

Qualitative Methods and Quantitative Data

Talk in the UW QUAL Series

Steven I. Wilkinson

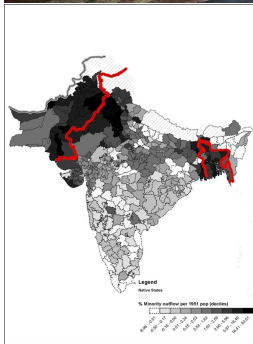
Yale University

May 26, 2021

How to think about QAM in light of increasing quantification?

- Drive for quantification and causal identification has accelerated the hunt for big data, historical and contemporary.
- Articles in PLSC, SOC & Econ in *Jstor* with keywords 'natural experiment' 'data' and 'archive' increased from 227 in 1990-1999 to 464 in 2000-09 to 981 in 2010-19.
- In *Google Scholar* cites with keywords 'natural experiment,' 'data,' 'archive' and 'political science' increased from 10,600 in 1980-1989 to 18,100 in 1990-99, 24,700 in 2000-09 and 20,100 in 2010-19.
- South Asia a particular focus: e.g. in Economics Dave Donaldson's (2017) paper on "Railroads of the Raj" *AER*; Banerjee and Iyer's, (2005) paper on Land tenure.

What's the role of QAM in this?



Yale QAM field

- 1 Fieldwork Course
- 2 Archival Methods Course
- 3 Mixed-Methods Course/Philosophy of Science Course

We need to be clear and confident about what fieldwork and qualitative methods give us that no other methods provide

- Historical and contemporary data don't exist on many key issues, often on purpose. When Leopold had to hand back the Congo to Belgium in 1908, a military aide at the palace wondered why the radiators were hot, and asked the janitor: "...he replied, "Sorry, but they're burning the state archives." The furnaces burned for eight days, turning most of the Congo state records to ash and smoke in the sky over Brussels. "I will give them my Congo," Leopold told Stinglhamber, "but they have no right to know what I did there." (Hochschild, 1998)
- Unmatched insights of deep ethnographic, interview and historical work

But I want to make three arguments on why qualitative methods are also absolutely vital for good quantitative methods

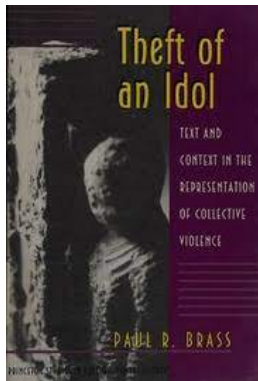
1. QAM are *crucial* to understanding whether big data are good data

TABLE 1A.3
Total IPC & SLL Crimes (State/UT-wise) - 2017-2019

S. No	State/UT	2017	2018	2019	Percentage Share of State (2019)	Mid-Year Projected Population (in Lakhs) (2019)	Rate of Cognizable Crimes (IPC+SLL) (2019)+
1	2	3	4	5	6	7	8
STATES:							
1	Andhra Pradesh	148002	144703	145751	2.8	523.2	278.6
2	Arunachal Pradesh	2746	2817	2877	0.1	15.1	190.9
3	Assam	109952	120573	132783	2.6	344.2	385.8
4	Bihar	236055	262815	269109	5.2	1201.1	224.0
5	Chhattisgarh	90516	98233	96561	1.9	288.5	334.7
6	Goa	3943	3884	3727	0.1	15.4	241.5
7	Gujarat	334799	393194	431066	8.4	682.5	631.6
8	Haryana	224816	191229	166336	3.2	288.1	577.4
9	Himachal Pradesh	17796	19594	19924	0.4	73.2	272.4
10	Jammu & Kashmir	25608	27276	25408	0.5	135.3	187.8
11	Jharkhand	52664	55664	62206	1.2	375.8	165.5
12	Karnataka	184063	163416	163691	3.2	659.7	248.1
13	Kerala	653500	512167	453083	8.8	351.9	1287.7
14	Madhya Pradesh	379682	405129	395619	7.7	826.1	478.9
15	Maharashtra	467753	515674	509433	9.9	1225.3	415.8
16	Manipur	4250	3781	3661	0.1	31.1	117.7
17	Meghalaya	3952	3482	3897	0.1	32.3	120.6
18	Mizoram	2738	2351	2880	0.1	12.0	241.0
19	Nagaland	1553	1775	1661	0.0	21.6	77.1
20	Odisha	103866	107408	121525	2.4	437.3	277.9
21	Punjab	70673	70318	72855	1.4	299.4	243.3
22	Rajasthan	245553	250546	304394	5.9	776.0	392.3
23	Sikkim	979	869	821	0.0	6.7	123.5
24	Tamil Nadu	420876	499188	455094	8.8	758.1	600.3
25	Telangana	133197	126858	131254	2.5	372.8	352.0
26	Tripura	4238	6078	5988	0.1	40.0	149.6
27	Uttar Pradesh	600082	585157	628578	12.2	2259.7	278.2
28	Uttarakhand	28861	34715	28268	0.5	111.8	252.8
29	West Bengal	195537	188063	188063	3.6	971.1	193.7
TOTAL STATE(S)		4748250	4796957	4826513	93.6	13135.2	367.4
UNION TERRITORIES:							
30	A&N Islands	3014	3699	4034	0.1	4.0	1013.6
31	Chandigarh	5462	5967	4518	0.1	11.8	381.6
32	D&N Haveli	309	315	290	0.0	5.6	52.2

- South Asian Crime Data (CII, 2019)
- Number of communal riots in a state (Wilkinson, 2004)
- Infrastructure data (National Statistics vs CAG)
- Number of Covid-19 deaths in India: 307k? 600k? 1.6 million? 6 Million? (NYT, 5/26/21)

2. Using QAM to understand what's 'wrong' with the numbers generates new questions, new ideas for better data, and new theories



3. QAM are *essential* to testing the core assumptions being made by our statistical models

Many statistical models assume that some *as random* process is at work, which allows us to assess the affect of some treatment of interest on similar territories or units, or that selection into treatment was independent of the characteristics of a unit.

