



工作场所健康项目 效果评估

随机对照试验的实证应用



参考文献: Jones, Molitor & Reif (2019)



讲课人: 何濯羽 (国际教育学院)

目 录

1

工作场所健康项目

2

研究问题与困难

3

随机对照试验

4

研究结果：哪类人参与了WWP

5

因果推断：WWP的作用

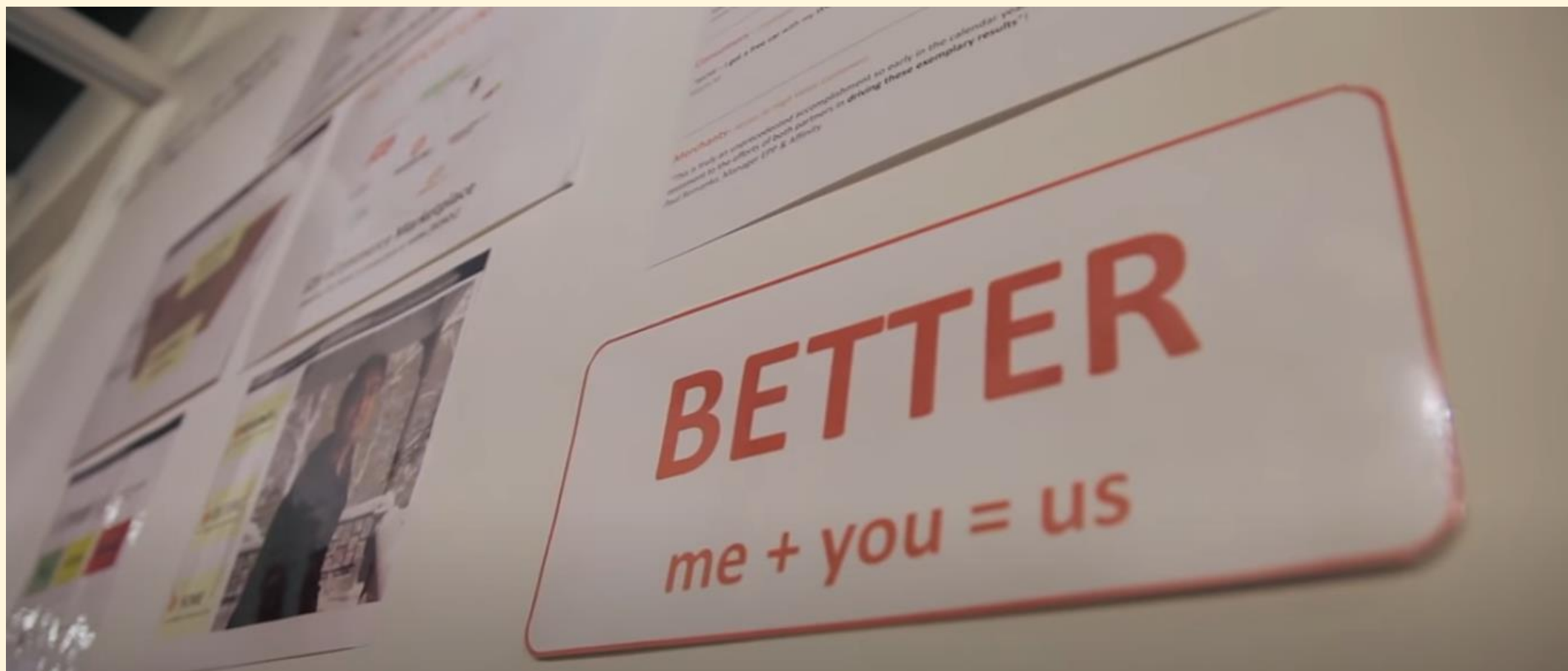


1

工作场所
健康
项目

什么是工作场所健康项目？

观看视频：约翰霍普金斯大学的调研



什么是工作场所健康项目？



Workplace Wellness Program

摘自美国俄克拉荷马州塔尔萨卫生局官网：

如今，美国工人每天有超过三分之一的时间花在工作上，雇主在促进员工的健康和安全方面处于独特的地位。使用有效的**工作场所健康项目**和政策可以降低健康风险并提高员工的生活质量。此外，这些项目还能增强同事友情和团队合作能力，同时降低与健康相关的成本，并且可以成为一个极好的招聘工具。

工作场所健康项目为何在美国流行？



雇主们相信该项目可以降低员工的医疗开销并提高员工的生产力。



雇主们利用该项目吸引求职者，留住已就职的员工。





2

研究

问题

与困难

工作场所健康项目真的有效吗？

- 可以降低员工的医疗开销吗？
- 可以提高员工的生产力吗？
- 可以降低员工的缺勤率，降低员工离职率，促进员工的升职率吗？
- 可以促使员工采取更健康的生活方式吗？

传统实证方法的漏洞

方法1：平均值求差

$$\frac{1}{N_1} \sum_{i=1}^{N_1} Y_i - \frac{1}{N_0} \sum_{i=1}^{N_0} Y_i$$

参与WWP的结果

不参与WWP的结果

方法2：普通最小二乘回归模型

$$Y_i = \alpha + \gamma P_i + \Gamma \mathbf{X}_i + \varepsilon_i$$

指标变量：是否参与了WWP

漏洞：选择性偏差 (Selection Bias)

