

Cycling to School: Increasing Secondary School Enrollment for Girls in India

By Karthik Muralidharan and Nishith Prakash

Team: Hitomi, Reina, Mao, Asahi, Xihang

Thesis Summary

2017年に「American Economic Journal: Applied Economics」で発表された論文である。

インドのビハール州で実施された自転車プログラムが女子の中等教育への進学率に与えた影響を調査している。

※この調査の中で出てくる「サイクル・プログラム」は、女子生徒が中学校9年生)に進学する際に自転車を提供し、学校へのアクセスを改善することを目的としている。

[主なRQ]インドのビハール州で実施されたサイクルプログラムは、女子の中等学校への進学率や学業成果にどのような影響を与えたのか。

著者らはトリプル・ディファレンス・アプローチを用いて、女子の進学率、卒業試験の受験率および合格率に対するプログラムの影響を分析した。

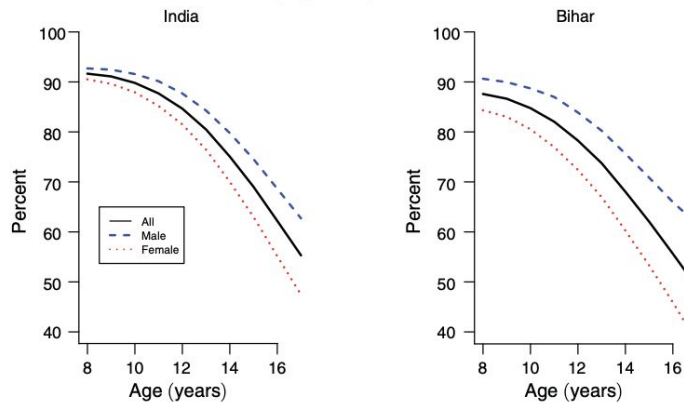
Figure 1: Enrollment in School by Age and Gender: 年齢別および性別による学校入学率を示す。

Figure 2: Non-parametric Double and Triple Difference Estimates: 学校までの距離に応じたプログラムの効果を非パラメトリックに示す。

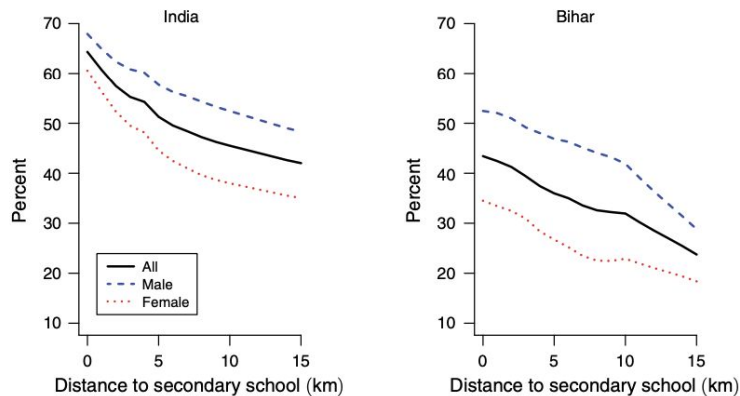
Table 2: Triple Difference (DDD) Estimate of the Impact: サイクルプログラムが女子生徒の中等学校進学率に与えた影響をトリプル・ディファレンス法を用いて推定した結果を示す。

Figure 1

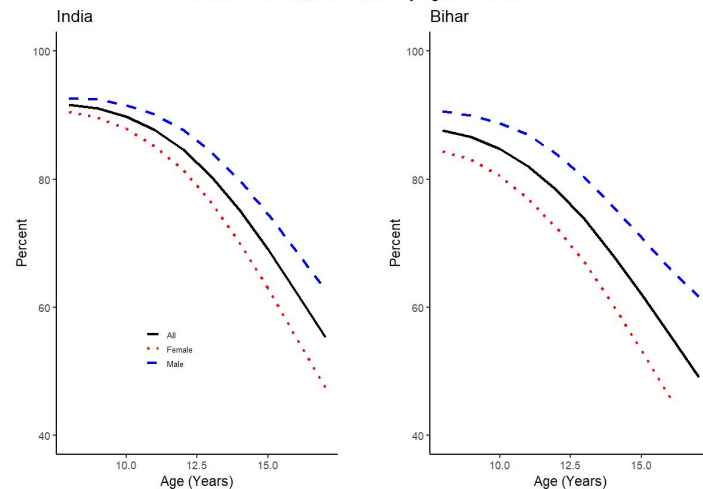
Panel A: Enrollment in school by age and gender



