

# Recent Advances in Social Program Evaluation : Use of Permutation Testing for Randomized Trial

**Seong Hyeok Moon**

**Economic Research Center  
Department of Economics  
University of Chicago**

April 4, 2011

**First Meeting of Marie Curie International Exchange Program  
St. Justine's Hospital, Montreal, Canada**

## Common Challenges in Evaluating Randomized Trial

- Sample size is usually **small**
- Randomization is commonly **compromised**
- **“Cherry Picking”**
- **Non-compliance/Attrition** is often serious
- Control group is often **contaminated**
- **Intensity of Treatment** may vary across treated individuals
- Difficult to establish **external validity**
- etc etc

## What is Permutation Testing?

- Originated from Fisher(1935) and Pitman(1937, 1938)
- Often used in medical researches
- Introduced to ECI evaluation literature by Heckman, Moon, Pinto et al (2010)
- **Key** : Under the null  $H_0$ , the labels (T or C) must be exchangeable
- Notationally, we write  $(Y, D) \stackrel{d}{=} (Y, gD)$  under the null
- Computation Protocol:
  - 1 Permute the labels
  - 2 Compute test statistic on each of permuted replications
  - 3 Construct a hypothetical distribution of the test statistic
  - 4 Compute p-value which is now exact even with a small sample