选择性偏差是指主动进入研究样本中的人与未主动进入样本中的人之间存在着特征上差异的情况。

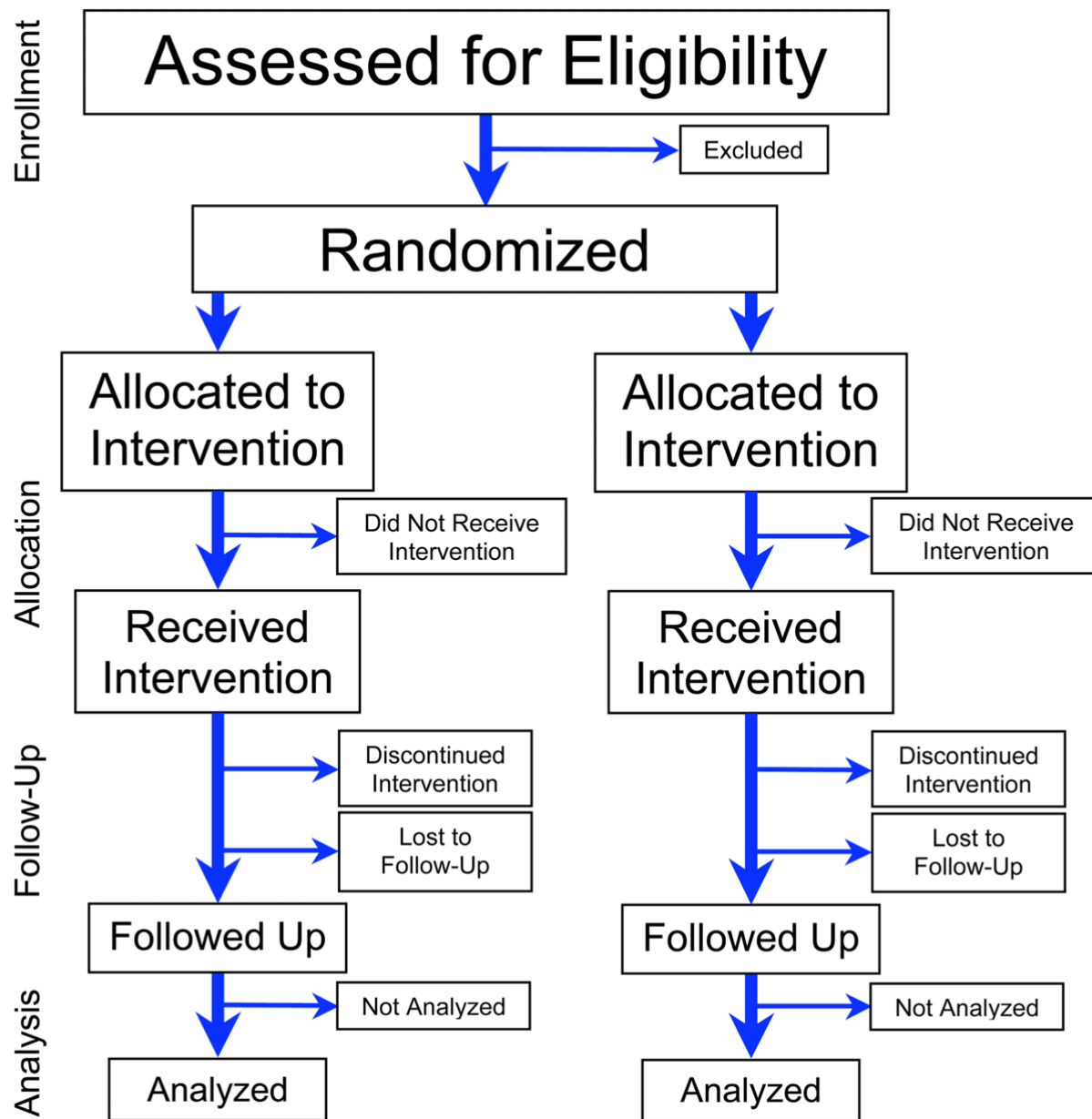
例：参与WWP的人可能原本就比不参与WWP的人更加重视个人健康，因而原本就拥有更好的身体状况，医疗开销也更少。



3

随机
对照
试验

随机对照试验 (Randomized Control Trial)



RCT

随机对照试验是一种源自医学临床研究的科学实验方法，常用于比较特定药物、手术、医疗设备、饮食或其他医学治疗手段的效果。

随机对照试验的参与者之间存在已知和未知的差异，这些差异可能会影响研究结果，但无法直接控制。如果设计良好、实施得当并招募到足够多的参与者，随机对照试验可以对这些混杂因素进行充分的控制，从而对其所研究的对象进行有效且无偏差的比较。

伊利诺伊大学香槟分校的随机对照试验



David Molitor

Laura Payne

Julian Reif

伊利诺伊州工作场所健康研究 (Illinois Workplace Wellness Study) 是一场大规模的随机对照试验，旨在调查工作场所健康项目对员工医疗支出、生产力和个人健康的影响。作为研究的一部分，三位伊利诺伊大学的学者与校园健康服务总监合作，在**伊利诺伊大学厄巴纳-香槟分校 (UIUC)** 设计并推出了名为 **iThrive** 的综合健康项目。他们的目标是创建一个具有代表性的工作场所健康项目，其中包括健康专家推荐的所有关键组成部分：生物识别筛查 (biometric screening)、健康风险评估 (health risk assessment, HRA)、各种健康活动 (健身房健身、马拉松比赛等)、金钱激励、带薪休假。



