*** (0.007)	0.063** (0.026)	0.085*** (0.028)
English language	0.068*** (0.006)	0.029*** (0.007)	0.008 (0.018)	0.010 (0.019)
Non-UK born	0.008 (0.009)	0.002 (0.010)	-0.009 (0.037)	-0.025 (0.037)

(cont.)

# Continued...

Ln(hexo)	All ASHE-Census matched jobs		Central Govt only	
	(1)	(2)	(3)	(4)
Mixed/multiple ethnic groups ( <i>vs White</i> )	-0.003 (0.010)	-0.019* (0.011)	-0.005 (0.031)	-0.003 (0.032)
Asian/ Asian British	-0.045*** (0.005)	-0.045*** (0.006)	0.052*** (0.016)	0.046*** (0.016)
Black/African/Caribbean/Black British	-0.086*** (0.007)	-0.074*** (0.007)	-0.042*** (0.015)	-0.061*** (0.016)
Other ethnic group	-0.032* (0.017)	-0.021 (0.019)	0.118*** (0.046)	0.117** (0.048)
Caring responsibilities	-0.016*** (0.004)	-0.010*** (0.004)	-0.007 (0.008)	0.000 (0.008)
Good health in general	0.011 (0.015)	0.014 (0.017)	0.018 (0.030)	0.034 (0.034)
Disability	-0.046*** (0.005)	-0.043*** (0.006)	-0.035*** (0.012)	-0.027** (0.012)

(Cont.)

# Continued...

Ln(hexo)	All ASHE-Census matched jobs		Central Govt only	
	(1)	(2)	(3)	(4)
Couple family	0.019*** (0.004)	0.017*** (0.005)	0.013 (0.012)	0.010 (0.013)
Partner also in employment (proxy)	-0.011*** (0.004)	-0.010** (0.004)	-0.011 (0.012)	-0.011 (0.012)
Youngest child under 4 years old	0.048*** (0.004)	0.032*** (0.004)	0.033*** (0.010)	0.020* (0.010)
1 dependent child ( <i>vs no children</i> )	0.007** (0.003)	0.014*** (0.003)	0.014 (0.009)	0.021** (0.009)
2 dependent children	0.034*** (0.004)	0.035*** (0.004)	0.023** (0.010)	0.039*** (0.010)
3 or more dependent children	0.025*** (0.007)	0.030*** (0.007)	0.035* (0.019)	0.052*** (0.019)
Observations/Jobs	96,141	73,000	11,395	11,137
SOC 1-digit	Y	Y	Y	Y
SIC 1-digit	Y	N	Y	N
Home region	Y	Y	Y	Y
Firm (enterprise) fixed effects	N	Y	N	Y
R <sup>2</sup>	0.587	0.717	0.535	0.593