Treatment

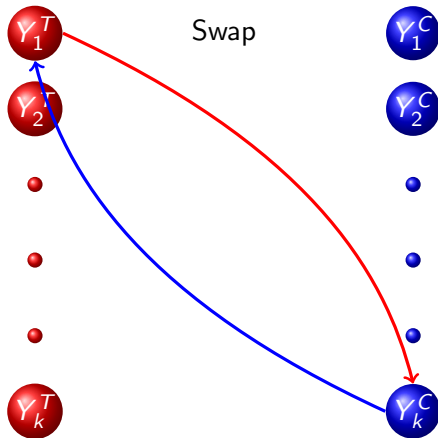


Control



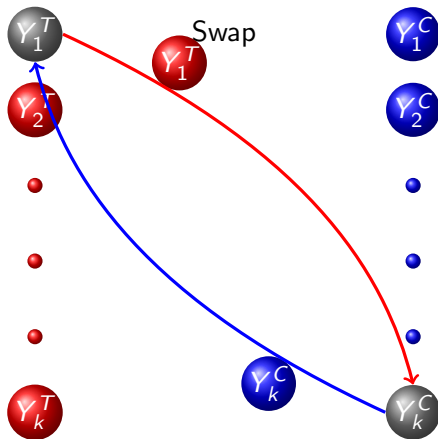
Treatment

Control



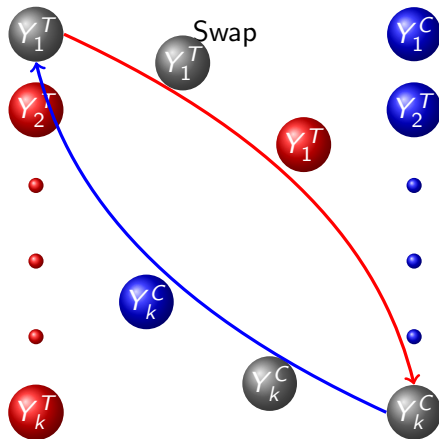
Treatment

Control



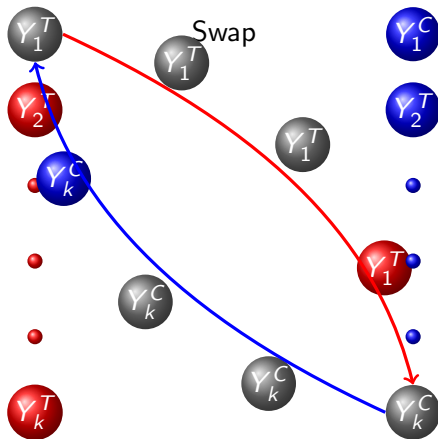
Treatment

Control



Treatment

Control





Treatment



Control



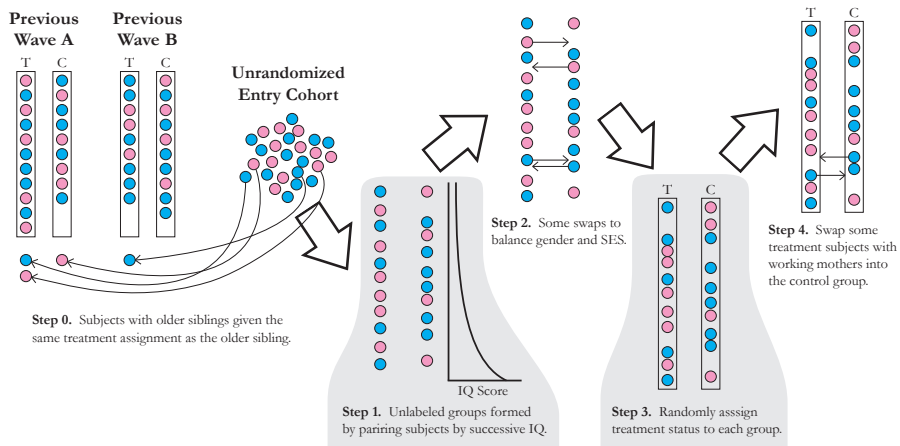
## Why is Permutation Testing useful?

- Produces the exact p-value free from small sample bias
- Free from distributional assumption
- Can be applied to any test statistic
- For technical details, see Heckman, Moon, Pinto et al (2010)

## What if the labels are not exchangeable?

- When the randomization was compromised or corrupted, the labels are not exchangeable so that a simple permutation testing does **NOT** guarantee the exact p-value.
- **Compromised/Corrupted Randomization** is quite common in many social experiments, rather than exceptional.
- Compromise or Corruption in randomization can occur due to various reasons such as
  - 1 Post-randomization Swap (Perry Preschool, HSIS)
  - 2 Non-compliance/Attrition (Abecedarian, Nurse-Family Partnership)
  - 3 Contaminated Controls (Abecedarian, HSIS)

Figure 1: Perry Preschool Program : Randomization Diagram



Source: Heckman, Moon, Pinto et al (2010)

## Imbalance due to Post-randomization Swap : Perry Preschool

**Table 1:** Mother's Activity Status and Father's presence at Study Entry

		<b>Maternal Empl.</b>			<b>Paternal Presence</b>		
		Yes	No	Prob. <sup>a</sup>	Yes	No	Prob. <sup>a</sup>
Males	Trt.	2	31	.06	15	18	.45
	Ctl.	11	28	.28	22	17	.56
Females	Trt.	3	22	.12	17	8	.68
	Ctl.	9	17	.35	11	15	.42

**Notes:** Measured at study entry (age 4 for wave 1; age 3 for other waves).

(a) Conditional probability estimate (# yes / total).

**Source:** Heckman, Moon, Pinto et al (2010)

## Imbalance due to Post-randomization Swap (?) : Abecedarian

**Table 2:** Mother's Activity Status and Father's presence at Birth

### Mother's Activity Status at Birth

	Daycare Trt		Daycare Ctrl	
Working	25	46.3%	15	28.3%
Schooling	21	38.9%	18	34.0%
Neither	8	14.8%	20	37.7%
Total	54		53	

### Father's Presence at Birth

	Daycare Trt		Daycare Ctrl	
Presence	10	18.5%	19	35.8%
Absence	44	81.5%	34	64.2%
Total	54		53	

**Source:** Heckman, Moon, Pinto et al (2011)

## Permutation Testing under Conditional Exchangeability

- The *Conditional Exchangeability* property enables inference of treatment effects under compromised randomization.
- Even though assignments might not be exchangeable across all background measures, they might be exchangeable conditional on the pre-program measures that influenced randomization.
- Notationally, we write  $(Y, D) | X \stackrel{d}{=} (Y, gD) | X$  under the null, where  $X$  is the set of pre-program measures.
- To bypass “curse of dimensionality”, use Freedman and Lane (1983)
- For technical details, see Heckman, Moon, Pinto et al (2010)

