Community Monitors vs. Leakage:
Experimental Evidence from Afghanistan

Pacific Conference for Development Economics
Government Performance
San Diego | March 2015

Eli Berman
UCSD, NBER, PDEL, IGCC, ESOC

Michael Callen
Harvard, ESOC

Luke Condra
U. Pittsburgh, ESOC

Mitch Downey
UCSD, ESOC

Mohammad Isaqzadeh
American Univ. of Afghanistan, ESOC

Tarek Ghani
UC Berkeley, ESOC

Abstract

Corruption and leakage are pervasive in many developing countries, necessitating audits and monitoring of project spending. Yet international aid increasing flows to violent, unstable spaces where foreign donors cannot access projects.

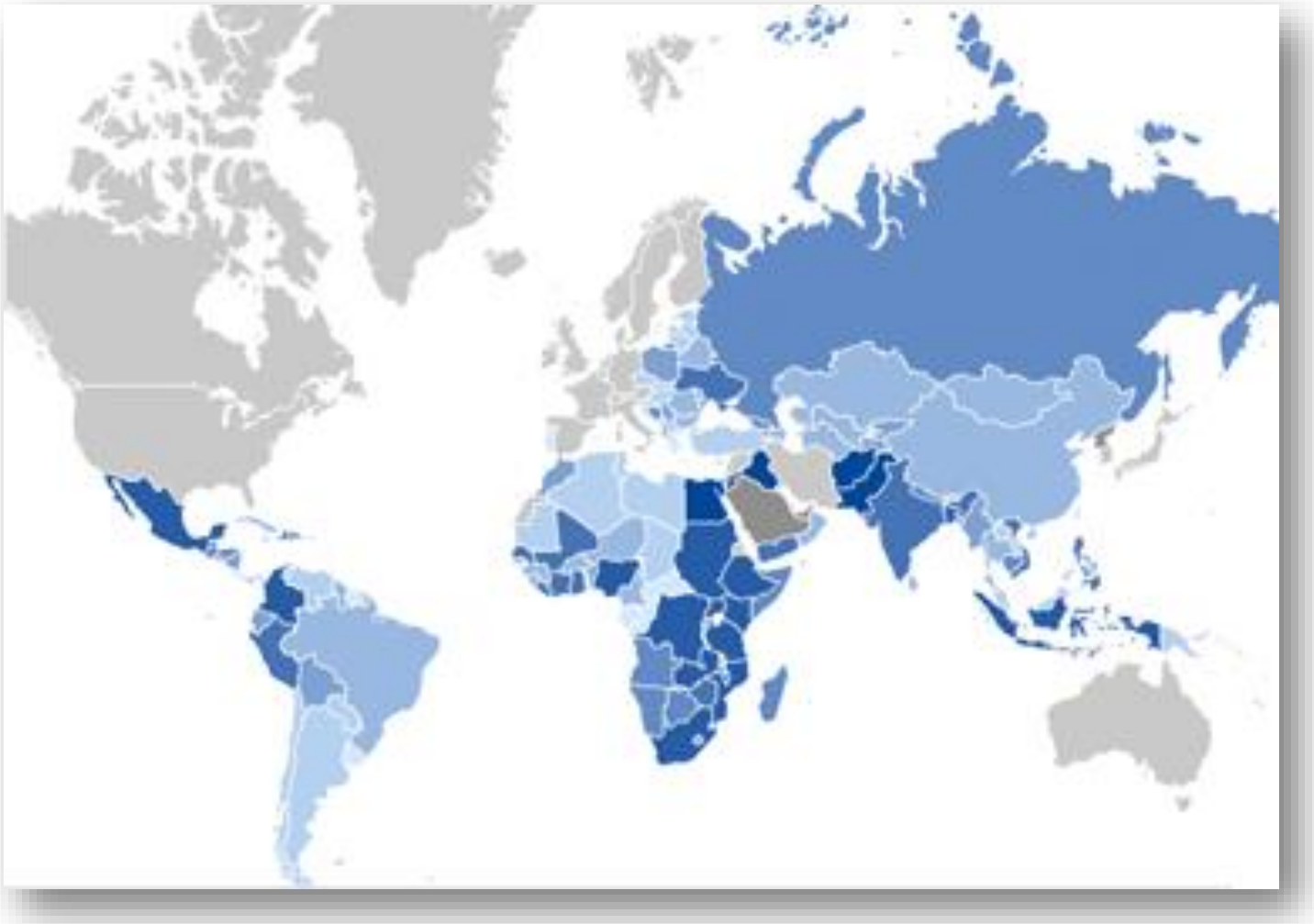
We experimentally evaluated a program that trained Afghans to monitor the quality of rural road construction adjoining their villages, funded by the World Bank.

Road quality was measured independently by separately trained Afghan technical teams.

Local monitors substantially improved road quality: a road on which all villages were treated experienced a statistically significant improvement about 75% (1.4 std. deviations) greater than that of pure control roads.

These results indicate both substantial leakage and a relatively inexpensive remedy which is feasible in poorly governed spaces.

Where does development spending go?



