The HILDA Survey: The First 5 Years (or so)

Mark Wooden
Project Director, HILDA Survey

The Need for a Household Panel Survey

- Social science research about change …
  yet most data are cross-sections!
  - Retrospective data are problematic
  - Repeated cross-sections only partial solution
- Panel data useful for evaluating effects of policy
- Almost every other developed country has one!

The Australian Panel Data Experience

- Focus on sub-groups (eg, LSAY, LSIA)
- Limited info about household structure
- Weak on economic data
- Often short-lived (SEUP, LSIA)
- Small samples (Negotiating the Life Course)

HILDA: Introductory Remarks

- Funded by the Commonwealth through FaCSIA
- Now have funding for up to 8 waves
- Intended to be a resource for academia
- Melbourne Institute responsible for design and management
- Want to know more?
  - see Sept 2002 and Sept 2004 issues of AustER
  - but for EVERYTHING, see our website:

www.melbourneinstitute.com/hilda/

What's In It?

- Three main areas of interest
  - **Labour market activity and employment**
  - **Income**
  - **Family life and household composition**
- Lots of background characteristics
- SCQ = health measures, lifestyle behaviours, attitudes, personality (w5), hh expenditure (w5-w6)

- Special modules:
  - W2 = Wealth
  - W3 = Retirement
  - W4 = Youth issues; Health insurance
  - W5 = Fertility
  - W6 = Wealth
  - W7 = Retirement; Skills (?)
  - W8 = Fertility

Key Design Features

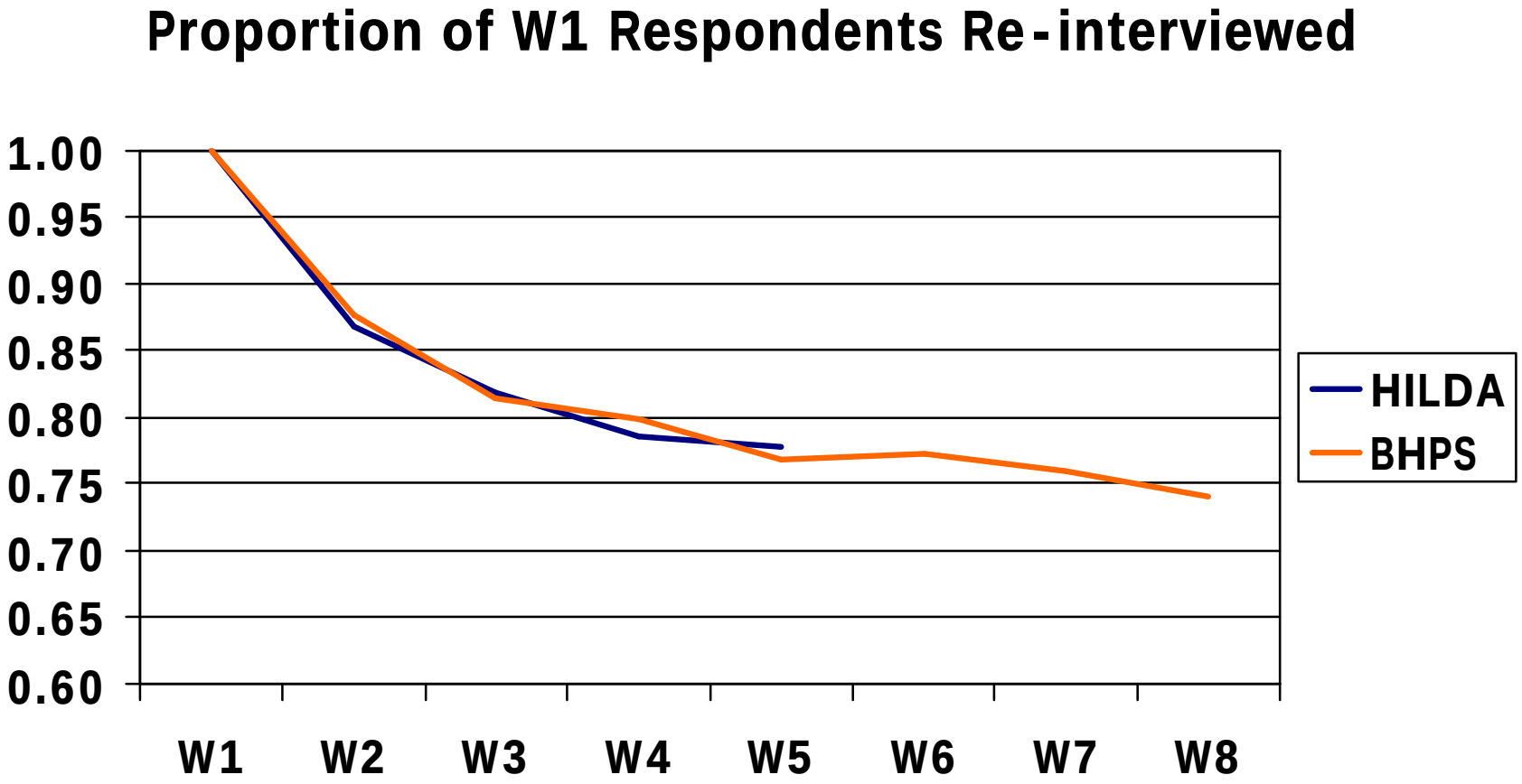
- Indefinite life panel
- Conducted annually: 5 waves now complete
- Sampling unit = the household …
  but it is the individuals we follow
- Reference population = residents of **private** dwellings
- All household members in sample …
  but only persons aged 15+ interviewed
- F2F interview, with a self-completion supplement
  (Telephones used as last resort)
- We offer bribes