Table 3: Program Effect on Adolescent Behavior : Abecedarian, Age 12

Achenbach CBCL : Age 12	Obs		Control	Effect		Plain t-test			Uncond. Perm. Test			Cond. Perm. Test		
	Ctrl	Total	Mean	Uncond.	Cond.	Pr(T<t)	Pr(T> t)	Pr(T>t)	Pr(T<t)	Pr(T> t)	Pr(T>t)	Pr(T<t)	Pr(T> t)	Pr(T>t)
<b>Females</b>														
Internalizing Problem	25	50	57.920	-3.080	-3.099	<b>0.076</b>	0.152	0.924	<b>0.082</b>	0.142	0.924	<b>0.022</b>	<b>0.092</b>	0.978
Somatic Complaint	25	50	62.680	-2.960	-2.663	<b>0.042</b>	<b>0.083</b>	0.958	<b>0.040</b>	<b>0.078</b>	0.964	<b>0.038</b>	<b>0.074</b>	0.962
Schizoid	25	50	62.080	-1.440	-1.028	0.225	0.451	0.775	0.214	0.444	0.798	0.102	0.602	0.898
Externalizing Problem	25	50	59.360	-3.520	-3.603	<b>0.058</b>	0.115	0.942	<b>0.060</b>	0.140	0.944	<b>0.062</b>	<b>0.076</b>	0.938
Delinquent Behavior	25	50	63.960	-2.920	-3.145	<b>0.043</b>	<b>0.087</b>	0.957	<b>0.046</b>	0.118	0.956	<b>0.038</b>	<b>0.044</b>	0.962
Aggressive Behavior	25	50	62.480	-2.920	-2.379	<b>0.063</b>	0.127	0.937	<b>0.056</b>	0.154	0.956	0.178	0.208	0.822
Social Problems	25	49	43.760	1.490	0.621	0.740	0.519	0.260	0.740	0.508	0.270	0.448	0.774	0.552
Social Competence	25	48	40.800	3.635	2.997	0.948	0.104	<b>0.052</b>	0.946	0.118	<b>0.060</b>	0.842	0.182	0.158
School Competence	25	49	34.720	7.988	7.406	0.999	<b>0.003</b>	<b>0.001</b>	0.998	<b>0.006</b>	<b>0.002</b>	0.994	<b>0.006</b>	<b>0.006</b>
Activities	25	50	47.160	1.520	1.334	0.784	0.432	0.216	0.812	0.430	0.196	0.804	0.456	0.196
<b>Males</b>														
Internalizing Problem	22	48	59.909	-2.024	-2.092	0.229	0.458	0.771	0.238	0.438	0.770	<b>0.076</b>	0.476	0.924
Somatic Complaint	22	48	62.045	-0.815	-1.071	0.345	0.690	0.655	0.356	0.690	0.658	0.116	0.594	0.884
Schizoid	22	48	62.136	-1.175	-0.832	0.260	0.520	0.740	0.312	0.564	0.716	0.188	0.594	0.812
Externalizing Problem	22	48	58.182	-1.490	-1.408	0.305	0.610	0.695	0.332	0.594	0.678	0.140	0.630	0.860
Delinquent Behavior	22	48	61.591	-0.976	-1.023	0.300	0.600	0.700	0.330	0.600	0.680	0.146	0.580	0.854
Aggressive Behavior	22	48	61.182	-1.836	-1.779	0.141	0.282	0.859	0.146	0.270	0.866	<b>0.070</b>	0.270	0.930
Social Problems	21	46	42.857	0.263	0.800	0.543	0.913	0.457	0.546	0.934	0.460	0.608	0.716	0.392
Social Competence	20	45	38.800	4.120	4.585	0.932	0.136	<b>0.068</b>	0.938	0.138	<b>0.068</b>	0.978	<b>0.066</b>	<b>0.022</b>
School Competence	22	48	38.773	5.112	5.726	0.972	<b>0.056</b>	<b>0.028</b>	0.976	<b>0.034</b>	<b>0.024</b>	0.996	<b>0.020</b>	<b>0.004</b>
Activities	21	47	43.571	2.544	2.576	0.833	0.334	0.167	0.854	0.336	0.154	0.862	<b>0.020</b>	0.138