USAID development spending per capita

Spending in locations unsafe for development?

	Average: 2003-2012			
	Annual ODA per capita (2011 $) World Bank + USA + UK	GDP per capita	Corrupt. Rank (of 212)	Political instability rank
Iraq	$149	$2,280	6	2
Afghanistan	90	352	5	3
Haiti	54	525	9	23
Somalia	36	-	1	1
Zambia	34	789	70	130
Rwanda	28	362	123	65
Ghana	27	656	120	103
Colombia	27	4,284	108	9
Mozambique	26	399	80	123
Tanzania	24	458	75	78

Problems with development programs in poorly governed spaces:

A. Not grounded in welfare theorems
- since people and property not safe.

B. May increase violence,
- since it creates lootable rents (Collier).

C. Programs suffer **capture** and **leakage**,
- since oversight by funders is hard.

D. Empirical **research** has same measurement issues as interveners.

Community Monitoring vs. Capture and Leakage?

- Dual goals:
  - Reduce theft and leakage
  - Helps build evidence-base
- Evidence: Olken (2007)
  - Government audits reduced theft by 20%
  - Village monitoring did not work



- Problems with audits
  - Expensive
  - Dangerous
- Problems with village monitoring
  - Local elite capture
  - Can amateurs evaluate roads?

Outline

1. Motivation: Why community monitors?
2. Institution: Integrity Watch Afghanistan
3. Experimental Design
4. Estimating Equations
5. Results – Balance and misbalance
6. Results – ITT and TOT
7. Conclusions:

Can we evaluate in poorly governed spaces?

- Development projects particularly valuable in dangerous places
  - People are overwhelmingly poor
  - Reinforce governance
  - But extorted projects may be wasteful or even counterproductive
  - So evaluation is critical to effective programming

- Evaluation challenges:
  1. Implement monitoring
  2. Successfully randomize monitoring
  3. Implement quality measurement

- Research questions:
  - Can we even evaluate, given measurement challenge?
  - Do community monitors improve road quality?

Implementing Partner: Integrity Watch Afghanistan (IWA)



- Trains community members in basic engineering and integrity standards.
- Teaches people how to monitor local reconstruction projects.



Research design (I)

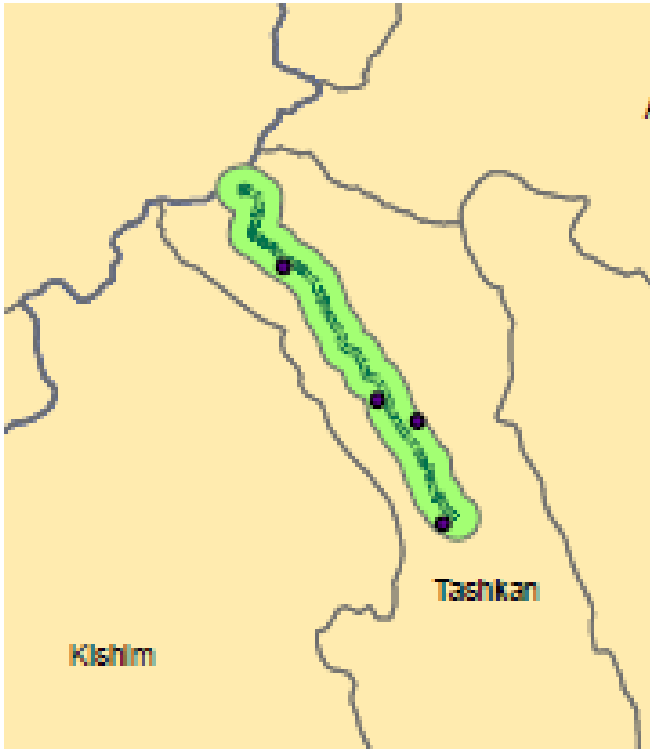
- Two questions:

  A. Do treated villages get better roads?

  B. Does monitoring spill over to neighboring villages ?
  - Positive spillover due to contractor feeling observed ?
  - Negative spillover due to contractor shifting resources from C to T ?

- Two level randomization:
  - Roads stratified into 3 quality terciles
  - Within strata, roads randomized into:
    - Pure control
    - 70% saturation
    - 95% saturation
  - Within roads, villages randomized to treatment and control





Implementation Timeline

	2011	2012	2013	Responsibility
Roads built	------|----------	--------- ? ---------	--------- ? ---------	Ministry, Contractors
Monitoring	----|----------	----------	----------	IWA
Tech team measurement	------|----------	---------|----------	---------|----------	Us
Village surveys	-|----------	----------	---------|----------	Us
Randomization	---|----------	----------	----------	Us

- Two types of data collection
  - Villager surveys: demographics, villager-assessed road quality
  - Tech team measurements: multiple measurements of road quality, incl. camber, gravel, safety panel of points on roads, GPS located

Prelude: ITT estimating equation

- Outcomes in village *v* on road *r* at time *t*:

$$dy_{vrt} = \beta_1 Treat_{vr} + \beta_2 D(S_r = .7) + \beta_3 D(S_r = .95) + \alpha_{q(r)} + \delta_t + \varepsilon_{vrt}$$

- For village v, road r, year t, where $Treat_{vr}$ is village assignment,

  $D(S_r = j)$ is dummy for saturation j

  $\alpha_{q(r)}$ are tercile fixed effects.