伊利诺伊大学厄巴纳-香槟分校 (UIUC) 是一所位于美国伊利诺伊州香槟-厄巴纳都会区的公立研究型大学，成立于1867年。截至2023年秋季学期，它拥有56000多名学生，其中有40000多名国际学生，是美国招生规模最大的公立大学之一。

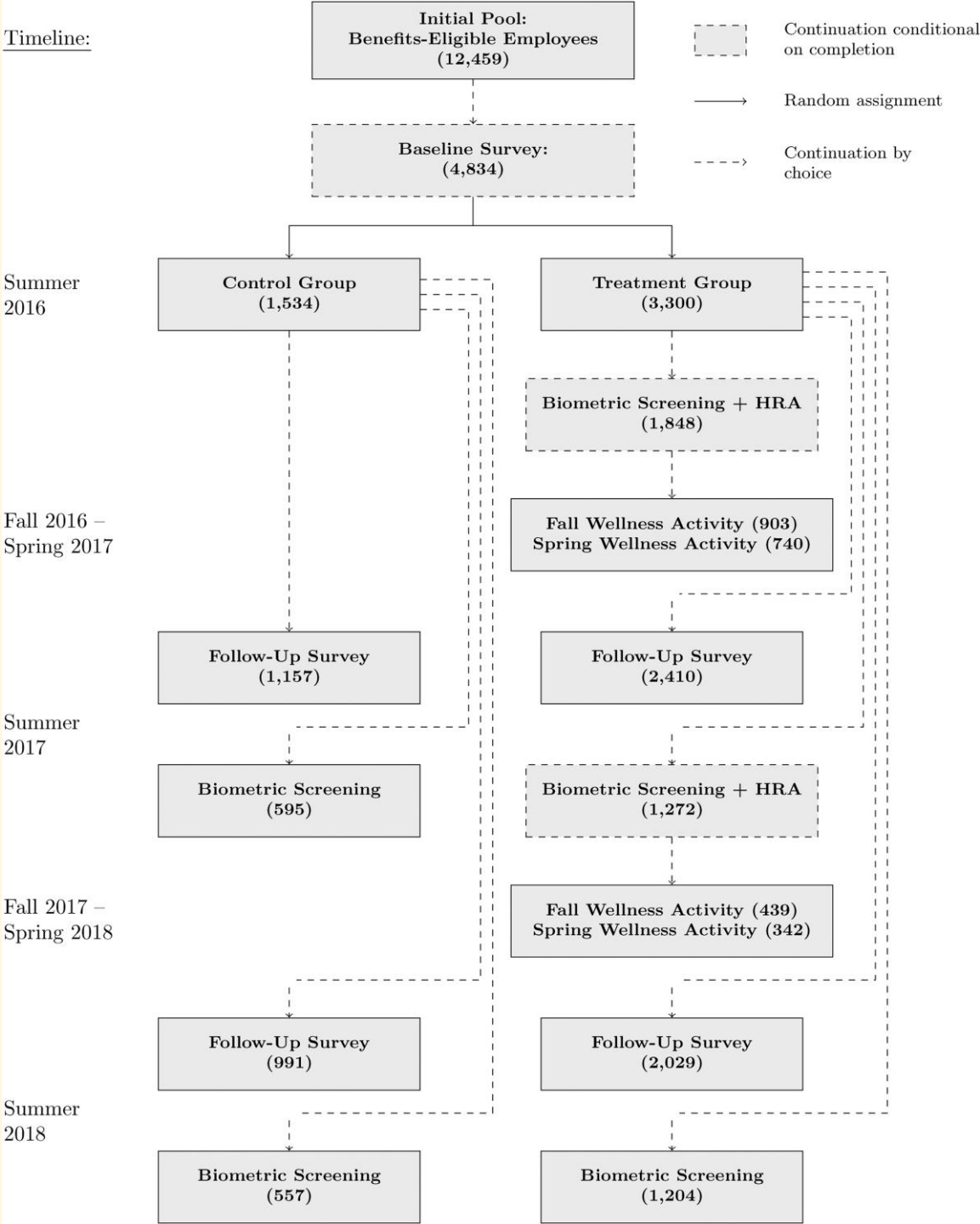
学校官网: <https://illinois.edu>



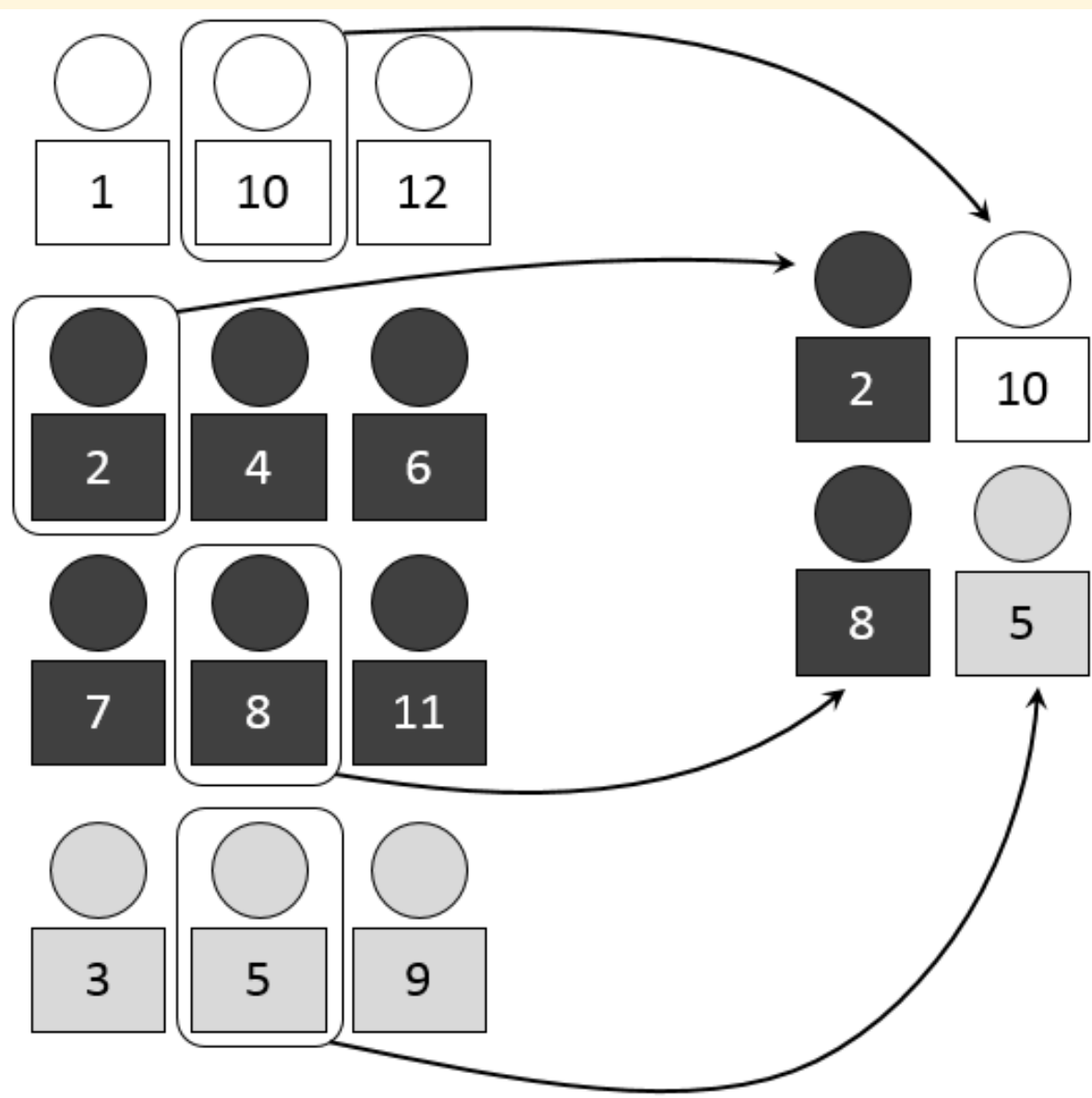
随机对照试验的设计

- 1) 2016年8月-9月，伊利诺伊大学的员工可以在校园多个地点接受**生物识别筛查**。筛查结束的几天后，他们收到一封电子邮件邀请函，请他们完成旨在评估他们的生活习惯的在线**健康风险评估（HRA）**。完成后，参与者会收到生物特征筛查的结果，并获得个人健康的改善建议。只有完成生物识别筛选和HRA的员工才有资格参加该试验的第二步。
- 2) 符合资格的参与者有机会参加**秋季学期**和**春季学期**的多项**健康活动**。参加春季活动的资格并不取决于秋季活动的注册或完成。
健康活动包括：慢性病管理、体重管理、太极拳、健康的工作习惯等线下课程；线上自定进度的健康挑战；戒烟热线。
活动时长从6周到12周不等，“完成活动”指的是参加至少四分之三时长的活动。选择活动时，参与者被鼓励要根据HRA的反馈进行选择。
- 3) 参与者完成每一步都会获得**金钱奖励**，这些奖励根据他们被分配到的**实验组（treatment group）**而有所不同。实验组A、B、C中的个体在第一年完成生物识别筛查和HRA时分别获得\$0、\$100、\$200的奖励。根据每项完成的健康活动（每学期最多一次），实验组被进一步划分：\$25或\$75的奖励。因此，总共有六个治疗组。第一年完成所有步骤（两个学期的生物识别筛查、HRA、健康活动）的总奖励为\$50至\$350不等，取决于参与者所在的实验组。

Timeline:



分层随机化



分层随机化 (stratified randomization)