Source: Heckman, Moon, and Pinto (2011)



Table 4: Program Effect on Adolescent Behavior : Abecedarian, Age 15

Achenbach CBCL : Age 15	Obs		Control	Effect		Plain t-test			Uncond. Perm. Test			Cond. Perm. Test		
	Ctrl	Total	Mean	Uncond.	Cond.	Pr(T<t)	Pr(T> t)	Pr(T>t)	Pr(T<t)	Pr(T> t)	Pr(T>t)	Pr(T<t)	Pr(T> t)	Pr(T>t)
<b>Females</b>														
Internalizing Problem	27	50	54.519	-2.301	-2.279	0.177	0.353	0.823	0.166	0.346	0.846	<b>0.098</b>	0.318	0.902
Somatic Complaint	27	50	63.630	-3.412	-2.981	<b>0.046</b>	<b>0.092</b>	0.954	<b>0.038</b>	<b>0.070</b>	0.970	<b>0.036</b>	0.118	0.964
Schizoid	27	50	58.778	-0.126	-0.149	0.469	0.937	0.531	0.446	0.930	0.608	0.258	0.928	0.742
Externalizing Problem	27	50	56.111	-4.589	-4.165	<b>0.023</b>	<b>0.045</b>	0.977	<b>0.026</b>	<b>0.046</b>	0.976	<b>0.026</b>	<b>0.044</b>	0.974
Delinquent Behavior	27	50	62.296	-3.731	-3.309	<b>0.009</b>	<b>0.018</b>	0.991	<b>0.010</b>	<b>0.022</b>	0.990	<b>0.024</b>	<b>0.032</b>	0.976
Aggressive Behavior	27	50	59.593	-2.506	-2.108	<b>0.025</b>	<b>0.051</b>	0.975	<b>0.038</b>	<b>0.062</b>	0.964	<b>0.068</b>	<b>0.088</b>	0.932
Social Problems	27	50	44.222	0.604	0.814	0.606	0.788	0.394	0.628	0.788	0.394	0.748	0.678	0.252
Social Competence	26	49	37.923	3.686	3.546	0.943	0.114	<b>0.057</b>	0.956	<b>0.098</b>	<b>0.046</b>	0.920	<b>0.090</b>	<b>0.080</b>
School Competence	26	49	36.962	<b>6.560</b>	<b>6.009</b>	0.991	<b>0.017</b>	<b>0.009</b>	0.994	<b>0.018</b>	<b>0.006</b>	0.964	<b>0.036</b>	<b>0.036</b>
Activities	27	50	42.481	1.171	1.140	0.712	0.577	0.288	0.734	0.550	0.274	0.640	0.546	0.360
<b>Males</b>														
Internalizing Problem	22	50	57.227	-1.370	-0.740	0.302	0.604	0.698	0.330	0.638	0.684	0.228	0.788	0.772
Somatic Complaint	22	50	61.045	-1.438	-1.624	0.212	0.424	0.788	0.226	0.426	0.790	<b>0.050</b>	0.390	0.950
Schizoid	22	50	60.364	-0.578	-0.341	0.378	0.757	0.622	0.412	0.780	0.606	0.292	0.850	0.708
Externalizing Problem	22	50	56.773	-1.487	-1.011	0.282	0.565	0.718	0.296	0.588	0.724	0.262	0.736	0.738
Delinquent Behavior	22	50	61.227	0.023	-0.045	0.505	0.990	0.495	0.524	1.000	0.492	0.386	0.982	0.614
Aggressive Behavior	22	50	58.818	-0.140	0.041	0.461	0.923	0.539	0.470	0.934	0.550	0.480	0.978	0.520
Social Problems	22	50	43.864	2.494	1.400	0.876	0.249	0.124	0.878	0.240	0.128	0.696	0.518	0.304
Social Competence	22	50	39.000	3.571	2.890	0.900	0.199	<b>0.100</b>	0.930	0.202	<b>0.080</b>	0.898	0.248	0.102
School Competence	22	50	36.773	5.120	5.679	0.951	<b>0.098</b>	<b>0.049</b>	0.950	0.118	<b>0.054</b>	0.984	<b>0.042</b>	<b>0.016</b>
Activities	22	50	43.273	0.799	0.400	0.615	0.770	0.385	0.628	0.774	0.384	0.644	0.842	0.356