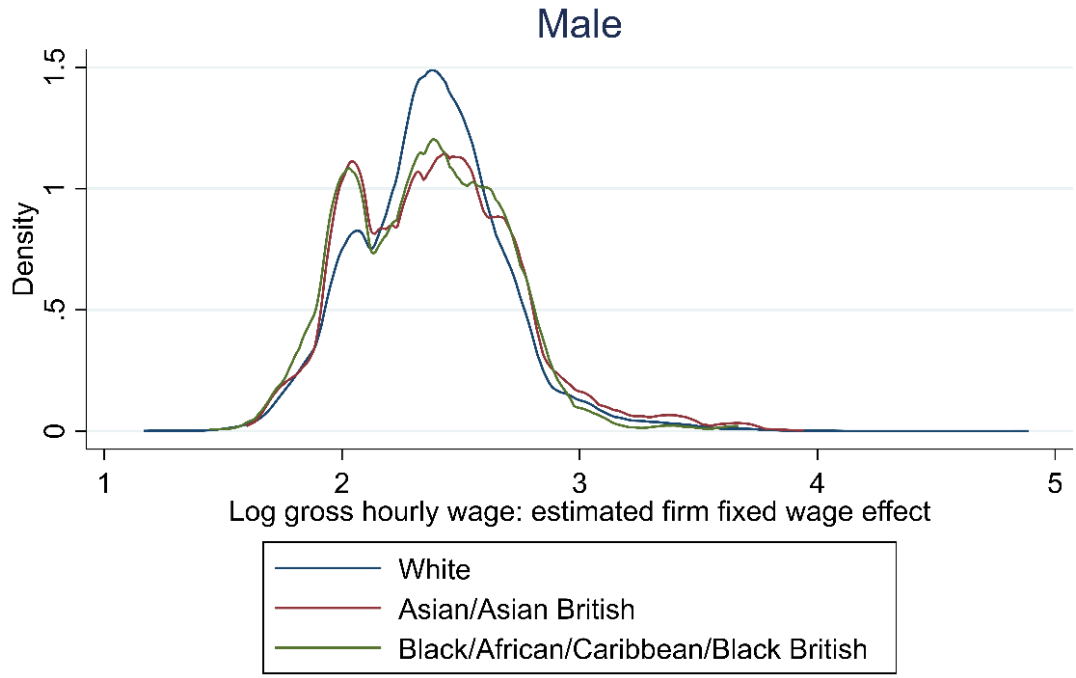
Note: Robust standard errors in parentheses; Control for other ASHE variables \* p<0.10, \*\* p<0.05, \*\*\* p<0.01. Includes all previous ASHE control variables also

# Using ASHE-Census to estimate Gender-Ethnicity wage gaps (within firms)

Ln(hexo)	(1)	(2)	(3)
Female ( <i>vs Male</i> )	-0.210*** (0.004)	-0.194*** (0.004)	-0.163*** (0.004)
Mixed/ multiple ethnic groups ( <i>vs White</i> )	-0.123*** (0.025)	-0.136*** (0.023)	-0.134*** (0.022)
Asian/ Asian British	-0.054*** (0.011)	-0.071*** (0.010)	-0.132*** (0.010)
Black/African/Caribbean/Black British	-0.142*** (0.015)	-0.134*** (0.013)	-0.173*** (0.013)
Other ethnic group	-0.051 (0.036)	-0.026 (0.034)	-0.037 (0.032)
Female##Mixed/ multiple ethnic groups	0.100*** (0.032)	0.086*** (0.029)	0.080*** (0.027)
Female##Asian/ Asian British	0.099*** (0.015)	0.068*** (0.014)	0.064*** (0.013)
Female##Black/African/Caribbean/Black British	0.183*** (0.019)	0.113*** (0.017)	0.113*** (0.016)
Female##Other ethnic group	0.099** (0.048)	-0.015 (0.046)	-0.044 (0.044)
Observations/Jobs	98,521	74,285	74,268
Education controls	N	N	Y
Firm (enterprise) fixed effects	N	Y	Y
R <sup>2</sup>	0.035	0.462	0.557

Note: Robust standard errors in parentheses, \* p<0.10, \*\* p<0.05, \*\*\* p<0.01.

# Gender-Ethnicity "Sorting" over Firms - Distributions of estimated firm-specific wage effects in England and Wales, 2011



Epanechnikov kernel, 0.05 bandwidth.  
N: White=32,044; Asian/Asian British=2,354; Black/African/Caribbean/Black British=960  
Controls: gender#Ethnicity & Highest Education, only... for now...



Epanechnikov kernel, 0.05 bandwidth.  
N: White=34,072; Asian/Asian British=2,263; Black/African/Caribbean/Black British=1,284  
Controls: gender#Ethnicity & Highest Education, only... for now...



# Summary

- ONS have done a **first-run** at linking 2011 E&W Census to ASHE
- Adds education, personal and family characteristics to around 60% of sample
- Probability of being matched from ASHE to Census is so far very much not random, especially by gender and region
- WED team assisting with QA and feedback to improve the match....but the potential is clear
- Will soon be able to revisit the determinants of pay and inequality in the E&W labour market, in many dimensions, esp. addressing the **influence of firms, jobs and workplaces.**
- This is important for policy (e.g., pay gap reporting, pay audits, low pay work)