Sample / Response

- Selected sample = 12,252 dwellings from 488 CDs
- 7682 households cooperated (66%) = individual sample of 13,969

	RESPONSE RATES %	
	Wave-on-wave	W1 respondents
W2	86.8	86.8
W3	90.4	81.9
W4	91.6	78.5
W5	94.4	77.8

Attrition: HILDA vs BHPS



W1 Sample Representativeness (%)

	ABS	HILDA
Men	49.3	47.4*
Women	50.7	52.6
15-19	8.8	8.4
20-24	8.9	7.3*
25-34	18.7	18.6
35-44	19.0	21.5*
45-54	17.1	17.4
55-64	11.8	12.1
65+	15.6	14.7
Aust-born	72.4	74.5*
MES country	10.2	10.9
Other country	17.5	14.5*

	ABS	HILDA
Employed FT	42.1	41.5
Employed PT	17.4	19.5*
Unemployed	4.3	4.4
Not in LF	36.3	34.6
Sydney	21.5	17.3*
Married	58.7	63.3*
Indigenous	1.7	1.8
Employee (employed persons only)	86.0	86.5

Statistical Issue: Attrition Bias

Respondent characteristics		
	C	*R|C*
Moved	-	
Female	+	ns
Age	+	+/-
Poor English ability	-	-
Indigenous		ns
Single	-	ns
# children in hh*	-	ns
# adults in hh*	ns	-

	C	*R|C*
Education		+
HH equiv income		ns
Employed	+	-
LT health condition	ns	-
Unit / other dwelling	-	
Renter	-	ns
Remoteness	-	ns
SES of region	+	ns

Statistical Issue: Attrition Bias (cont'd)

Design features / Interview situation		
	C	*R|C*
Ivwr workload	+/-	ns
Ivwr continuity	ns	+(?)
# calls made	-	-
Part responding hh	-	-
Cooperative	+	+
Understanding	ns	+
Suspicious	ns	-

	C	*R|C*
SCQ not returned	-	-
Item NR		-
PQ length		+/-
HQ length		+
Telephone ivw		+