Source: Heckman, Moon, and Pinto (2011)

Table 5: Program Effect on Adult Outcomes : Perry Preschool, Females

Outcome	Age	Effect				p-Values					Available Observations
		Ctl. Mean	Uncond. <sup>b</sup>	Cond. (Full) <sup>c</sup>	Cond. (Part.) <sup>d</sup>	Naïve <sup>e</sup>	Full Lin. <sup>f</sup>	Partial Lin. <sup>g</sup>	Part. Lin. (adj.) <sup>h</sup>	Gender D-in-D <sup>i</sup>	
<b>Education</b>											
Mentally Impaired?	≤19	0.36	-0.28	-0.29	-0.31	<b>0.008</b>	<b>0.009</b>	<b>0.005</b>	<b>0.017</b>	0.337	46
Learning Disabled?	≤19	0.14	-0.14	-0.15	-0.16	<b>0.009</b>	<b>0.016</b>	<b>0.009</b>	<b>0.025</b>	<b>0.029</b>	46
Yrs. of Special Services	≤14	0.46	-0.26	-0.29	-0.34	<b>0.036</b>	<b>0.013</b>	<b>0.013</b>	<b>0.025</b>	0.153	51
Yrs. in Disciplinary Program	≤19	0.36	-0.24	-0.19	-0.27	<b>0.089</b>	0.127	<b>0.074</b>	<b>0.074</b>	0.945	46
High School Graduation	19	0.23	0.61	0.49	0.56	<b>0.000</b>	<b>0.000</b>	<b>0.000</b>	<b>0.000</b>	<b>0.003</b>	51
Grade Point Average	19	1.53	0.89	0.88	0.95	<b>0.000</b>	<b>0.001</b>	<b>0.000</b>	<b>0.001</b>	<b>0.009</b>	30
Highest Grade Completed	19	10.75	1.01	0.94	1.19	<b>0.007</b>	<b>0.008</b>	<b>0.002</b>	<b>0.006</b>	<b>0.052</b>	49
# Years Held Back	≤19	0.41	-0.20	-0.14	-0.21	<b>0.067</b>	0.135	<b>0.097</b>	0.178	0.106	46
Vocational Training Certificate	≤40	0.08	0.16	0.13	0.16	<b>0.070</b>	0.106	0.107	0.107	0.500	51
<b>Health</b>											
No Health Problems	19	0.83	0.05	0.12	0.07	0.265	0.107	0.137	0.576	0.308	49
Alive	40	0.92	0.04	0.04	0.06	0.273	0.249	0.197	0.675	0.909	51
No Treat. for Illness, Past 5 Yrs.	27	0.59	0.05	0.14	0.10	0.369	0.188	0.241	0.690	0.806	47
No Non-Routine Care, Past Yr.	27	0.00	0.04	0.02	0.03	0.484	0.439	0.488	0.896	0.549	44
No Sick Days in Bed, Past Yr.	27	0.45	-0.05	-0.04	0.06	0.623	0.597	0.529	0.781	0.412	47
No Doctors for Illness, Past Yr.	19	0.54	-0.02	-0.01	-0.05	0.559	0.539	0.549	0.549	0.609	49
No Tobacco Use	27	0.41	0.11	0.08	0.08	0.208	0.348	0.298	0.598	0.965	47
Infrequent Alcohol Use	27	0.67	0.17	0.07	0.12	0.103	0.336	0.374	0.587	0.924	45
Routine Annual Health Exam	27	0.86	-0.06	-0.09	-0.05	0.684	0.751	0.727	0.727	0.867	47
<b>Family</b>											
Has Any Children	≤19	0.52	-0.12	-0.05	-0.07	0.218	0.419	0.328	0.601	—	48
# Out-of-Wedlock Births	≤40	2.52	-0.29	0.51	0.05	0.652	0.257	0.402	0.402	—	42