随机对照试验中，参与者被分配到每个组的概率是相等的。随机化是按员工类别（教学岗、行政岗等）、性别、年龄、年薪四分位数和种族进行分层的。

实证分析的数据来源

伊利诺伊大学香槟分校行政管理数据

大学员工的个人信息、联系方式、升迁或离职记录、病假天数、使用运动馆次数。



线上问卷数据

大学员工参加 iThrive 前的健康状况、医疗服务使用情况、职位满意度、工作能力。



健康保险索赔数据

大学员工的医疗开销。



伊利诺伊州马拉松数据

大学员工是否参加马拉松比赛，以及参加者的比赛结果。

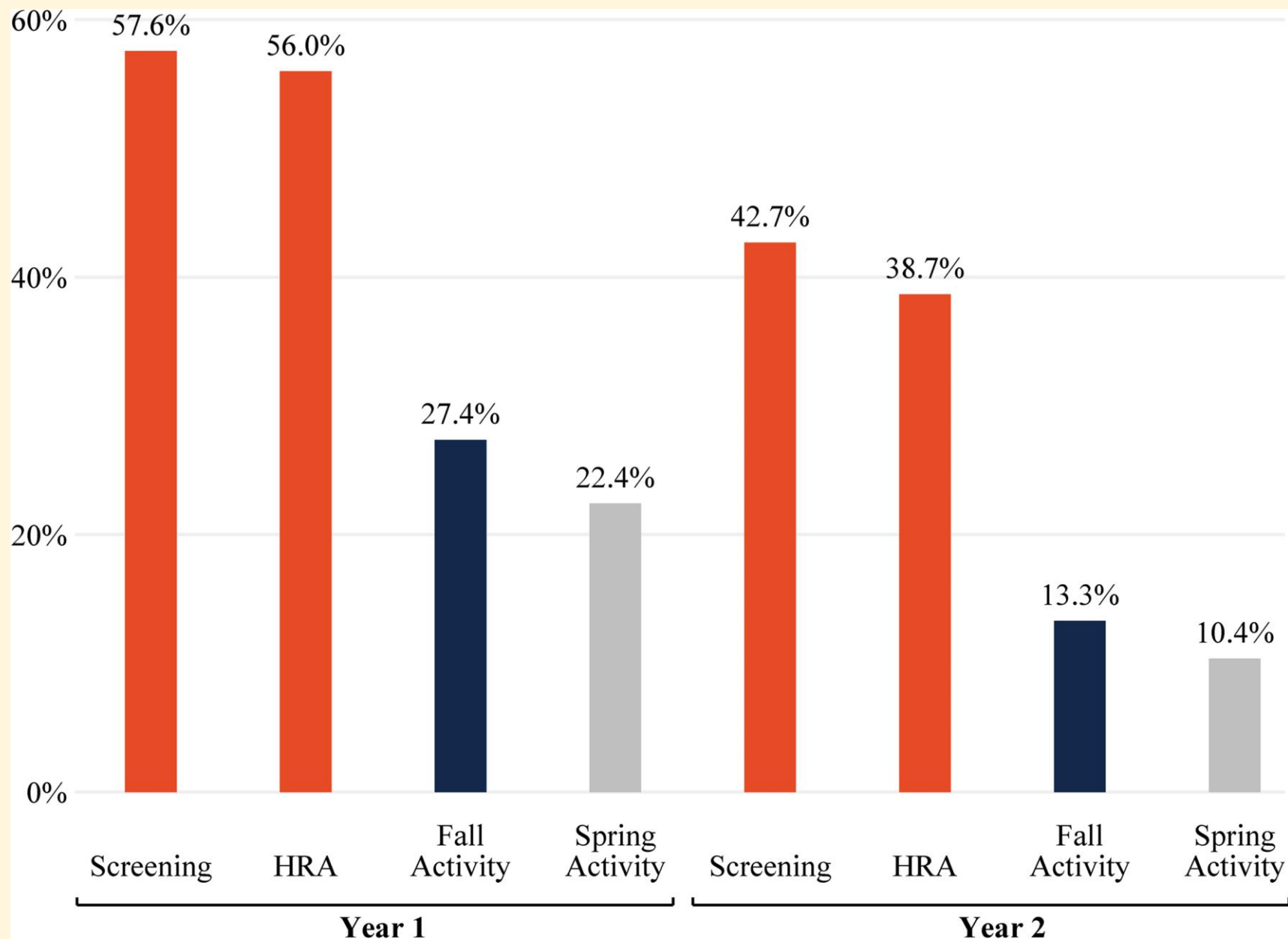


实验组与对照组的相似度

Means of Study Variables at Baseline					
	Enrolled in study				
	Not in			<i>p</i> -	Sample
	study	Control	Treatment	value	size
	(1)	(2)	(3)	(4)	(5)
Panel A: Stratification variables					
Male	0.536	0.426	0.428	.902	12,459
Age 50+	0.430	0.323	0.327	.818	12,459
Age 37–49	0.362	0.340	0.332	.591	12,459
White	0.774	0.841	0.836	.648	12,459
Salary Q1 (bottom quartile)	0.234	0.244	0.242	.881	12,459
Salary Q2	0.189	0.255	0.259	.773	12,459
Salary Q3	0.197	0.249	0.250	.924	12,459
Faculty	0.298	0.196	0.201	.721	12,459
Academic staff	0.324	0.443	0.437	.712	12,459