Panel B: 16- and 17-year-olds enrolled in or completed grade 9 by distance and gender



Panel A: Enrollment in School by Age and Gender



Panel B: 16 and 17 Year Olds Enrolled in OR Completed Grade 9 by Distance and Gender

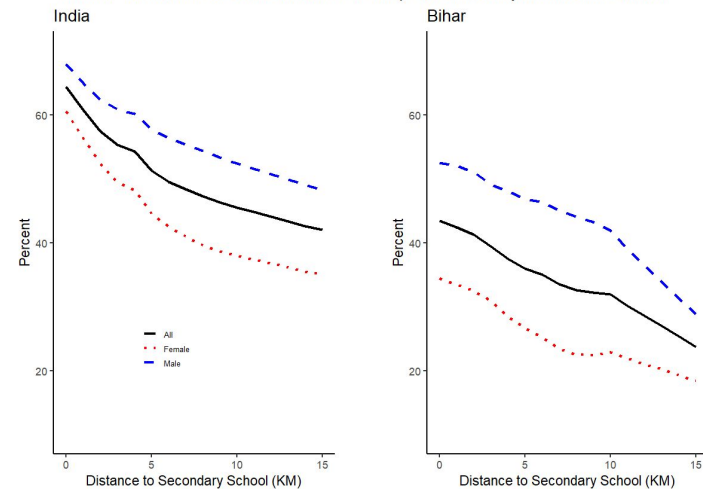
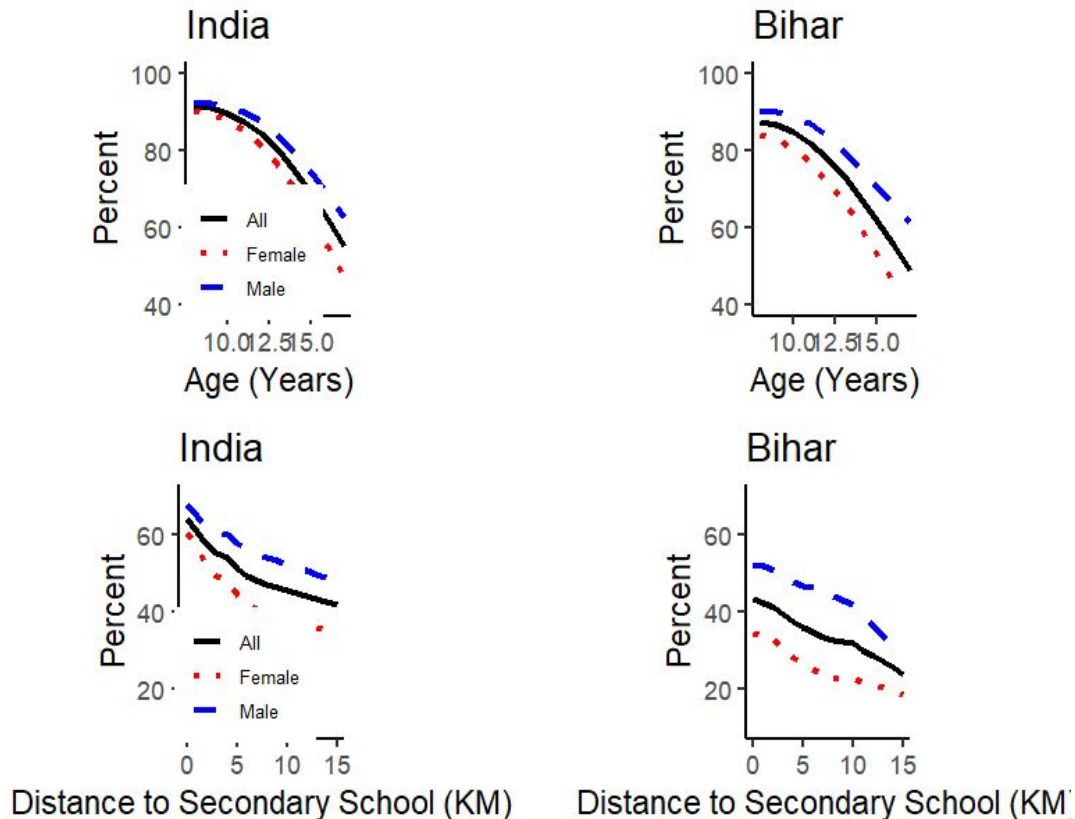