We need QAM to see if these assumptions are correct.

- Was selection to one kind of land tenure or other in 19th century India really independent of characteristics of a district? (Banerjee and Iyer, 2005: 1205).
- Was princely state selection into British rule really as random? (Iyer 2010)

A few examples from Jha and Wilkinson's work on War and Political Change (2012-)

We try to explain veterans' role in ethnic cleansing in the partition of India



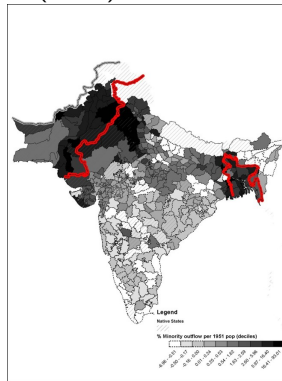
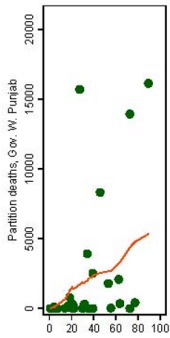
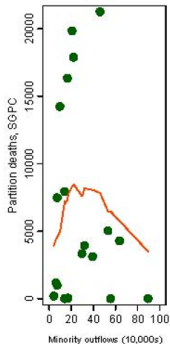
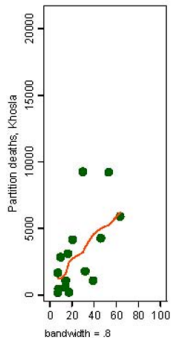
Margaret Bourke-White, 1947

Several Issues:

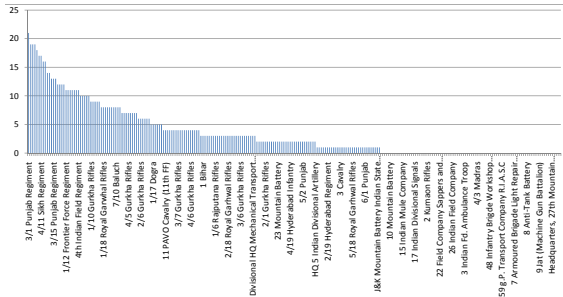
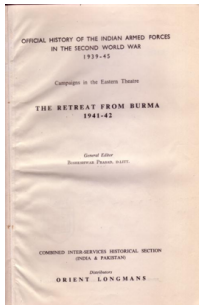
- How good are the data on deaths?
- No systematic data on soldiers' military experience in WW2 and their home districts
- Was this military experience 'as random'

Data on Deaths and Partition Violence

1947-50 counts and Bharadwaj, Khwaja, & Mian (2008)



Data on Soldier's Military Experience in WW2

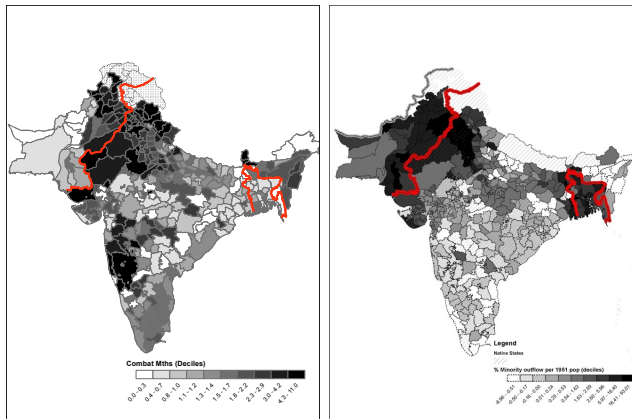


How can we match a unit's military experience to districts?



Name:	KHAN AFSAR
Nationality:	Indian
Rank:	Sepoy
Regiment/Service:	1st Punjab Regiment
Unit Text:	2nd Bn.
Age:	27
Date of Death:	17/06/1945
Service No:	22253
Additional information:	Son of Abdullah; husband of Rahim Jan, of Banitti, Rawalpindi, Pakistan.
Casualty Type:	Commonwealth War Dead
Grave/Memorial Reference:	Face 31.
Memorial:	RANGOON MEMORIAL

We can then use our measure of district combat experience to see if it predicts degree of ethnic cleansing in each district



Jha and Wilkinson (2012) and Bharadwaj, Khwaja, & Mian (2008)

QAM is vital to checking key assumptions of this quantitative work

- Was deployment of soldiers really 'as random' required trips to NAI, BL, and other archives, e.g. Slim Papers, 2/2, 6/6, Churchill Archives, Cambridge
- Do archive, press and secondary sources confirm soldiers were involved and using military organization and skills?

Jullundur Aug. 24. The [Punjab] mail-train had been diverted south from Ludhiana, and east of Ferozepur it was halted by obstructions on the line. All the Sikh passengers had been tipped off at a previous station and had left the train. After being halted the train was surrounded by armed Sikhs. The major noticed that they had advanced and retired in military formation, with a Sikh in blue commanding each section of about 12 men. They were driven off by shots fired by the British passengers and a few others who had arms, and by the train's armed escort, which consisted of two Muslim sepoy, two Hindus and one Sikh. The Times, 25 Aug 1947