Source: Heckman, Moon, Pinto et al (2010)

**Table 6: Program Effect on Adult Outcomes : Perry Preschool, Males**

Outcome	Age	Effect				p-Values					Available Observations
		Ctl. Mean	Uncond. <sup>b</sup>	Cond. (Full) <sup>c</sup>	Cond. (Part.) <sup>d</sup>	Naïve <sup>e</sup>	Full Lin. <sup>f</sup>	Partial Lin. <sup>g</sup>	Part. Lin. (adj.) <sup>h</sup>	Gender D-in-D <sup>i</sup>	
Any Non-Juv. Arrests	≤40	0.92	-0.14	-0.12	-0.12	<b>0.090</b>	0.124	<b>0.078</b>	0.192	0.463	72
Any Fel. Arrests	≤40	0.44	-0.16	-0.15	-0.16	<b>0.047</b>	0.133	<b>0.083</b>	0.191	—	72
Any Arrests	≤40	0.95	-0.13	-0.11	-0.09	<b>0.072</b>	0.142	0.123	0.181	0.824	72
Any Misd. Arrests	≤40	0.87	-0.11	-0.08	-0.07	0.166	0.281	0.191	0.191	0.519	72
# Misd. Arrests	≤40	8.46	-3.13	-3.42	-3.64	<b>0.037</b>	<b>0.043</b>	<b>0.021</b>	<b>0.039</b>	0.549	72
# Non-Juv. Arrests	≤40	11.72	-4.26	-4.45	-4.85	<b>0.039</b>	<b>0.053</b>	<b>0.025</b>	<b>0.041</b>	0.458	72
# Total Arrests	≤40	12.41	-4.20	-4.44	-4.88	<b>0.056</b>	<b>0.073</b>	<b>0.036</b>	<b>0.053</b>	0.566	72
# Fel. Arrests	≤40	3.26	-1.14	-1.03	-1.20	0.112	0.173	<b>0.092</b>	<b>0.092</b>	—	72
# Non-Victimless Charges <sup>j</sup>	≤40	3.08	1.59	1.65	1.65	<b>0.029</b>	<b>0.048</b>	<b>0.027</b>	<b>0.061</b>	0.175	72
# Total Charges	≤40	13.38	4.38	5.08	5.08	<b>0.063</b>	<b>0.081</b>	<b>0.041</b>	<b>0.075</b>	0.637	72
Total Crime Cost <sup>k</sup>	≤40	775.90	-351.22	-515.10	-515.10	0.153	0.108	<b>0.070</b>	<b>0.070</b>	0.858	72
Any Non-Victimless Charges <sup>l</sup>	≤40	0.62	0.16	0.15	0.15	0.105	0.179	0.112	0.259	0.957	72
Ever Incarcerated	≤40	0.23	-0.08	-0.11	-0.12	0.260	0.159	0.114	0.202	0.563	72
Any Charges	≤40	0.95	0.13	0.09	0.09	<b>0.072</b>	0.142	0.125	0.125	0.799	72

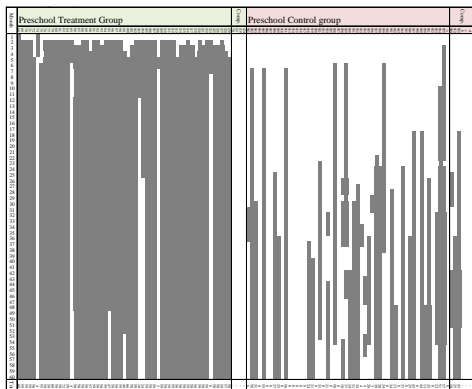
**Source:** Heckman, Moon, Pinto et al (2010)