Results I: Balance and misbalance

- Forced treatment-control balance on several village-level variables
  - Single ethnicity, female participation, CDC.
- Some roads are outliers in some dimensions
- Big stick balancing effectively forced them into certain assignment groups
- Test: Re-randomize roads 390 times
  - 3 roads were given same assignment in 50% or more of these iterations
  - They were large roads and outliers in some dimension
- We don't think of these as randomly assigned

Misbalance of road assignment: 10000 Replications of RA program

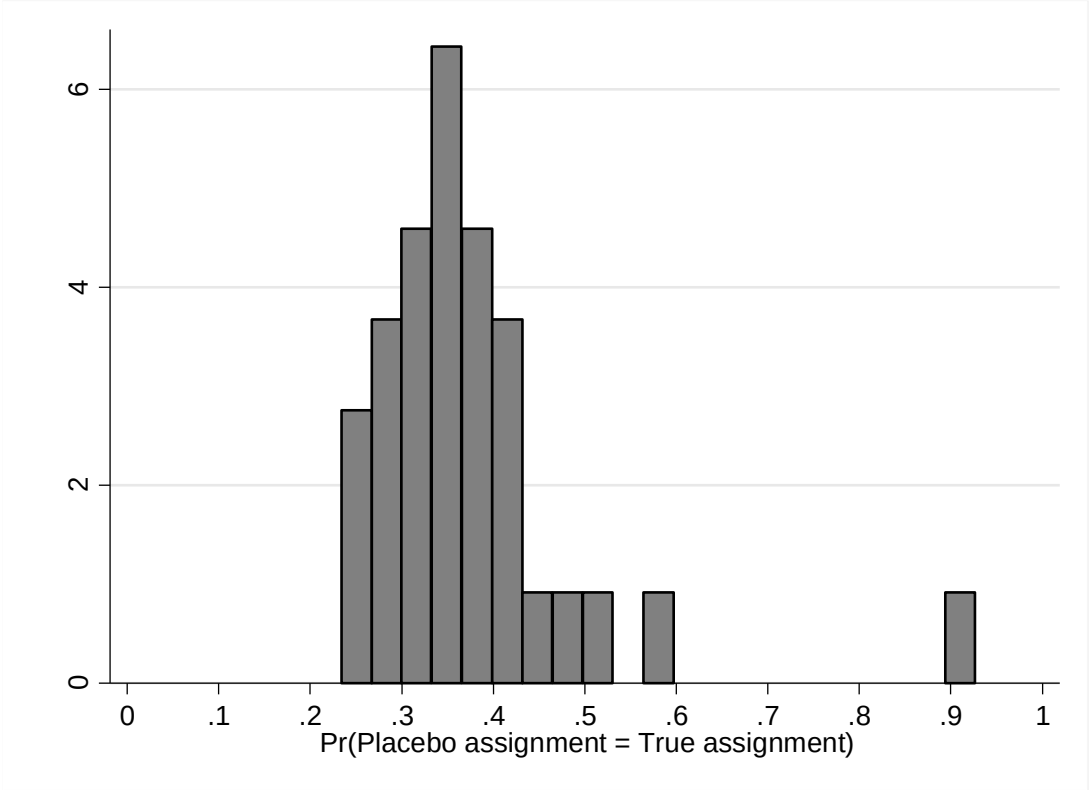
True RA would have allocated roads to “True assignment” 1/3 of times.

Of 10000 replications, only 390 met balance criteria.

Realized distribution forced some roads into particular saturation categories with p>1/3 in order to satisfy balance.

Our solution: Use only a “randomized” sample that has .2 < p <.5 of road assignment to realized category.

“Randomized” sample includes 24 roads.



Baseline balance (road saturation)

	Baseline means			p-value 95% vs. cont.	p-value 70% vs. cont.
	95% sat.	70% sat.	Pure control		
Road safety (3km buffer)	.069	.352	.536	.192	.520
Tech aggregate (3km buffer)	-.990	-.791	-.241	.112	.164
Road safety (2km buffer)	.044	.421	.644	.080*	.224
Road quality (village survey)	.760	.941	1.49	.620	.560
Number of villages on road	6.51	5.05	3.30	.064*	.304
CDC in village	.713	.634	1.00	.332	.024**
N. of villages (3km buffer)	21 (7 roads)	38 (10 roads)	15 (5 roads)		
N. of villages (2km buffer)	18 (5 roads)	35 (10 roads)	12 (4 roads)		
N. of villages (survey data)	35 (8 roads)	45 (11 roads)	15 (5 roads)		

