2A 2020 Schield ECOTS Controlling for Context by Standardizing 1

Controlling for Context By Standardizing

Milo Schield
Editor of www.StatLit.org
Consultant Univ. New Mexico
Fellow, American Statistical Association
President, National Numeracy Network (NNN)
US Rep: International Statistical Literacy Project

ECOTS May 2020

www.StatLit.org/pdf/2020-Schield-eCOTS.pdf
www.StatLit.org/pdf/2020-Schield-eCOTS-Slides.pdf
www.StatLit.org/v/2020-Schield-eCOTS-Slides.mp4

2A 2020 Schield ECOTS Controlling for Context by Standardizing 2

Today's students want to engage in social issues

Most social issues involve social statistics: typically averages, counts and rates.

Most social statistics are crude statistics: they don't take anything else into account.

To really understand social statistics, students need to "see" how to *take something into account.*

Students get engaged in learning that social statistics may have a *story behind the story.*

2A 2020 Schield ECOTS Controlling for Context by Standardizing 3

Most Social Statistics are Observational Statistics

Why is the Covid-19 infection rate much higher in Italy (1,333/M) than in the US (279/M)? [3/25]

- Older people are a bigger share of the population in Italy (23%) than in the US (17%).
- Population density is higher in Italy (533 per sq. mile) than in the US (94 per sq. mile).

To compare Italy's infection rate with US's, such confounders may need to be controlled for.

2A 2020 Schield ECOTS Controlling for Context by Standardizing 4

"Taking into Account": "Controlling for": Mental

Computer methods of controlling for confounders are powerful, but they may obscure the process.

Manual methods are easy to do (weighted average) and can "show" students the key ideas (graphical).

CONTROLLING FOR CONFOUNDERS Take into account (mental)			
Can do by hand			*Calculator/Computer*
1	Select/Stratify	4	Linear Regression
2	Form Ratios	5	Logistic Regression
3	Standardize	6	Multivariate Regress

2A 2020 Schield ECOTS Controlling for Context by Standardizing 5

Standards for Standardizing: Std. Group & Combined Group

Standardizing (adjusting) requires a standard for matching the mixtures (weights) of the two groups.

Standard-group matching means selecting one group as the standard and adjusting the other group mixture to match that standard. (C.f., demography)

Combined-group matching adjusts both group mixtures to their combined values. (C.f., regression)

Calculations can be done algebraically or graphically. Two standards and two calculations = 4 combinations.

2A 2020 Schield ECOTS Controlling for Context by Standardizing 6

Standard Group: Algebra #1A Adjust Rural Mix to Match City

Patients' Death Rate (Mix: Percentage in this condition)			
Hospital	Good Cond.	Poor Cond.	All
City	1% (10%)	6% (90%)	5.5%
Rural	3% (70%)	7% (30%)	4.2%
All: City	= 0.1*1% + 0.9*6%		1.3 points
All: Rural	= 0.7*3% + 0.3*7%		City higher

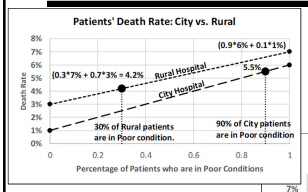
Crude compare Mixed-fruit

Match Rural to City. After *controlling for patient condition*, the death rate is higher at Rural than at City.

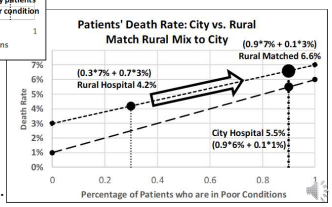
Match Rural Mix to City; Apply City Mix to Rural Patients' Death Rate (Mix: Percentage in this condition)			
Hospital	Good Cond.	Poor Cond.	All
City	1% (10%)	6% (90%)	5.5%
Rural	3% (10%)	7% (90%)	**6.6%**
All: City	= 0.9*6% + (1-0.9)*1%		**-1.1 pts**
All: Rural	**= 0.9*7% + (1-0.9)*3%**		City lower

2A 2020 Schield ECOTS Controlling for Context by Standardizing 7

Standard Group: Graph #1G: Adjust Rural Mix to Match City



Wainer (2002), Schield (2006) Patient death rate lower at Rural hospital than at City. [Mixed-fruit comparison]



Match Rural mixture to City. After *controlling for* patient condition, death rate higher at Rural than at City. [Reversal] Apples and apples comparison.