Statistical Issue: Imputation of Missing Data

- Item non-response also a source of bias
- Mainly an issue for monetary variables
- Missing income data:
  - FY personal income (resp. p) 15% W1 14% W2
  - FY personal income (enum. p) 21% W1 20% W2
  - FY household income 29% W1 28% W2
- Missing wealth data (W2):
  - HQ wealth components 20%
  - PQ wealth components 15%
  - Total household wealth 39%

Statistical Issues: Imputation (contd)

- Main income / wealth items imputed
  - 2 methods: both employ nearest-neighbour technique
    - Parametric
    - Simple multiplicative (columns and row) model
- Cross-section estimates seem reasonable
  - Income estimates higher than ABS
  - Aggregate wealth estimates plausible

How Good are the HILDA Income Estimates?

Mean FY (2002-03) income per person	ABS ($) (1)	HILDA ($) (2)	(2)/(1)
Wages and salaries	20902	21909	1.048
Government pensions (excl CCB, FTB)	2606	2740	1.051
Business income	1792	1600	0.893
Investment income	1267	1679	1.325
Other income	1052	1811	1.721
Total FY income	27619	29740	1.077

How Good are the HILDA Wealth Estimates?

2002 (qtr 4)	Nat Accts ($b)	HILDA ($b)
Financial assets	1236.7	1125.1
Non-financial assets	1955.4	2440.3
Total assets	3192.1	3565.4
Financial liabilities	630.8	516.5
Net worth	2561.3	3048.9

Data Use

- Growing community of registered users
  - 655 licensed users
  - 265 users of latest release (and growing)
- Over 70 papers using HILDA data have already been published / accepted
- Many other working papers / conference papers / reports

www.melbourneinstitute.com/hilda/hbiblio.html

What Topics Have Been Covered?

- Patterns of cohabitation
- Children's living arrangements
- Post-separation contact with children
- Familial status and housework
- Childlessness
- Distribution of household wealth
- Residential property investors
- Retirement savings
- Income, wealth and happiness
- Time stress
- Jobless households
- Underemployment
- Employment transitions
- Non-standard employment and job satisfaction
- Part-time employment and wages
- Health and labour force participation
- Forgone earnings of mothers
- Poverty dynamics

Example 1: Poverty Persistence

Annual poverty rates (%)