Means in levels and differences

Levels	mean	SD	min	max	Village(N)-years
Road safety (3km)	.997	.707	0	2.30	213
Tech aggregate (3km)	-.154	.750	-1.13	1.68	213
Road safety (2km)	1.03	.724	0	2.33	193
Survey quality	.974	.783	0	3	181
Long difference	**mean**	**SD**	**min**	**max**	**N**
Road safety (3km)	1.254	.577	-.167	2	70
Tech aggregate (3km)	1.239	.614	-.750	2.431	70
Road safety (2km)	1.240	.648	-.5	2	63
Survey quality	.113	1.11	-3.00	2.40	86
First difference (’11 - ’12)	**mean**	**SD**	**min**	**max**	**N**
Road safety (3km)	.883	.422	-.167	2.00	55
Tech aggregate (3km)	.693	.561	-.233	2.81	55
Road safety (2km)	.888	.470	-.500	2	48
First difference (’12 - ’13)	**mean**	**SD**	**min**	**max**	**N**
Road safety (3km)	.332	.456	-.889	1.09	60
Tech aggregate (3km)	.673	.562	-1.15	1.44	60
Road safety (2km)	.297	.495	-1.00	1.71	54

Estimating equation (I) - ITT

- Outcomes in village *v* on road *r* at time *t*:

$$dy_{vrt} = \beta_1 Treat_{vr} + \beta_2 D(S_r = .7) + \beta_3 D(S_r = .95) + \alpha_{q(r)} + \delta_t + \varepsilon_{vrt}$$

- For village v, road r, year t, where $Treat_{vr}$ is village assignment,

  $D(S_r = j)$ is dummy for saturation j

  $\alpha_{q(r)}$ are tercile fixed effects.

ITT effects (Long Differences)

Dependent variable:	Safety (3km)	Aggregate (3km)	Safety (2km)	Survey qual.
Village treatment	0.027	0.052	0.035	0.070
	(0.032)	(0.042)	(0.031)	(0.069)
70% saturation	0.406*	0.331	0.419**	-0.121
	(0.136)	(0.162)	(0.133)	(0.169)
95% saturation	0.619**	0.450**	0.698**	-0.203
	(0.150)	(0.194)	(0.184)	(0.142)
Number of villages	0.016	-0.057	-0.032	0.087
	(0.027)	(0.047)	(0.035)	(0.032)
y_{vrt-1}	-0.909***	-0.992***	-0.813***	-0.600***
	(0.094)	(0.125)	(0.087)	(0.056)
Constant	1.074	0.458	1.099	0.311
	(0.211)	(0.168)	(0.156)	(0.194)
R^2	0.89	0.89	0.89	0.38
Tercile fixed effects	Yes	Yes	Yes	Yes
N. Roads	22	22	19	24
N. Villages	70	70	63	95

Note: Randomization sample, clustered std. errors, stars denote clustered wild bootstrap p values