2A 2020 Schield ECOTS Controlling for Context by Standardizing 8

Combined Mix: Algebra #2A: Adjust All Mixes to Combined

Patients' Death Rate (Mix: Percentage in this condition)			
Hospital	Good Cond.	Poor Cond.	All
City	1% (10%)	6% (90%)	5.5%
Rural	3% (70%)	7% (30%)	4.2%
All: City	= 0.1*1% + 0.9*6%		1.3 points
All: Rural	= 0.7*3% + 0.3*7%		City higher

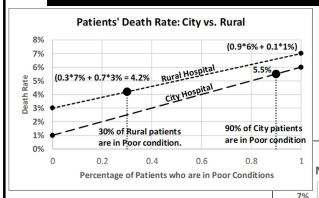
Crude compare Mixed-fruit

Match to combined: 70% After controlling for patient condition, death rate is higher at Rural than at City. [Reversal] Apples & apples compare

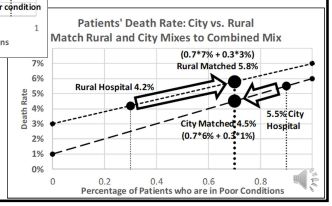
Match City & Rural Mixes to Combined Mix: 70%			
Patients' Death Rate (Mix: Percentage in this condition)			
Hospital	Good Cond.	Poor Cond.	All
City	1% **(30%)**	6% **(70%)**	**4.5%**
Rural	3% **(30%)**	7% **(70%)**	**5.8%**
All: City	**= 0.3*1% + 0.7*6%**		**-1.3 pts**
All: Rural	**= 0.3*3% + 0.7*7%**		City lower

2A 2020 Schield ECOTS Controlling for Context by Standardizing 9

Combined Mix: Graph #2G: Adjust All Mixes to Combined



Patient death rate higher at City hospital than at Rural. [Mixed-fruit comparison]



Standardize on combined mix. After controlling for patient condition, death rate is higher at Rural than at City. [Reversal] Apples and apples comparison.

2A 2020 Schield ECOTS Controlling for Context by Standardizing 10

Compare the Calculations: Algebraic vs. Graphical

PLUS: *Algebraic techniques* seem to be
- simpler for teachers to teach (graphs take time),
- simpler for students to do (graphs are tricky), and
- more applicable (applies to more than two groups)

as compared with the Wainer (2002) graph approach.

MINUS: *Algebraic techniques*
- are calculation based (not visual. Students can't see)
- are very sensitive to the wording (match, apply) and
- give different results depending on the standard.

2A 2020 Schield ECOTS Controlling for Context by Standardizing 11

Both let Students work Problems. Might Teaching Both Be Best?

I've taught combined-group graphs (#2G) for almost 10 years. This past year I taught standard-group algebra (#1A).

Suppose you start with *standard-group algebra* (#1A): it is simpler to teach and simpler to do. Then have students show their results using Wainer's graph (#1B).

Depending on time, introduce *combined-group algebra* (#2A). Have students show their results in a graph (#2B).

Doing the visual graphical technique should help students **see and understand** what the algebraic technique is doing.

2A 2020 Schield ECOTS Controlling for Context by Standardizing 12

Bibliography: The Graphical Technique

Wainer, H. (2002). "The BK-Plot: Making Simpson's Paradox Clear to the Masses." *CHANCE*, 15(3):60-62. www.statlit.org/Wainer.htm www.statlit.org/CP/Cornfield/2002-Wainer-Visual-Revelation-BK-plot.pdf

Schield, M. (2006). Understanding Confounding from Lurking Variables using Graphs. *STATS Magazine* ASA. Fall 2006. pp. 14-18. See www.StatLit.org/pdf/2006SchieldSTATS.pdf

Thanks to Marc Isaacson for the title of – and comments on – these slides.

Controlling for Context By Standardizing

Milo Schield

Editor of www.StatLit.org

Consultant Univ. New Mexico

Fellow, American Statistical Association

President, National Numeracy Network (NNN)

US Rep: International Statistical Literacy Project

ECOTS May 2020

www.StatLit.org/pdf/2020-Schield-eCOTS.pdf

www.StatLit.org/pdf/2020-Schield-eCOTS-Slides.pdf

www.StatLit.org/v/2020-Schield-eCOTS-Slides.mp4

Today's students want to engage in social issues

Most social issues involve social statistics: typically averages, counts and rates.

Most social statistics are crude statistics: they don't take anything else into account.

To really understand social statistics, students need to "see" how to *take something into account*.

Students get engaged in learning that social statistics may have a *story behind the story*.

Most Social Statistics are Observational Statistics

Why is the Covid-19 infection rate much higher in Italy (1,333/M) than in the US (279/M)? [3/25]

- Older people are a bigger share of the population in Italy (23%) than in the US (17%).
- Population density is higher in Italy (533 per sq. mile) than in the US (94 per sq. mile).