2000-01	14.2
2001-02	13.2
2002-03	12.6
2003-04	13.5

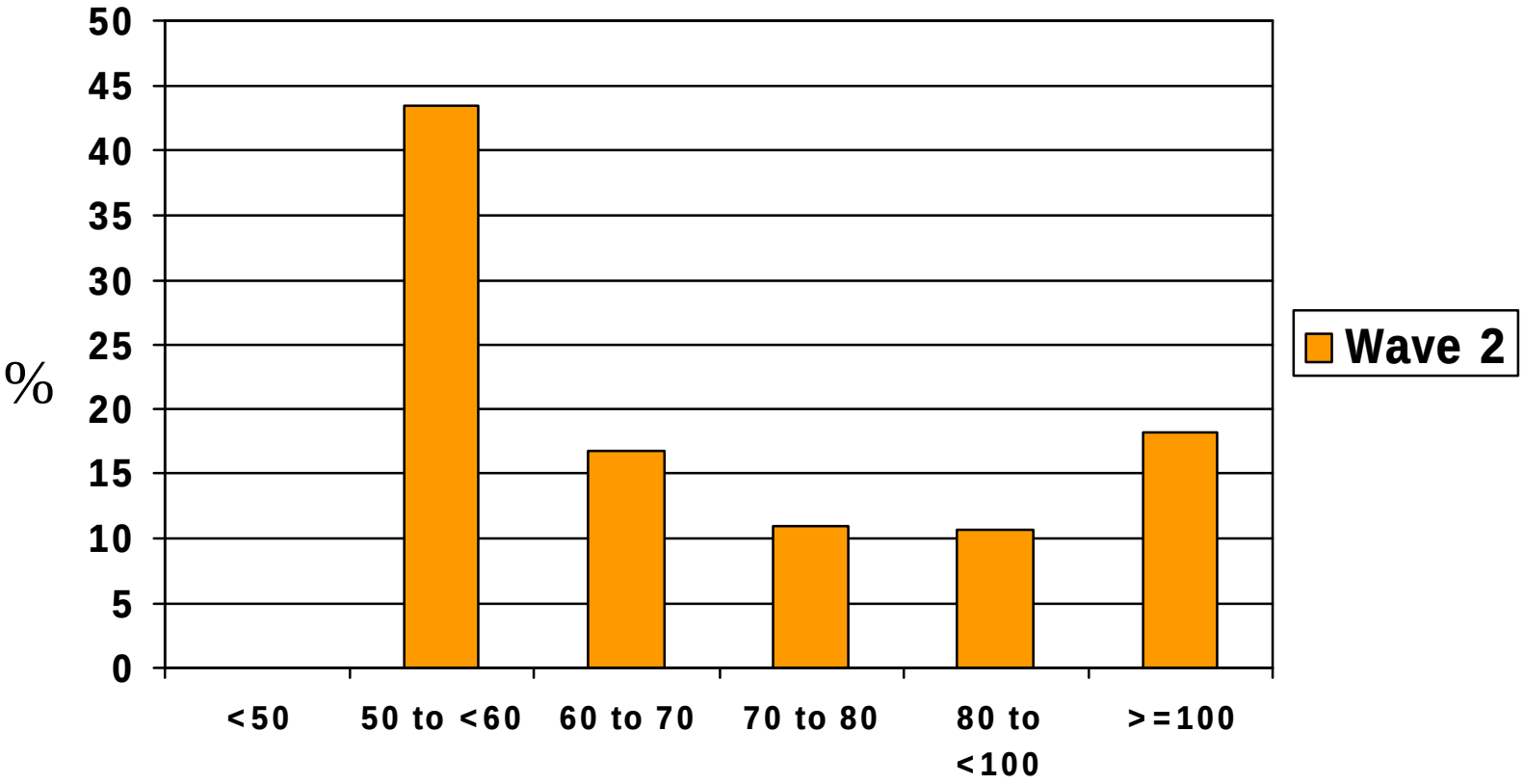
Poverty persistence (% distribution)

Never poor	75.7
Poor in only one year	7.8
Poor in two out four years	6.8
Poor in three out of four years	4.7
Poor in all four years	5.0