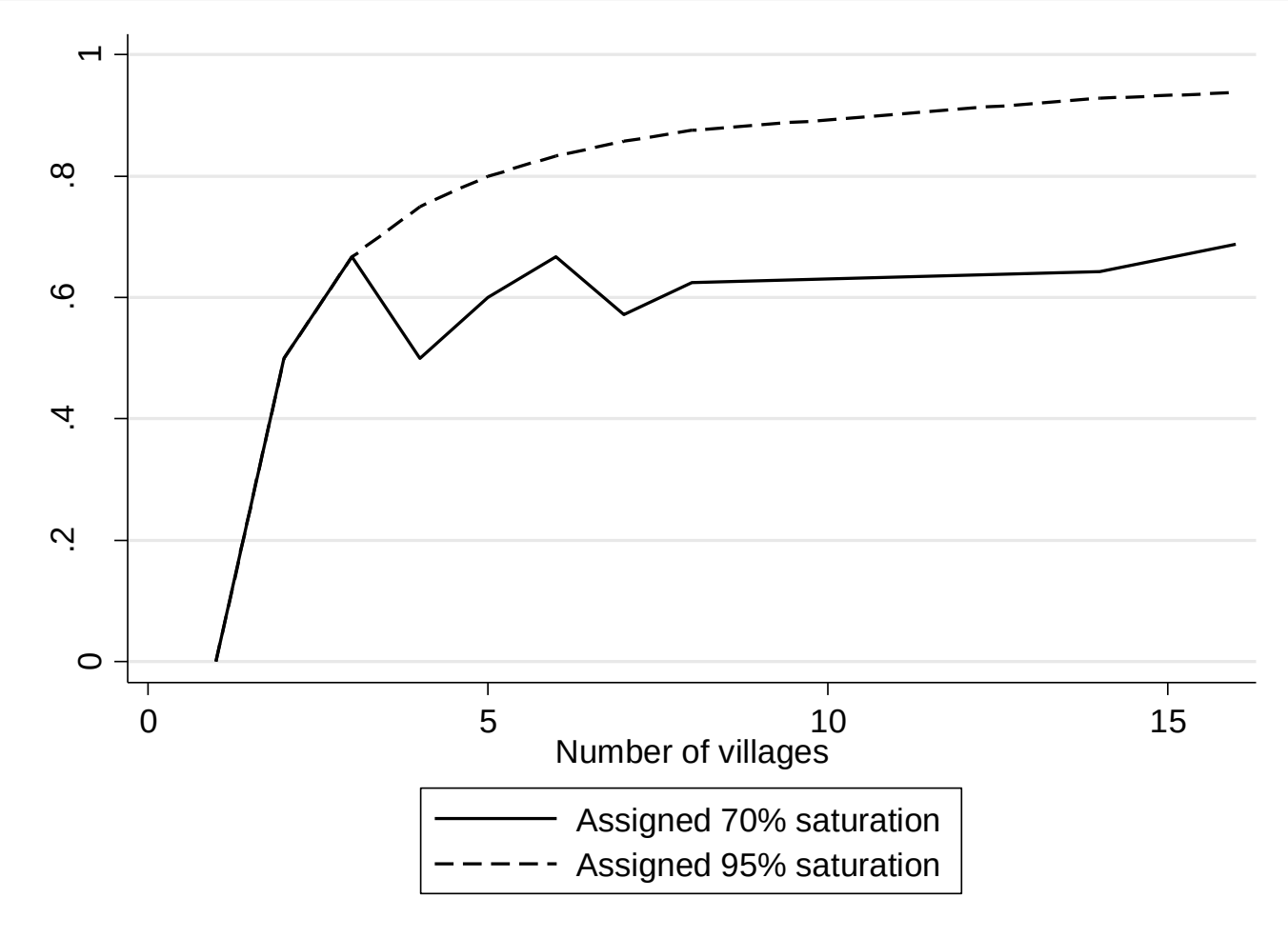
Estimating equation (II) - TOT

- We can enhance precision with a simpler, linear estimating equation
- Outcomes in village *v* on road *r* at time *t*:

$$dy_{vrt} = \beta_1 Treat_{vr} + \beta_2 S_r^E + \alpha_{q(r)} + \delta_t + \varepsilon_{vrt}$$

- For village v, road r, year t, where..
  $Treat_{vr}$ is village assignment,
  S_r^E is the effective saturation rate,
  and $\alpha_{q(r)}$ are tercile fixed effects.
- We have an instrument for S_r^E .

Assigned S instruments for Effective S



Note: In this “first stage” $R^2 = 1$.

TOT effects (long differences)

Dependent variable:	Safety (3km)	Aggregate (3km)	Safety (2km)	Survey qual.
Village treatment	0.017	0.035	0.030	0.033
	(0.035)	(0.038)	(0.030)	(0.067)
Realized saturation	0.838**	0.669**	0.891**	-0.103
	(0.197)	(0.237)	(0.237)	(0.184)
Number of villages	-0.001	-0.077	-0.042	0.082**
	(0.026)	(0.048)	(0.037)	(0.035)
y_{vrt-1}	-0.954***	-1.020***	-0.862***	-0.600***
	(0.091)	(0.102)	(0.091)	(0.060)
Constant	1.135	0.494	1.128	0.274
	(0.169)	(0.149)	(0.165)	(0.137)
R^2	0.89	0.90	0.89	0.37
Tercile fixed effects	Yes	Yes	Yes	Yes
N. roads	22	22	19	24
N. villages	70	70	63	95