To compare Italy's infection rate with US's, such confounders may need to be controlled for.

"Taking into Account": "Controlling for": Mental

Computer methods of controlling for confounders are powerful, but they may obscure the process.

Manual methods are easy to do (weighted average) and can "show" students the key ideas (graphical).

CONTROLLING FOR CONFOUNDERS			
Take into account (mental)			
	Can do by hand		*Calculator/Computer*
1	Select/Stratify	4	Linear Regression
2	Form Ratios	5	Logistic Regression
3	Standardize	6	Multivariate Regress

Standards for Standardizing: Std. Group & Combined Group

Standardizing (adjusting) requires a standard for matching the mixtures (weights) of the two groups.

Standard-group matching means selecting one group as the standard and adjusting the other group mixture to match that standard. (C.f., demography)

Combined-group matching adjusts both group mixtures to their combined values. (C.f., regression)

Calculations can be done algebraically or graphically. Two standards and two calculations = 4 combinations

Standard Group: Algebra #1A Adjust Rural Mix to Match City

Patients' Death Rate (Mix: Percentage in this condition)			
Hospital	Good Cond.	Poor Cond.	All
City	1% (10%)	6% (90%)	5.5%
Rural	3% (70%)	7% (30%)	4.2%
All: City	= 0.1*1% + 0.9*6%		1.3 points
All: Rural	= 0.7*3% + 0.3*7%		City higher