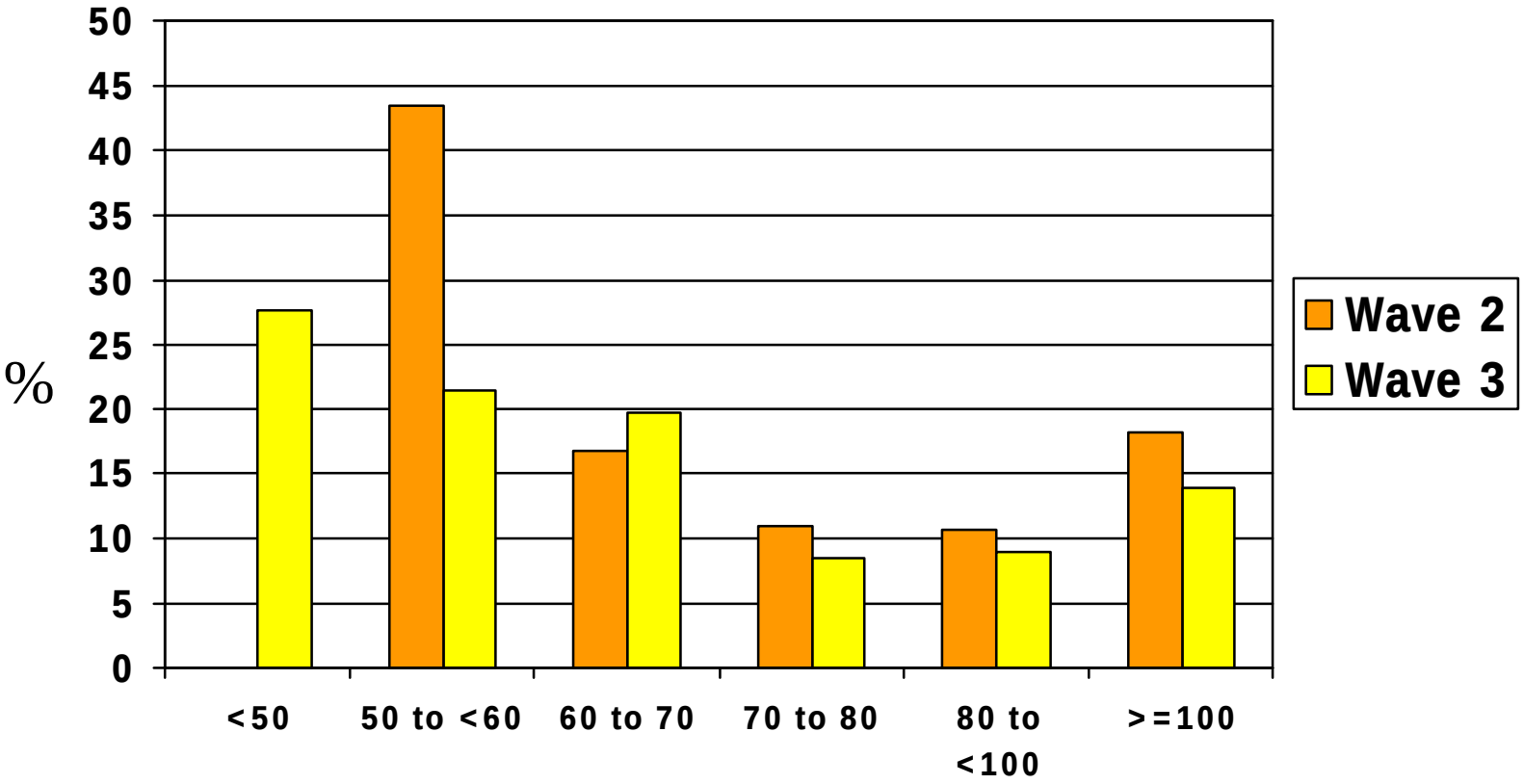
Poverty transitions, 2001 to 2004 (%)

P=Poor; N=Not poor	%	P=Poor; N=Not poor	%
PPPP	5.0	NPPP	2.3
PPPN	0.0	NPPN	0.0
PPNP	0.7	NPNP	0.5
PPNN	1.7	NPNN	2.8
PNPP	1.7	NNPP	4.0
PNPN	0.0	NNPN	0.0
PNNP	0.7	NNNP	1.1
PNNN	3.9	NNNN	75.7