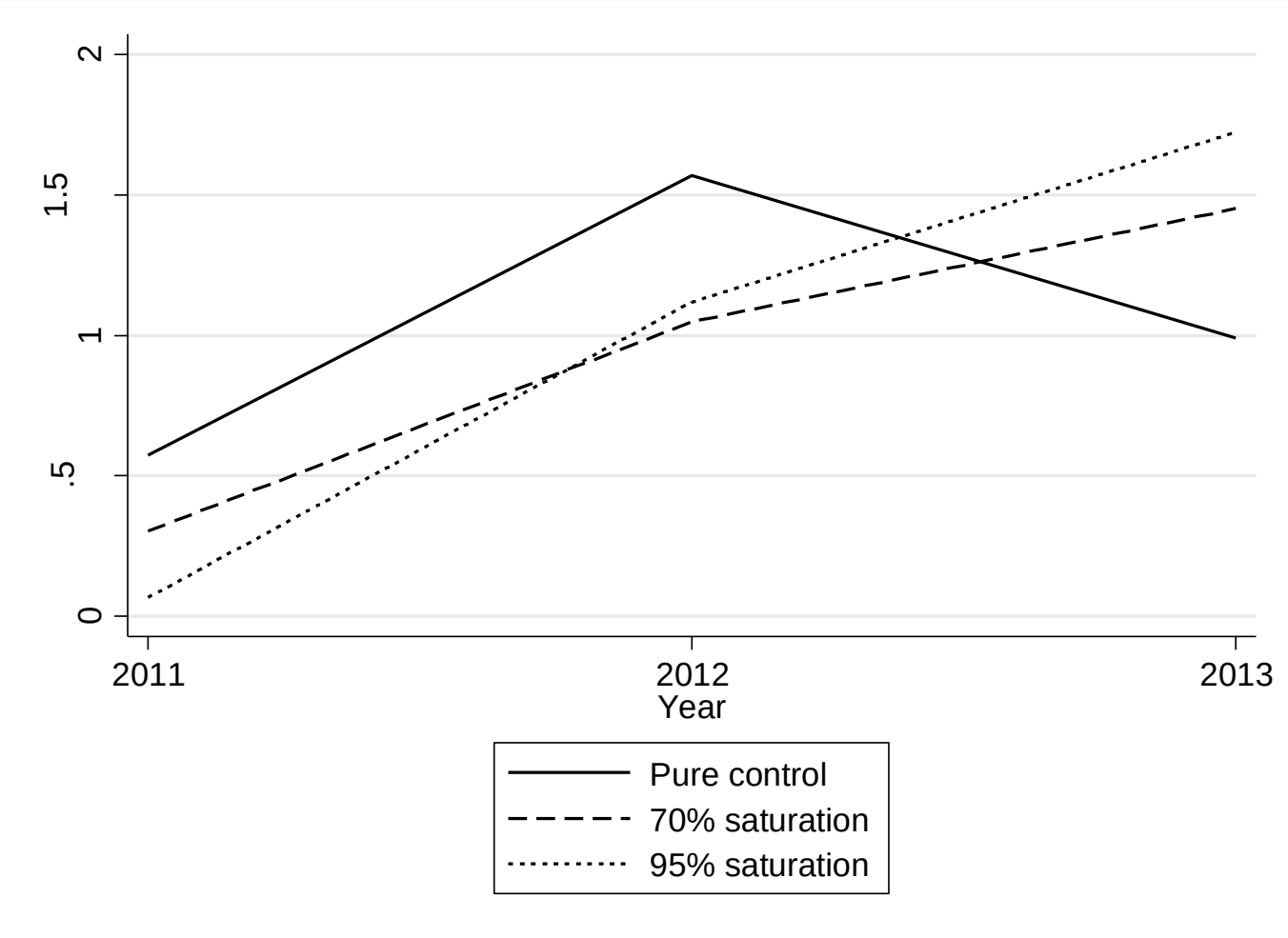
Note: Randomization sample, clustered std. errors, stars denote clustered wild bootstrap p values

TOT effects (separately by year)

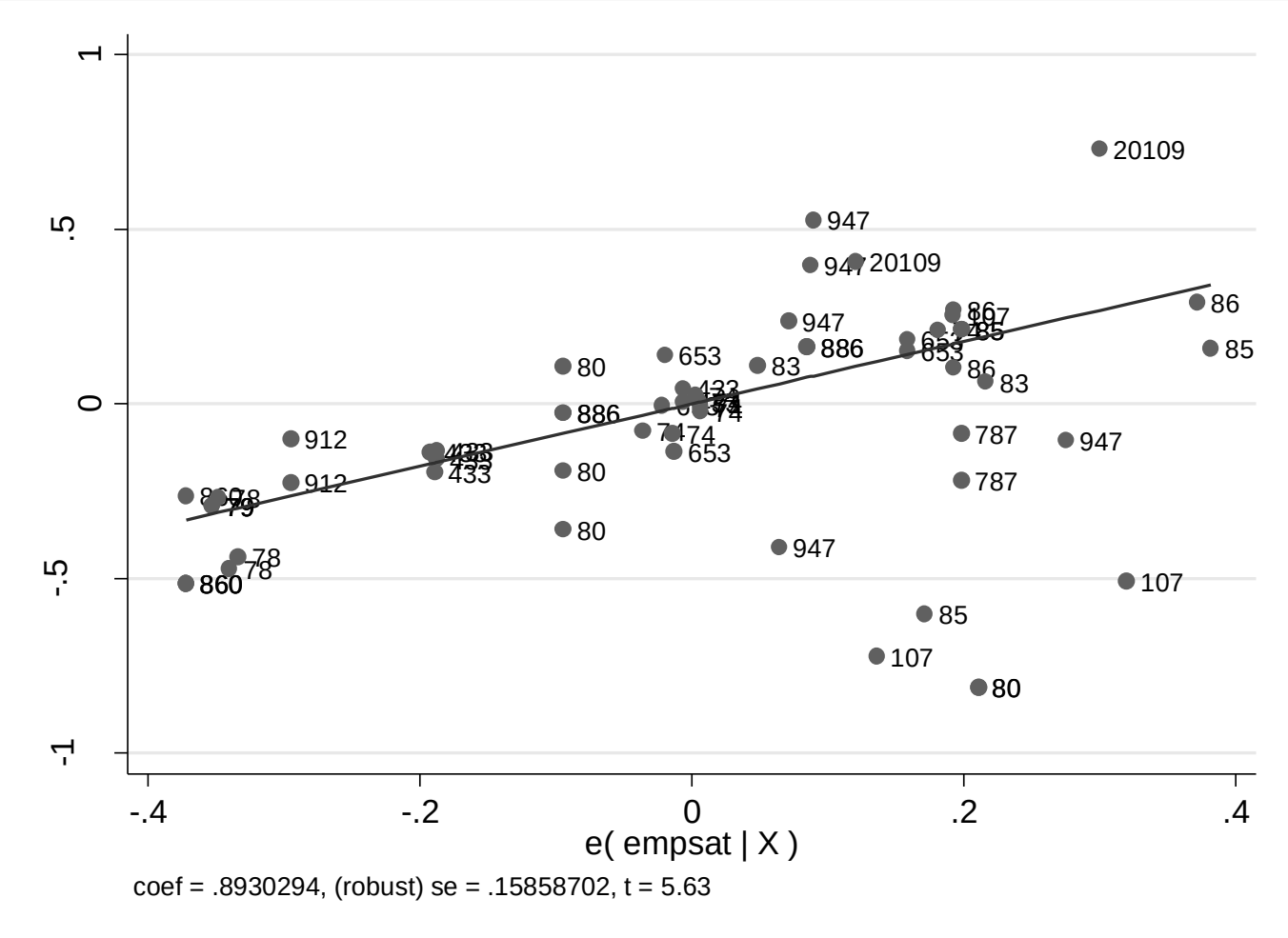
	2012		2013	
Dependent variable:	Safety (3km)	Safety (2km)	Safety (3km)	Safety (2km)
Village treatment	0.068	0.066	0.029	0.054
	(0.084)	(0.114)	(0.054)	(0.062)
Realized saturation	-0.010	0.304	0.893***	0.671***
	(0.381)	(0.586)	(0.159)	(0.135)
Number of villages	0.012	0.029	-0.013	-0.029
	(0.054)	(0.051)	(0.017)	(0.026)
y_{vrt-1}	-0.632*	-0.962*	-0.710***	-0.719***
	(0.181)	(0.156)	(0.083)	(0.066)
Constant	1.553	1.536	0.529	0.724
	(0.337)	(0.304)	(0.186)	(0.184)
R^2	0.446	0.534	0.845	0.817
N	55	48	60	54
Tercile fixed effects	Yes	Yes	Yes	Yes
N. roads	16	14	16	15
N. villages	55	48	60	54