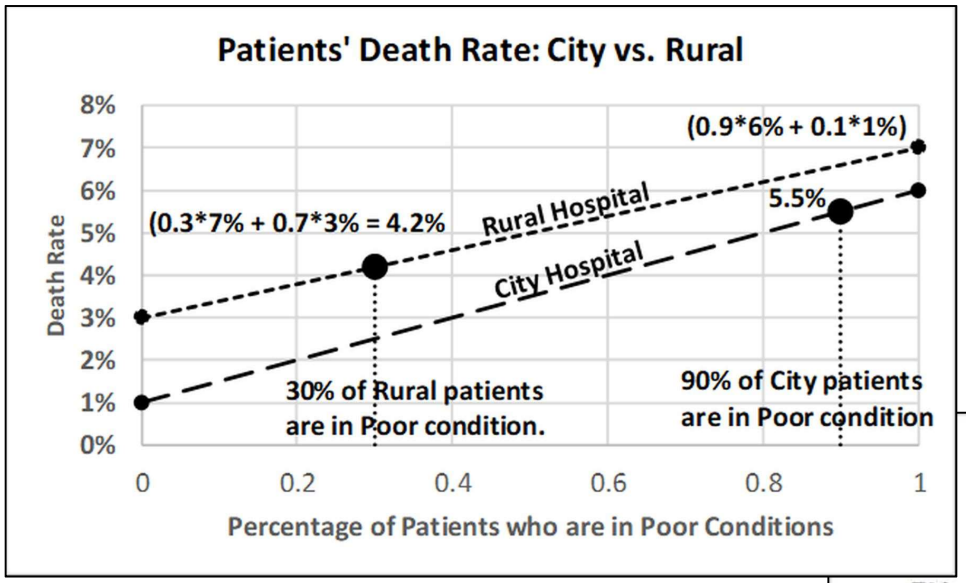
Crude compare

Mixed-fruit

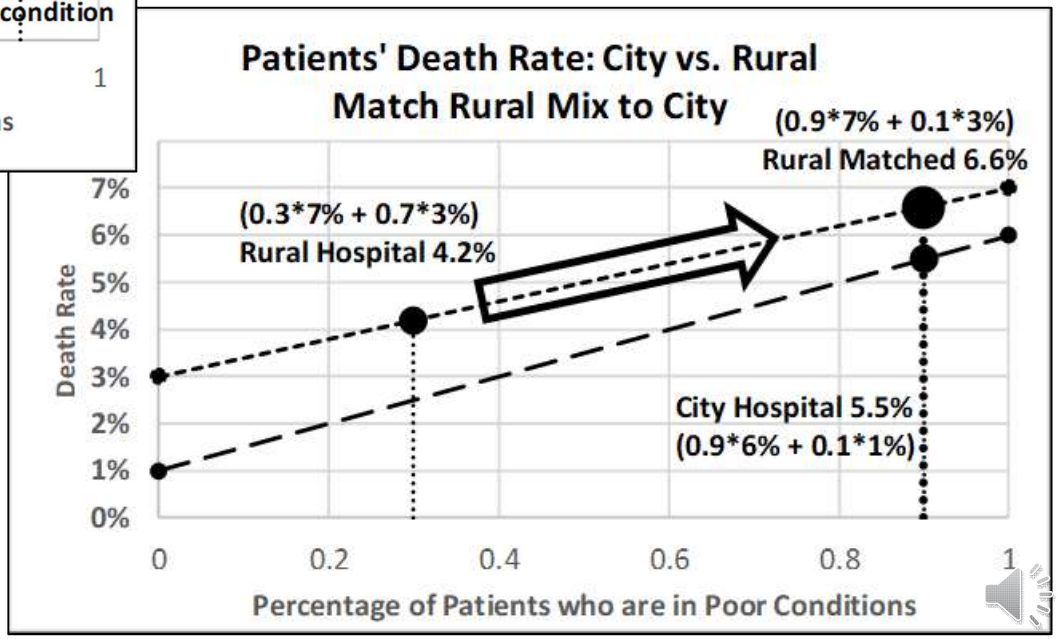
Match Rural to City. After *controlling for patient condition*, the death rate is higher at Rural than at City.

Match Rural Mix to City; Apply City Mix to Rural			
Patients' Death Rate (Mix: Percentage in this condition)			
Hospital	Good Cond.	Poor Cond.	All
City	1% (10%)	6% (90%)	5.5%
Rural	3% **(10%)**	7% **(90%)**	**6.6%**
All: City	= 0.9*6% + (1-0.9)*1%		**-1.1 pts**
All: Rural	**= 0.9*7% + (1-0.9)*3%**		City lower

Standard Group: Graph #1G: Adjust Rural Mix to Match City



Wainer (2002), Schield (2006)
Patient death rate lower at Rural hospital than at City.
[Mixed-fruit comparison]



Match Rural mixture to City.
After *controlling for* patient condition, death rate higher at Rural than at City. [Reversal]
Apples and apples comparison.

Combined Mix: Algebra #2A: Adjust All Mixes to Combined

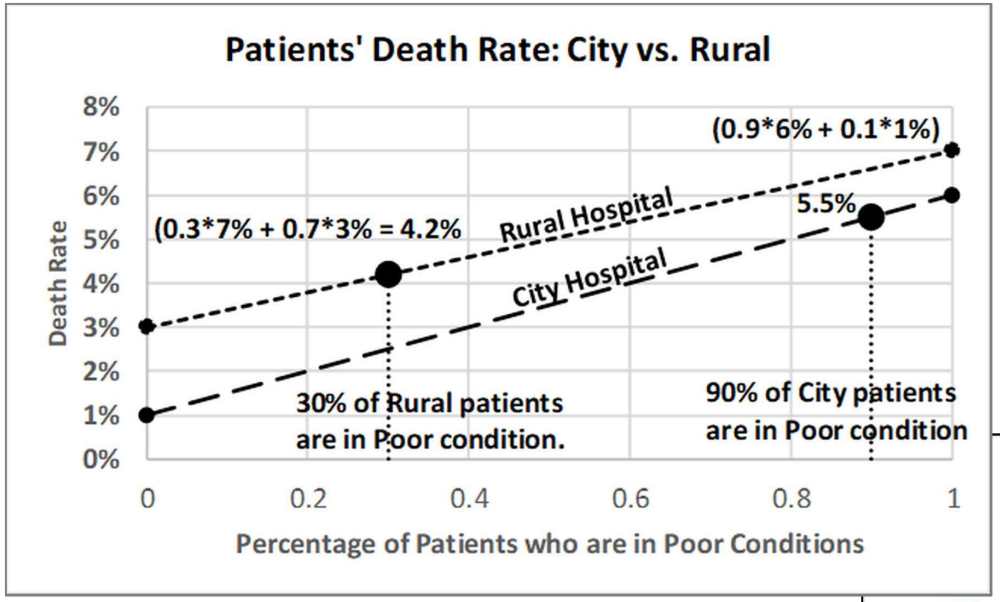
Patients' Death Rate (Mix: Percentage in this condition)			
Hospital	Good Cond.	Poor Cond.	All
City	1% (10%)	6% (90%)	5.5%
Rural	3% (70%)	7% (30%)	4.2%
All: City	= 0.1*1% + 0.9*6%		1.3 points
All: Rural	= 0.7*3% + 0.3*7%		City higher

Crude compare Mixed-fruit

Match to combined: 70%
After controlling for patient condition, death rate is higher at Rural than at City. [Reversal]
Apples & apples compare