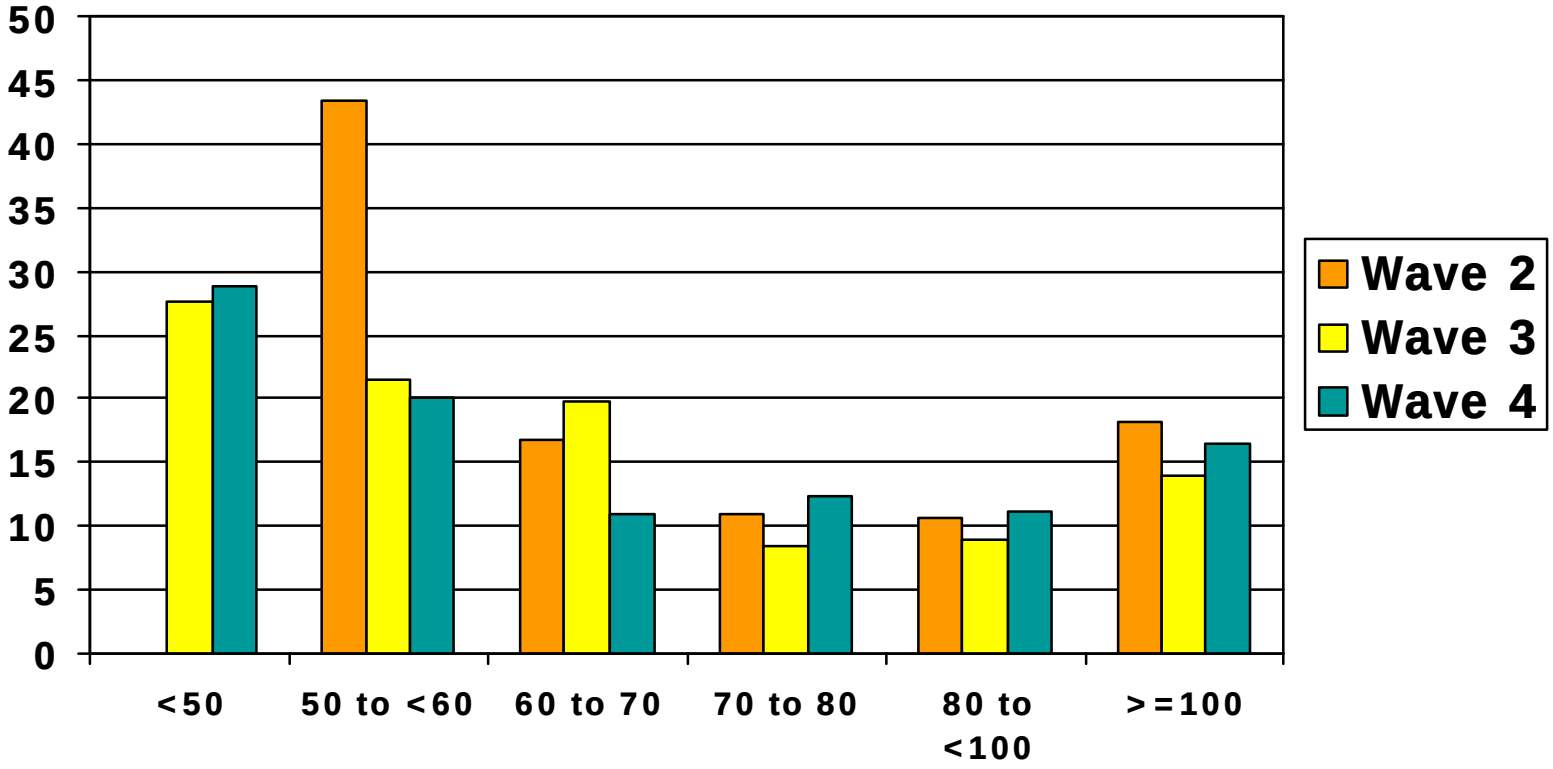
Equivalised Household Income of Persons Exiting Poverty in W2 *(% of median)*



Equivalised Household Income of Persons Exiting Poverty in W2 *(% of median)*



Equivalised Household Income of Persons Exiting Poverty in W2 *(% of median)*



Example 2: PT / FT wage differentials & unobserved heterogeneity

- In cross-section regressions unobservables = bias
- $\ln W_{it} = X_{it}'\beta + \alpha PT_{it} + \mu_i + \varepsilon_{it}$

  where μ_i = individual specific effect (eg, unobserved ability)
- If corr (μ_i, PT) < 0 and corr (μ_i, W) > 0 then estimate of α biased down if μ_i omitted

Estimated PT coefficients

- Unbalanced panel of employees
- All persons earning less than $1 per hr or more than $100 per hour were dropped
- Dep var = ln real hourly wage in main job

Spec.	*Men*		*Women*	
	Coeff	*t-stat*	*Coeff*	*t-stat*
Pooled	-0.044	2.78	-0.017	1.85
RE	0.031	2.13	0.055	5.68
FE	0.120	6.36	0.114	8.55

Future Developments

- Seeking new funding for W9-W12
- An immigrant top-up
- The CNEF
- CAPI