Note: Randomization sample, clustered std. errors, stars denote clustered wild bootstrap p values

Road quality, by road saturation status



Saturation and Δ quality, 2012-13 (column 3)



TOT effects by year – Full sample

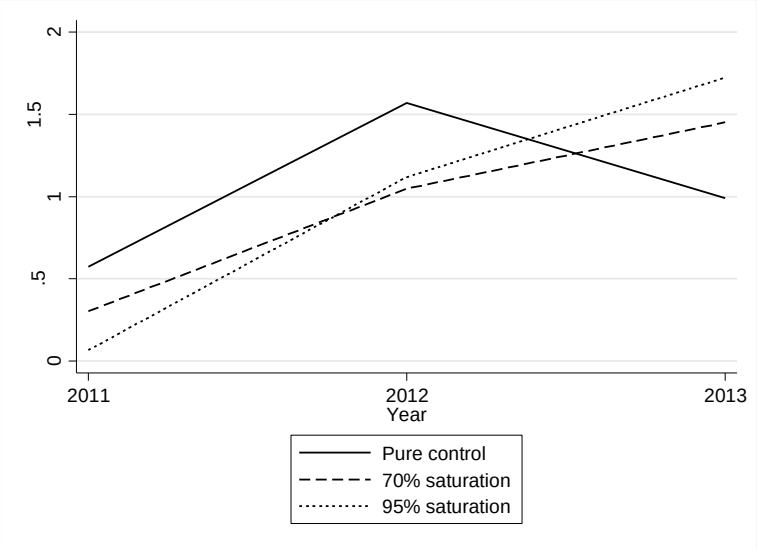
	2012		2013	
Dependent variable:	Safety (3km)	Safety (2km)	Safety (3km)	Safety (2km)
Village treatment	0.067	0.058	-0.016	-0.011
	(0.060)	(0.082)	(0.053)	(0.062)
Realized saturation	-0.233	-0.195	1.035***	0.943***
	(0.275)	(0.346)	(0.209)	(0.213)
Number of villages	0.052	0.074*	-0.077***	-0.078***
	(0.015)	(0.018)	(0.008)	(0.010)
y_{vrt-1}	-0.682**	-0.914**	-0.770***	-0.852***
	(0.148)	(0.122)	(0.101)	(0.078)
Constant	1.572	1.657	0.994	1.207
	(0.198)	(0.171)	(0.224)	(0.180)
R^2	0.445	0.501	0.85	0.832
N	82	74	99	89
Tercile fixed effects	Yes	Yes	Yes	Yes
N. roads	25	22	26	24
N. villages	82	74	99	89

Note: Randomization sample, clustered std. errors, stars denote clustered wild bootstrap p values

Conclusion: Community monitors vs. leakage

- **Can we measure and experimentally evaluate roads in rural Afghanistan?**
  YES

- Do community monitors improve road quality?
  YES
  - With persistence
  - With positive spillovers on roads
  - Inexpensive and safe
  - Despite weak formal incentives for contractors to heed monitors



- Broader implication
  - Community monitoring can enable effective development spending in poorly governed and dangerous spaces.