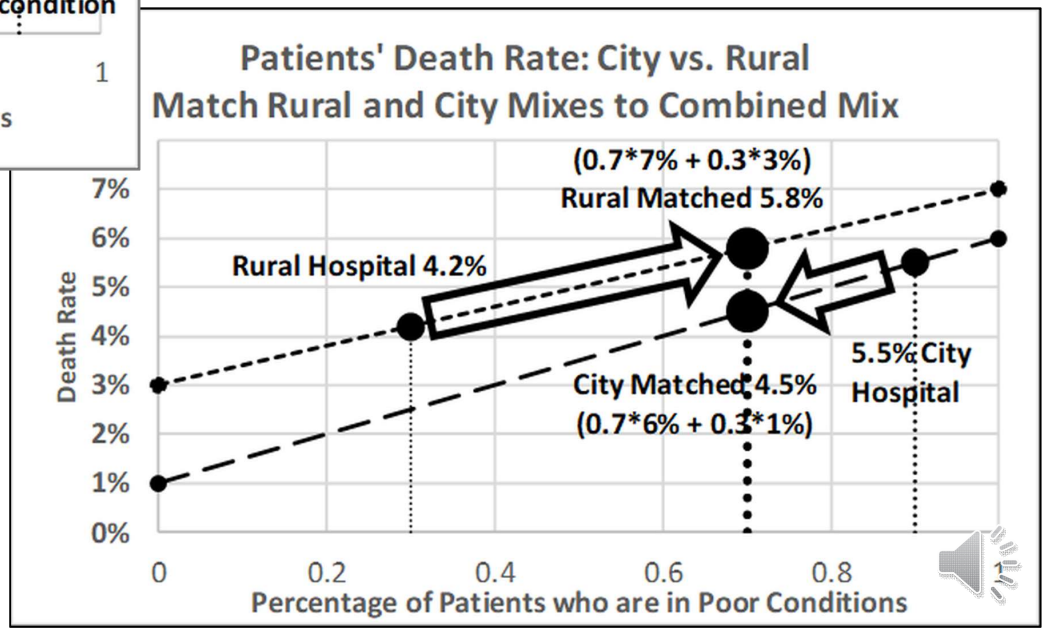
Match City & Rural Mixes to Combined Mix: 70%			
Patients' Death Rate (Mix: Percentage in this condition)			
Hospital	Good Cond.	Poor Cond.	All
City	1% **(30%)**	6% **(70%)**	**4.5%**
Rural	3% **(30%)**	7% **(70%)**	**5.8%**
All: City	**= 0.3*1% + 0.7*6%**		**-1.3 pts**
All: Rural	**= 0.3*3% + 0.7*7%**		City lower

Combined Mix: Graph #2G: Adjust All Mixes to Combined



Patient death rate higher at City hospital than at Rural. [Mixed-fruit comparison]

Standardize on combined mix. After controlling for patient condition, death rate is higher at Rural than at City. [Reversal] Apples and apples comparison.



Compare the Calculations: Algebraic vs. Graphical

PLUS: *Algebraic techniques* seem to be

- simpler for teachers to teach (graphs take time),
- simpler for students to do (graphs are tricky), and
- more applicable (applies to more than two groups)

as compared with the Wainer (2002) graph approach.

MINUS: *Algebraic techniques*

- are calculation based (not visual. Students can't see)
- are very sensitive to the wording (match, apply) and
- give different results depending on the standard.

Both let Students work Problems. Might Teaching Both Be Best?

I've taught combined-group graphs (#2G) for almost 10 years. This past year I taught standard-group algebra (#1A).

Suppose you start with *standard-group algebra* (#1A): it is simpler to teach and simpler to do. Then have students show their results using Wainer's graph (#1B).

Depending on time, introduce *combined-group algebra* (#2A). Have students show their results in a graph (#2B).

Doing the visual graphical technique should help students **see and understand** what the algebraic technique is doing.

Bibliography: The Graphical Technique

Wainer, H. (2002). "The BK-Plot: Making Simpson's Paradox Clear to the Masses." *CHANCE*, 15(3):60-62. www.statlit.org/Wainer.htm www.statlit.org/CP/Cornfield/2002-Wainer-Visual-Revelation-BK-plot.pdf

Schield, M. (2006). Understanding Confounding from Lurking Variables using Graphs. *STATS Magazine* ASA. Fall 2006. pp. 14-18. See www.StatLit.org/pdf/2006SchieldSTATS.pdf

Thanks to Marc Isaacson for the title of – and comments on – these slides.