## Challenges : Revisited and Unvisited

- Sample size is usually **small**
- Randomization is commonly **compromised**
- **“Cherry Picking”**
- **Non-compliance/Attrition** is often serious
- Control group is often **contaminated**
- **Intensity of Treatment** may vary across treated individuals
- Difficult to establish **external validity**
- etc etc

## Contaminated Controls : Abecedarian

Figure 2: Length of Daycare Treatment among ABC Controls



**Source:** Heckman, Moon, and Pinto (2011)

## Contaminated Controls : Head Start Impact Study

Table 7: Preschool Treatment among Controls

Child Care Arrangement	Treatment	Control
Head Start	76.7	13.8
Non-Head Start center	11.1	35.3
Home of non-relative	1.5	5.8
Home of relative	0.9	3.0
Own home with relative	0.4	2.3
Own home with non-relative	0.3	0.1
Parent care	9.1	39.7
Total	100.0%	100.0%

**Source:** Moon and Pinto (2011)

## Advertisement



ABOUT  
THE MFI

PEOPLE

EVENTS

RESEARCH

DONATE

CONTACT

SEARCH

## Human Capital and Economic Opportunity: A Global Working Group

Home

Members

Working Groups

Events

Working Papers

Contact

### About the Working Group

The Human Capital and Economic Opportunity Working Group is a global network of researchers from multiple disciplines that seeks to restore and strengthen human capital development research.

It establishes a community of researchers, educators and policymakers focused on human development (including health and psychology) and related issues in finance, macroeconomics and inequality, with the aim of to promoting easy, effective collaboration.

The working group will focus on the failure of markets to accurately price the economic value of human capital and the consequences of these failures, notably economic development distortions, declining growth, rising economic inequality and difficulty correcting fiscal imbalances. Currently seven topical working groups have been established to explore specific aspects of these issues.  
[more »](#)

### Human Capital Events

#### First Working Group Meeting Set for March 15

The first meeting of the INET Global Human Capital and Economic Opportunity Working Group was held by teleconference on Tuesday, March 15.  
[Hear conference recording »](#)

### News

#### Obama Names Olopade to National Cancer Advisory Board

Group member Funmi Olopade is among five named to the board that advises the heads of HHS and NCI.

#### Conference Explores Individual Differences, Economic Behavior

Economists and psychologists gathered Dec. 10–11 for the third meeting in this series.

### Supported by

Institute for New  
Economic Thinking



From the Director

© 2010 [The University of Chicago](#)  
1126 East 59th Street, Chicago IL 60637  
[mfi@uchicago.edu](mailto:mfi@uchicago.edu) | 773.702.5599



THE UNIVERSITY OF  
CHICAGO

Webpage: [www.hceconomics.org](http://www.hceconomics.org)





ABOUT  
THE MFI

PEOPLE

EVENTS

RESEARCH

DONATE

CONTACT

SEARCH

## Human Capital and Economic Opportunity: A Global Working Group

Home

Members

Working Groups

Events

Working Papers

Contact



**Early Childhood Interventions (ECI)**  
To investigate the early origins of inequality and their lifetime consequences.



**Family Inequality (FI)**  
FI will focus on the relationship between the family and the sources of inequality.



**Health Inequality (HI)**  
HI will build a comprehensive framework for the emergence and the evolution of health disparities.



**Identity and Psychology (IP)**  
To understand the role of personality in all of its varied forms on shaping and perpetuating inequality and to understand the evolution of personality over the lifecycle.



**Markets (M)**  
The interface between markets for financing human capital and markets for physical capital.



**Measuring and Interpreting Inequality (MIE)**  
To understand and interpret frameworks for inequality in all of its manifestations using economic theory and the philosophy of social justice.



**Youth Human Capital and Economic Development (YHC)**  
Investing in the intersection between youth human capital and economic growth.