- 列(1)-(3)展示了大学员工的平均值。
- 列(4)展示了第(2)-(3)列中的两个系数相等的联合假设检验的p值。

iThrive 的参与度



$$3300 \times 56\% = 1848$$

在第一年，56%的实验组成员（1848名成员）完成了生物识别筛查和健康风险评估。



4

研究结果：

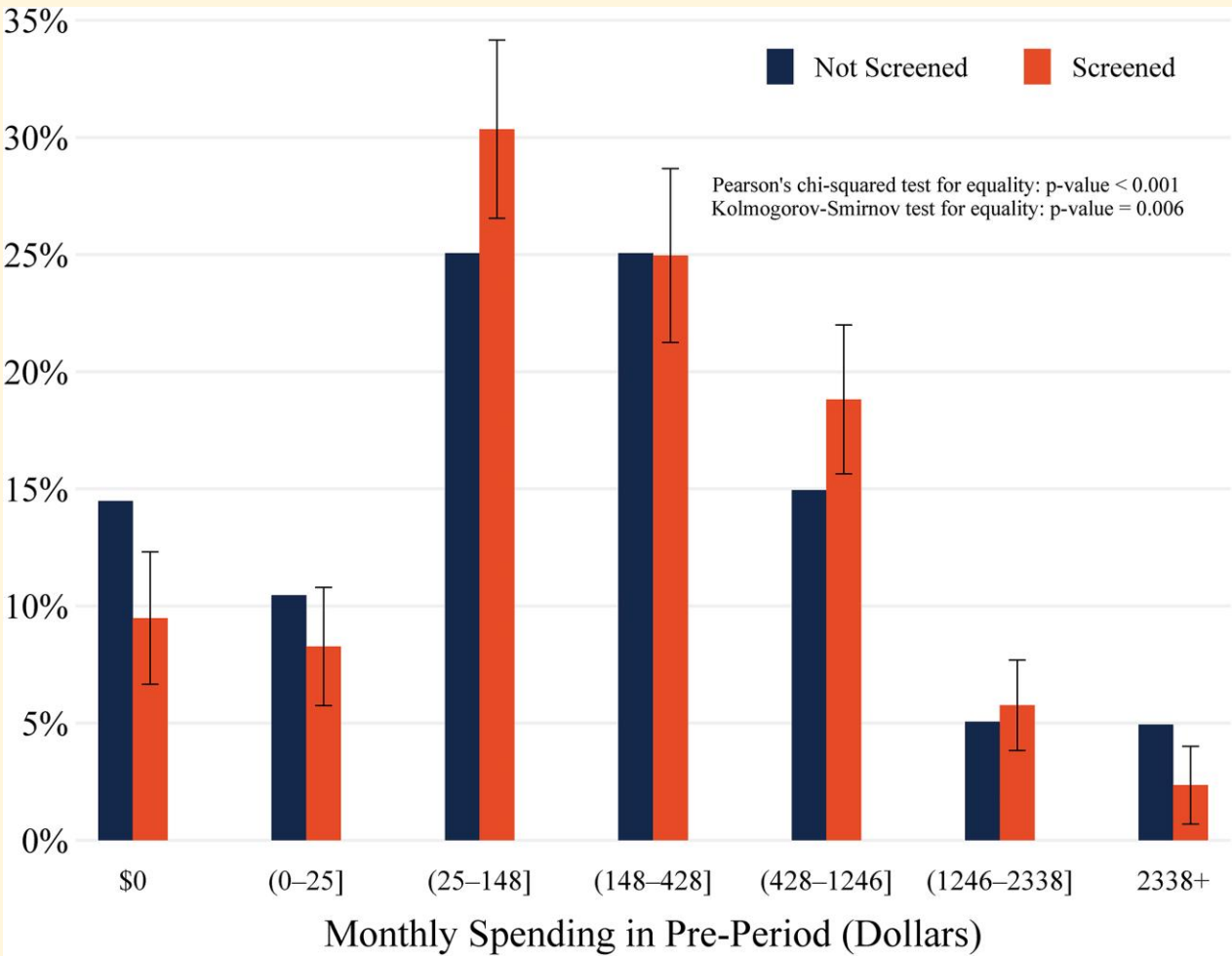
哪类人

参与了WWP

负担中等医疗开销的员工更愿意参与 iThrive

Selection variable	Mean	N	Completed		Completed
			screening	Completed	spring
			and HRA	fall activity	activity
	(1)	(2)	(3)	(4)	(5)
Panel A: Baseline medical spending					
Total spending (dollars/month)	479	2,188	-115.3**	-60.6	-62.5
[admin]			(52.2)	(43.6)	(44.3)
			[.082]	[.405]	[.273]
Nonzero medical spending [admin]	0.885	2,188	0.050***	0.049***	0.046***
			(0.014)	(0.014)	(0.014)
			[.008]	[.005]	[.020]

列(3)-(5)展示了完成活动的人和未完成活动的人之间的均值差异。



实验组成员的干预前医疗支出（按参与活动进行分组）

生产力低、请假频繁、薪资偏低的员工更愿意参与 iThrive

			Completed		Completed
			screening	Completed	spring
Selection variable	Mean	N	and HRA	fall activity	activity
Productivity index [survey/admin]	0.008	3,251	−0.077	−0.099**	−0.104**
			(0.047)	(0.050)	(0.052)
			[.096]	[.046]	[.044]
Sick leave (days/year) [admin]	6.274	3,296	0.473*	0.705**	0.617**
			(0.267)	(0.290)	(0.312)
			[.144]	[.015]	[.048]
Worked 50+ hours/week [survey]	0.173	3,297	−0.058***	−0.065***	−0.064***
			(0.013)	(0.014)	(0.014)
			[.000]	[.000]	[.000]
Annual salary (dollars) [admin]	61,736	3,257	−782.7	−3,363.9***	−3,429.1***
			(1,248.3)	(1,191.6)	(1,251.8)
			[.519]	[.009]	[.012]
Salary Q1 (bottom quartile) [admin]	0.242	3,300	−0.069***	−0.022	−0.036**
			(0.015)	(0.016)	(0.017)
			[.000]	[.398]	[.121]

逆向选择!
Adverse Selection!

参加过马拉松、频繁光顾校园运动馆的员工更愿意参与 iThrive

			Completed		Completed
			screening	Completed	spring
Selection variable	Mean	N	and HRA	fall activity	activity
IL Marathon/10K/5K (2014–2016)	0.118	3,300	0.089***	0.111***	0.090***
[admin]			(0.011)	(0.014)	(0.016)
			[.000]	[.000]	[.000]
Campus gym visits (days/year)	6.780	3,300	2.178**	1.006	1.629
[admin]			(0.885)	(1.024)	(1.132)
			[.013]	[.328]	[.153]

优势选择!
Advantageous Selection!



