

Join the Forces!

Enlisting citizens in the fight to transcribe military records

Evan Roberts, Sociology and Minnesota Population Center
University of Minnesota

@evanrobertsnz
@MeasuringANZACs

Measuring the ANZACs

Do your part to help transcribe first-hand accounts of New Zealanders from the Australian and New Zealand Army Corps circa World War I.

[START MARKING](#)

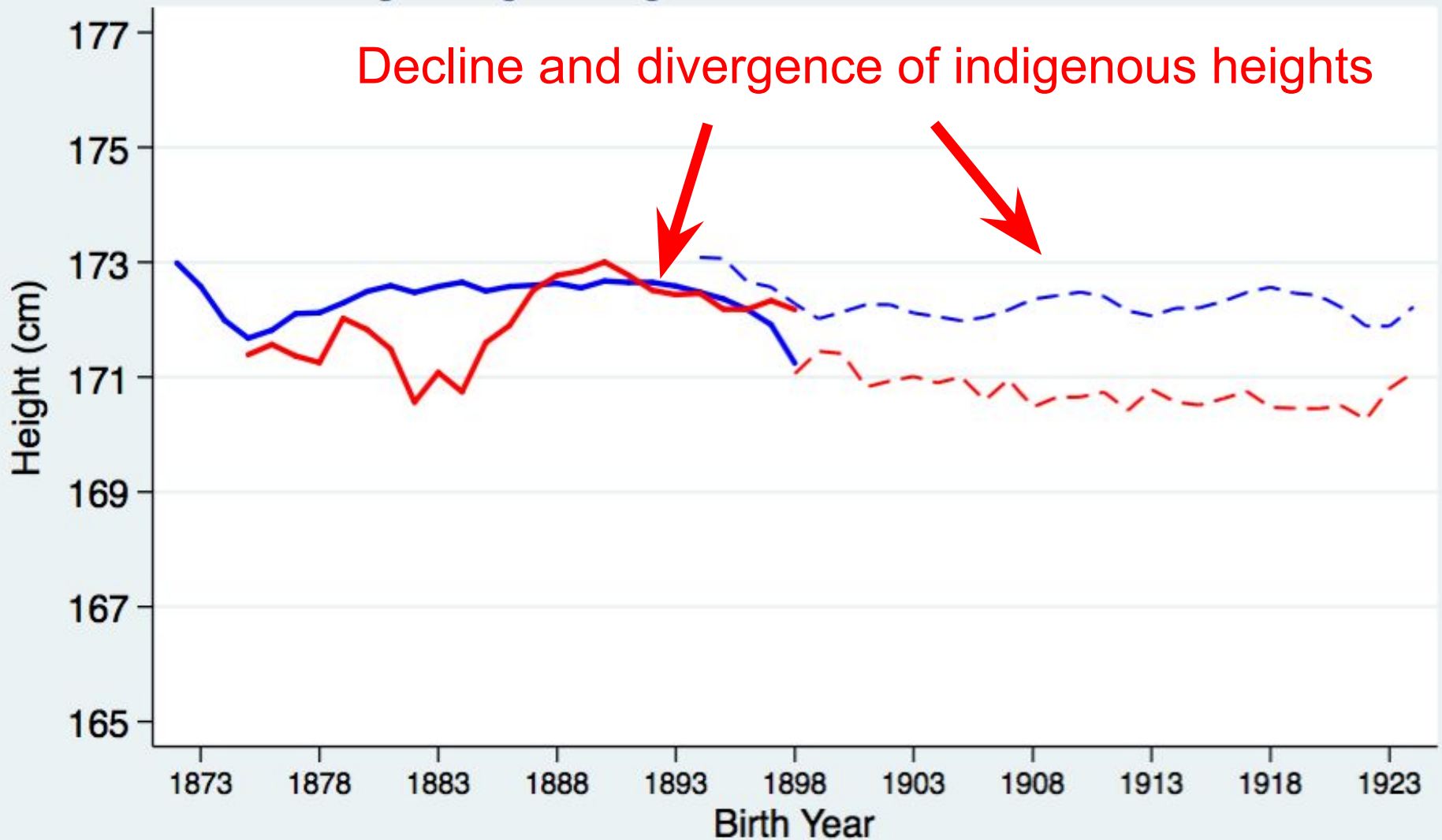
or

[START TRANSCRIBING](#)

SCRIBE

Rolling average of heights of New Zealand World War soldiers

Decline and divergence of indigenous heights



— Pakeha, WWI — Maori, WWI
- - - Pakeha, WWII - - - Maori, WWII

Limitations of our current sample

- ~20,000 records per war
- Can't do too many interactions in our models
- No information on combat exposure
- Nearly a thousand women served [as nurses]



What kind of page is this?