Figure 1

失敗例

- ・凡例の位置やサイズが悪い
- ・図そのもののサイズが小さく、破線が4個程度しかない
- ・Panel名が入っていない

Figure 1



Source: Author's calculations using the 2008 District Level Health Survey (DLHS).

```
# Arrange the plots in a 2x2 grid
p5 <- grid.arrange(
  p1, p2,
  ncol = 2,
  top = "Panel A: Enrollment in School by Age and Gender"
)
p6<- grid.arrange(
  p3, p4,
  ncol = 2,
  top = "Panel B: 16 and 17 Year Olds Enrolled in OR Completed Grade 9 by Distance and Gender"
)

p7<- grid.arrange(
  p5, p6,
  ncol = 1
)
```

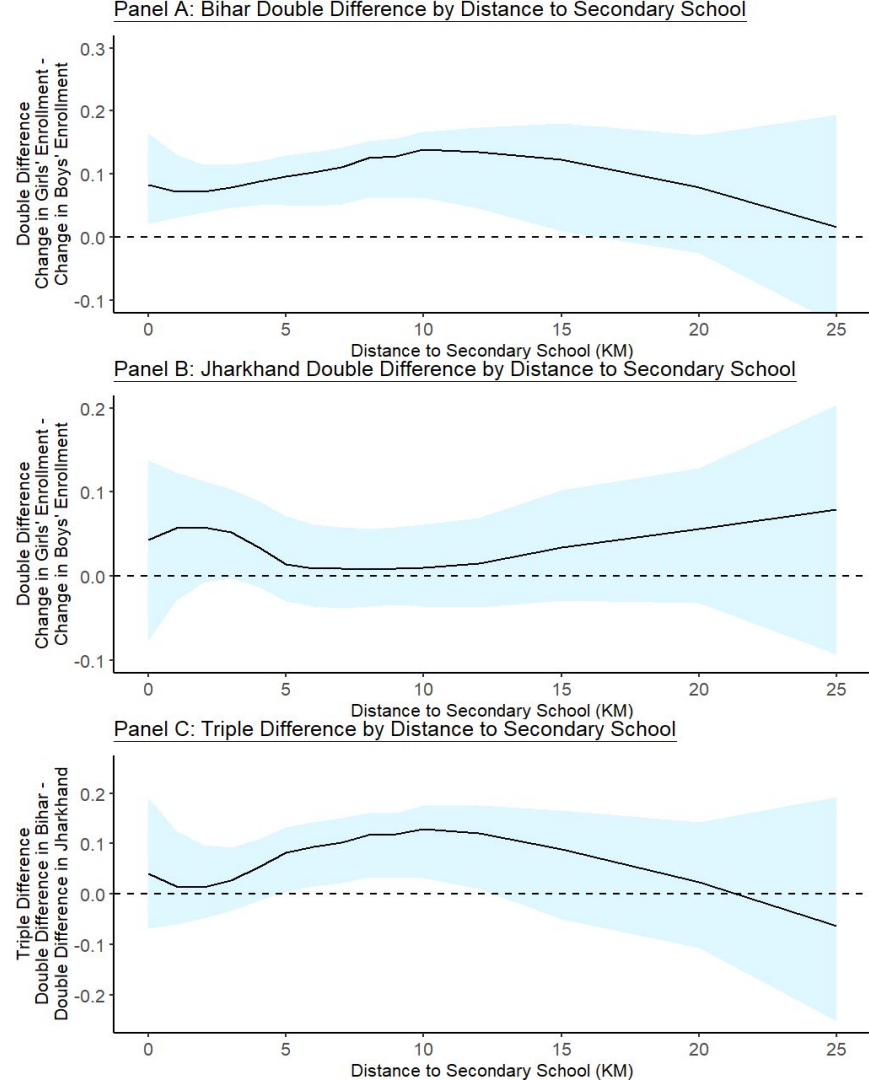
工夫した点

- ・プロットに使用する変数を事前に準備した
- ・論文と同じ図の構成を再現するために、`gridExtra::grid.arrange()`を活用した

反省点

- ・関数を用いて図のプロットを行うコードを短縮できた
- ・凡例を綺麗に表示することに時間がかかった

Figure 2



工夫した点

- ・baseR を tidyR に変換した
- ・論文の図に近づけるために theme_classic などを活用した
- ・データを確認しながら作業を進めることで、コードで何をしているかの対応関係が明確になった