5

因果推断： WWP 的作用



$$Y_i = \alpha + \gamma T_i + \Gamma \mathbf{X}_i + \varepsilon_i$$

指标变量：是否属于实验组

识别假设： T_i 与 ε_i 无关，即

$$E(\varepsilon_i | T_i = 1, \mathbf{X}_i) = E(\varepsilon_i | T_i = 0, \mathbf{X}_i)$$

RCT 确保了该等号的成立！

强大的RCT！

因果推断：WWP对员工的医疗开销没有显著影响

Outcome variable	First year (12 months)			Longer-run (24–30 months)		
	Mean (1)	No controls (2)	Post-Lasso (3)	Mean (4)	No controls (5)	Post-Lasso (6)
Panel A: Medical spending						
Total spending (dollars/month) [admin]	576.2	10.8 (48.5) [.937]	34.9 (36.9) [.859]	650.5	−74.7 (58.5) [.618]	−39.7 (47.9) [.754]
	<i>N</i> = 3,239	<i>N</i> = 3,239	<i>N</i> = 3,152	<i>N</i> = 3,307	<i>N</i> = 3,307	<i>N</i> = 3,155
Drug spending [admin]	132.0	−8.5 (26.5) [.937]	−6.1 (12.0) [.947]	148.8	−25.2 (27.7) [.836]	−22.2 (16.4) [.589]
	<i>N</i> = 3,239	<i>N</i> = 3,239	<i>N</i> = 3,152	<i>N</i> = 3,307	<i>N</i> = 3,307	<i>N</i> = 3,155
Office spending [admin]	69.5	−6.1 (10.0) [.937]	−2.0 (4.4) [.947]	74.2	−6.6 (8.6) [.836]	−4.8 (5.4) [.754]
	<i>N</i> = 3,239	<i>N</i> = 3,239	<i>N</i> = 3,152	<i>N</i> = 3,307	<i>N</i> = 3,307	<i>N</i> = 3,155
Hospital spending [admin]	313.0	22.2 (30.9) [.937]	24.6 (28.1) [.868]	353.5	−31.7 (35.6) [.836]	−20.3 (31.9) [.754]
	<i>N</i> = 3,239	<i>N</i> = 3,239	<i>N</i> = 3,152	<i>N</i> = 3,307	<i>N</i> = 3,307	<i>N</i> = 3,155
Nonzero medical spending [admin]	0.902	−0.008 (0.011) [.937]	0.002 (0.010) [.947]	0.950	0.005 (0.008) [.836]	0.007 (0.007) [.754]
	<i>N</i> = 3,239	<i>N</i> = 3,239	<i>N</i> = 3,152	<i>N</i> = 3,307	<i>N</i> = 3,307	<i>N</i> = 3,155

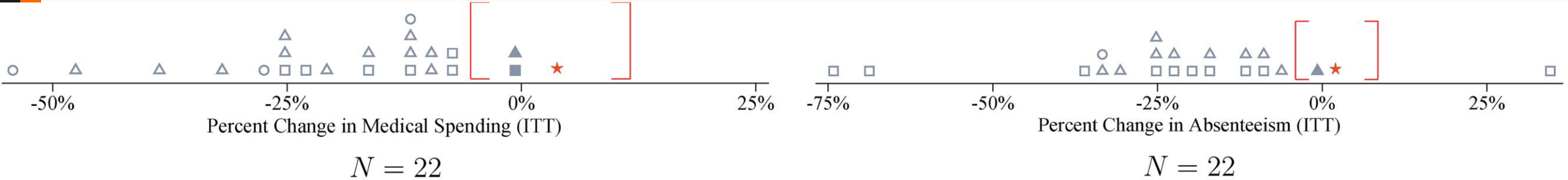
因果推断：WWP对员工的生产力没有显著影响

Outcome variable	First year (12 months)			Longer-run (24–30 months)		
	Mean (1)	No controls (2)	Post-Lasso (3)	Mean (4)	No controls (5)	Post-Lasso (6)
Panel B: Employment and Productivity						
Job promotion [admin]	0.176	−0.003 (0.013) [.952]	−0.004 (0.012) [.922]	0.360	0.006 (0.017) [.996]	0.006 (0.016) [.996]
Job terminated [admin]	<i>N</i> = 4,146 0.113	<i>N</i> = 4,146 −0.013 (0.010) [.630]	<i>N</i> = 4,130 −0.012 (0.009) [.571]	<i>N</i> = 3,635 0.204	<i>N</i> = 3,635 0.002 (0.012) [1.000]	<i>N</i> = 3,619 0.002 (0.012) [.998]
Sick leave (days/year) [admin]	<i>N</i> = 4,834 6.341	<i>N</i> = 4,834 0.186 (0.230) [.816]	<i>N</i> = 4,753 0.138 (0.200) [.880]	<i>N</i> = 4,834 6.066	<i>N</i> = 4,834 0.013 (0.204) [1.000]	<i>N</i> = 4,753 0.018 (0.169) [.998]
Management priority on health/safety [survey]	<i>N</i> = 4,782 0.790	<i>N</i> = 4,782 0.057*** (0.015) [.001]	<i>N</i> = 4,712 0.050*** (0.014) [.003]	<i>N</i> = 4,782 0.784	<i>N</i> = 4,782 0.028* (0.016) [.539]	<i>N</i> = 4,712 0.024 (0.015) [.632]
Productivity index [survey/admin]	<i>N</i> = 3,566 0.000	<i>N</i> = 3,566 −0.046 (0.061) [.450]	<i>N</i> = 3,514 −0.060 (0.056) [.283]	<i>N</i> = 3,018 0.000	<i>N</i> = 3,018 −0.015 (0.062) [.805]	<i>N</i> = 2,976 −0.054 (0.056) [.328]
	<i>N</i> = 3,309	<i>N</i> = 3,309	<i>N</i> = 3,300	<i>N</i> = 2,890	<i>N</i> = 2,890	<i>N</i> = 2,881