Bonus slides
including full sample results and other stuff for longer seminar

Baseline balance (village assignment)

	Baseline means		Unconditional difference	Conditional difference	Bootstrapped p-value
	Treatment	Control			
Road safety (3km buffer)	.198	.470	-.273	-.208	.036**
Tech aggregate (3km buffer)	-.901	-.533	-.367	-.175	.024**
Road safety (2km buffer)	.243	.512	-.269	.215	.052*
Road quality (village survey)	.692	1.01	-.323	.082	.284
Number of villages on road	7.39	4.36	3.04	1.76	0***
CDC in village	.775	.845	-.070	-.128	.116
N. of villages (3km buffer)	55	48			
N. of villages (2km buffer)	51	43			
N. of villages (survey data)	83	60			

Baseline balance (road saturation)

	Baseline means			p-value 95% vs. cont.	p-value 70% vs. cont.
	95% sat.	70% sat.	Pure control		
Road safety (3km buffer)	.221	.319	.670	.288	.396
Tech aggregate (3km buffer)	-.836	-.823	-.069	.124	.060*
Road safety (2km buffer)	.242	.383	.742	.124	.032**
Road quality (village survey)	.580	.851	1.42	.116	.076*
Number of villages on road	7.65	5.33	2.69	.184	.680
CDC in village	.754	.793	1	.028	.152
N. of villages (3km buffer)	38 (13 roads)	46 (11 roads)	19 (9 roads)		
N. of villages (2km buffer)	35 (11 roads)	43 (11 roads)	16 (8 roads)		
N. of villages (survey data)	55 (15 roads)	69 (13 roads)	19 (9 roads)		

ITT effects (pooled across years)

Dependent variable:	Safety (3km)	Aggregate (3km)	Safety (2km)	Survey qual.
Village treatment	0.089*	0.038	0.087	0.057
	(0.047)	(0.056)	(0.058)	(0.061)
70% saturation	-0.167	-0.339	-0.210	-0.072
	(0.196)	(0.256)	(0.163)	(0.156)
95% saturation	0.192	0.044	0.218	-0.147
	(0.196)	(0.225)	(0.193)	(0.122)
Year 2013	-0.047	0.163	-0.004	
	(0.209)	(0.194)	(0.213)	
Number of villages	-0.030	0.008	-0.029	0.024
	(0.015)	(0.033)	(0.018)	(0.008)
y_{vrt-1}	-0.809**	-0.733***	-0.878***	-0.567***
	(0.154)	(0.103)	(0.166)	(0.055)
Constant	1.503	0.569	1.619	0.500
	(0.264)	(0.317)	(0.238)	(0.167)
R^2	0.753	0.533	0.758	0.309
N	181	181	163	265
Tercile fixed effects	Yes	Yes	Yes	Yes
Years	2012, 2013	2012, 2013	2012, 2013	2012, 2013
N. roads	26	26	24	37
N. villages	100	100	92	143

ITT effects (separately by year)

Dependent variable:	Safety (3km)	Safety (2km)	Safety (3km)	Safety (2km)
Village treatment	0.097	0.090	0.081	0.092
	(0.052)	(0.065)	(0.061)	(0.078)
70% saturation	-0.565**	-0.547*	0.289	0.062
	(0.183)	(0.196)	(0.216)	(0.173)
95% saturation	-0.306	-0.254	0.573*	0.438*
	(0.167)	(0.219)	(0.192)	(0.199)
Number of villages	0.039*	0.063**	-0.053**	-0.053*
	(0.010)	(0.014)	(0.008)	(0.010)
y_{vrt-1}	-0.775***	-0.987***	-0.843***	-0.986***
	(0.123)	(0.120)	(0.137)	(0.117)
Constant	2.024	2.074	1.172	1.669
	(0.238)	(0.208)	(0.451)	(0.347)
R^2	0.566	0.618	0.822	0.81
N	82	74	99	89
Tercile fixed effects	Yes	Yes	Yes	Yes
Years	2012	2012	2013	2013
N. roads	25	22	26	24
N. villages	82	74	99	89

TOT effects (pooled across years)

Dependent variable:	Safety (3km)	Aggregate (3km)	Safety (2km)	Survey qual.
Village treatment	0.018	-0.012	0.019	0.036
	(0.041)	(0.051)	(0.046)	(0.055)
Realized saturation	0.544*	0.294	0.561*	-0.059
	(0.224)	(0.243)	(0.258)	(0.172)
Year 2013	-0.136	0.138	-0.131	
	(0.168)	(0.188)	(0.174)	
Number of villages	-0.043	0.000	-0.044	0.026
	(0.022)	(0.044)	(0.024)	(0.010)
y_{vrt-1}	-0.697***	-0.664***	-0.725***	-0.567***
	(0.124)	(0.092)	(0.126)	(0.057)
Constant	1.100	0.197	1.140	0.455
	(0.186)	(0.319)	(0.183)	(0.121)
R^2	0.73	0.49	0.727	0.306
N	181	181	163	265
Tercile fixed effects	Yes	Yes	Yes	Yes
Years	2012, 2013	2012, 2013	2012, 2013	2012, 2013
N. roads	26	26	24	37
N. villages	100	100	92	143