反省点

- ・使用したデータは同じはずだが、作成した図と論文の図でズレが生じている
- ・同じような図の作成をするときに、関数を使ってコードを簡潔にしたかった

DDD推定

目的変数: Enrolled in or completed grade 9

子供の性別、処置群 (age 14 -15)、ビハール州の交差項

β_1 の推定に興味がある

$$(1) \ y_{ihv} = \beta_0 + \beta_1 \cdot F_{ihv} \cdot T_{ihv} \cdot BH_{ihv} + \beta_2 \cdot F_{ihv} \cdot BH_{ihv} + \beta_3 \cdot T_{ihv} \\ \times BH_{ihv} + \beta_4 \cdot F_{ihv} \cdot T_{ihv} + \beta_5 \cdot F_{ihv} + \beta_6 \cdot T_{ihv} + \beta_7 \cdot BH_{ihv} + \varepsilon_{ihv},$$

Table 2

DDDの推定結果

属性(カーストや宗教)、
世帯(資産、家長の教
育)、村をコントロールし
た上でも、自転車プログラ
ムは女性生徒の進学に
寄与していることがわ
かった

	(1)	(2)	(3)	(4)
Treat*female*Bihar	0.103 (0.030)	0.091 (0.030)	0.052 (0.025)	0.052 (0.025)
Treat*female	0.020 (0.026)	0.024 (0.026)	0.038 (0.022)	0.039 (0.022)
Treat*Bihar	-0.044 (0.018)	-0.042 (0.018)	-0.029 (0.016)	-0.028 (0.016)
Female*Bihar	-0.094 (0.023)	-0.091 (0.023)	-0.067 (0.020)	-0.066 (0.020)
Treat	-0.148 (0.014)	-0.143 (0.014)	-0.138 (0.013)	-0.138 (0.013)
Female	-0.092 (0.020)	-0.088 (0.020)	-0.100 (0.017)	-0.101 (0.017)
Bihar	0.011 (0.016)	-0.044 (0.017)	-0.032 (0.015)	-0.044 (0.015)
Constant	0.464 (0.013)	0.771 (0.024)	0.593 (0.027)	0.562 (0.040)
N	30295	30295	30147	30112
R2	0.035	0.088	0.207	0.208
Demographic controls	NO	YES	YES	YES
HH socioeconomic controls	NO	NO	YES	YES
Village level controls	NO	NO	NO	YES

工夫した点

- ・行を一部表示し、順番を変えたい、スターを消したいなど、可能な作業の種類を知り、言語化することで、適切に調べたり質問したりすることができた
- ・コントロール変数を含めたかどうかの表示欄(YES、NO)の表はベクトルをつなげて作成し、推定値の表とマージさせた
- ・データフレームを画像やエクセルとして出力するのにgt パッケージを用いた
- ・時間をかけてじっくり考えることができた

反省点

- ・値を中央揃えにできなかった
- ・上下の細かい注記を入れられなかった
- ・時間の見積を間違えてしまった
- ・bind_rows問題

ファイル管理について

工夫した点

- ・ローファイル、コード、アウトプットのフォルダをそれぞれ分けて保存した
- ・別な方法を試したいときに、一旦コメントアウトしておくなど、復元しやすい形にできていた

反省点

- ・データファイルおよびコードファイルの関連や順番について、Readmeを読めばわかるのだが、オーガナイザーに任せてしまった部分があった
- ・オリジナルを残さずに編集してしまい、改めてファイルをダウンロードしなおすことになった

学んだこと・感想

- ・アンサンブルプログラミングを通して、知識の共有ができた
- ・一緒にstruggleしたことで知識・スキルが定着した気がする
- ・質問することが大事
- ・失敗から学ぶことが多い