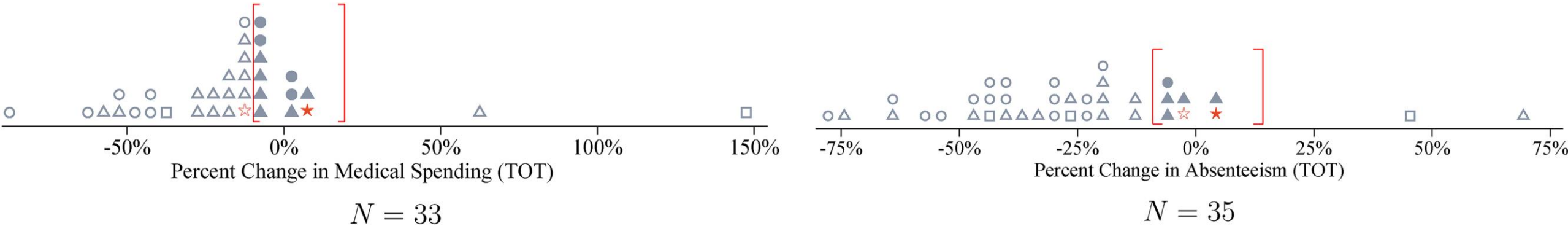
因果推断：WWP对员工的身体健康没有显著影响

Outcome variable	First year (12 months)			Longer-run (24–30 months)		
	Mean (1)	No controls (2)	Post-Lasso (3)	Mean (4)	No controls (5)	Post-Lasso (6)
Panel C: Health status and behaviors						
IL Marathon/10K/5K [admin]	0.066	0.002 (0.008) [.975]	−0.005 (0.006) [.471]	0.052	0.006 (0.007) [.625]	0.001 (0.006) [.995]
	<i>N</i> = 4,834	<i>N</i> = 4,834	<i>N</i> = 4,817	<i>N</i> = 4,834	<i>N</i> = 4,834	<i>N</i> = 4,817
Campus gym visits (days/year) [admin]	5.839	−0.062 (0.733) [.975]	0.401 (0.360) [.471]	5.047	−0.342 (0.660) [.625]	0.001 (0.367) [.998]
	<i>N</i> = 4,834	<i>N</i> = 4,834	<i>N</i> = 4,817	<i>N</i> = 4,834	<i>N</i> = 4,834	<i>N</i> = 4,817
Ever screened [survey]	0.942	0.039*** (0.009) [.001]	0.036*** (0.008) [.000]	0.962	0.029*** (0.008) [.006]	0.027*** (0.007) [.005]
	<i>N</i> = 3,567	<i>N</i> = 3,567	<i>N</i> = 3,557	<i>N</i> = 3,020	<i>N</i> = 3,020	<i>N</i> = 3,010

与先前研究结果的对比

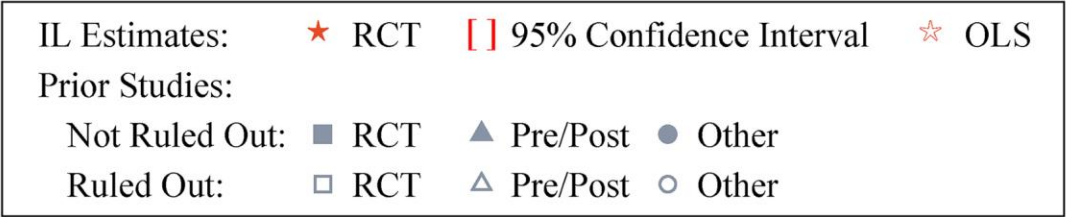


RCT的研究结果与先前78%的研究结果相悖!



(A) Medical Spending Estimates

(B) Absenteeism Estimates



为什么之前的研究结果与RCT的研究结果有显著差异？



选择性偏差 (Selection Bias)

之前的研究忽略了选择性偏差的存在！



发表性偏差 (Publication Bias)

在同类研究中，显著结果的论文比无显著结果的论文更可能被接受和发表。因此，论文作者倾向于只展示显著的研究结果。



研究发现的总结



不具显著性的结果凸显了通过大规模健康干预措施提升个人健康状况所面临的挑战。对这些令人失望的结果的一种解释是：**从WWP中受益最大的人通常拒绝参与其中，即使WWP提供了大量的金钱激励。**更好地了解这类人群的决策机制将有助于改进健康干预措施，从而更好地服务于这类人群。



谢谢大家的聆听!



参考文献: Jones, Molitor & Reif (2019)



讲课人: 何濯羽 (国际教育学院)