HISTORY SHEET

STATEMENT OF SERVICES

ATTESTATION: GENERAL
FORM

ATTESTATION: MEDICAL
FORM

DEATH NOTIFICATION

SOUTH AFRICAN WAR
ATTESTATION

OTHER

DONE

BAD PAGE

1. Classify pages to open up page-dependent menus and instructions

2. Citizen scientists identify fields and mark their location for transcription

1st 12/13

HISTORY-SHEET. E.F. Form No. 3A. Advice despatched to soldier: 20/1/10

Unit: 1st Battalion 5th Regiment T-1st

Rank: 2nd Lieut

Surname: Lt Cormick

Christian Name: Alexander Leslie

No. 23/334

Religion: Protestant

Last New Zealand address: Waipukuruan

Occupation: Grocer

Last employer: B.G. Mc Cormick Waipukuruan

Name, relationship, and address of next-of-kin (if not resident in New Zealand, insert also name and address of nearest relative in New Zealand):
 (Father) Mr B.G. Mc Cormick
 Waipukuruan

Service ...

No. 23/334 Rank: 2/Lt

Name: A.L. Cormick

Address: Waipukuruan

Periods of Service.		Theatres of Operation.	
In New Zealand:	years 192 days.	Australasian	
Overseas ...	3 years 104 days.	Egyptian	1915-16
Total service ...	3 years 296 days.	Egyptian E.F.	1916
Date commenced duty:	28/4/15	Balkans	
		Western	1916-17

Select a label and draw rectangle around any corresponding entries on the document.

- Unit (7) ?
- Rank (7)
- Surname (6)
- Christian name (8)
- No. (7)
- Occupation (6)
- Last Employer (6)
- Religion (6)
- Last New Zealand Address (8)
- Next-of-Kin (4)

NEXT

Y-SHEET. despatched to at E.F. Form No. 3A.

Authority to Headquarters at *Table*

Christian Name: *Aloysius*

No: *5/227*

James Aloysius John

No: *7/227*

Religion: *Roman*

Last New Zealand address: *Raikoura*

6 MAR 1902

Mackie (father)

ough

Mackie

Raikoura

Theatres of Operation.

What kind of page is this?

- HISTORY SHEET
- STATEMENT OF SERVICES
- ATTESTATION: GENERAL FORM
- ATTESTATION: MEDICAL FORM
- DEATH NOTIFICATION
- SOUTH AFRICAN WAR ATTESTATION
- OTHER
- DONE**
- BAD PAGE

- [View A Tutorial](#)
- Transcribe this page now!
- [Discuss this personnel record.](#)



Show All

3. Transcribe marked fields via three routes ...

Measuring the ANZACs

Do your part to help transcribe first-hand accounts of New Zealanders from the Australian and New Zealand Army Corps circa World War I.

[START MARKING](#)

or

[START TRANSCRIBING](#)

SCRIBE

Essentially random entries not yet transcribed

Rank	Surname.	Christian Name.	No.				
TRANSCRIBE	ABOUT	SEARCH	FIELD GUIDE	BLOG	DISCUSS	HELLO EVAN ROBERT	
PVTE	Macaw	James Courtney	38196				
Religion:	Last New Zealand address:						
Ireland C.E.	Hastings						
21-12-80							
Mrs K. Macaw (Mother)	T.G. D...						
Innisrush.							
Ireland.							
No. 38196	Rank: Pte						
Name: James Courtney Macaw							

Transcribe last New Zealand address. Press [Shift] + [Enter] to start a new line.

[Need some help?](#) [Bad mark](#) [Illegible?](#) [Next Entry](#)

Measuring the ANZACs data structure

- 147 variables/fields/columns from source material
 - Some variables describe categories of events that can occur multiple times
 - Sick, wounded, transfers between units, entering new combat zone
 - **You only die once**
- 6 different document types to identify
- Paradata : Data about the transcription of each element
 - 17 elements of paradata about each transcription or classification! Includes who, when, X/Y co-ordinates on page
- Significant automatically generated metadata

Measuring the ANZACs data structure

- 147 variables/fields/columns from source material
 - Some variables describe categories of events that can occur multiple times
 - Sick, wounded, transfers between units, entering new combat zone
 - **You only die once**
- 6 different document types to identify
- Paradata : Data about the transcription of each element
 - 17 elements of paradata about each transcription or classification! Includes who, when, X/Y co-ordinates on page
- Significant automatically generated metadata

Progress towards goals ...

- ~2 million fields transcribed
- 6,500 personnel files have been touched out of 140,000
- 58,518 pages of 3.7 million
- In short term: Shifts research questions towards the life course data

**Accuracy of paid research assistants
versus citizens?!**

	Attestations (NZ RAs)		Casualty Rolls (Canadian RAs)	
	Given	Last	Given	Last
Mean similarity score to truth (1 = absolute accuracy)	0.98	0.99	0.99	0.99
Proportion absolutely accurate	0.91	0.95	0.93	0.98
Proportion with similarity score > 0.9	0.96	0.98	0.95	0.99
Proportion with similarity score > 0.8	0.98	0.99	0.99	0.99
Proportion with > 3 words in string (more likely problematic transcriptions)	0.003	0.0002	0.030	0

	Citizen transcription		
	Given name	Surname	Serial number
Mean similarity score to truth (1 = absolute accuracy)	0.98	0.98	0.98
Proportion absolutely accurate	0.87	0.84	0.89
Proportion with similarity score > 0.9	0.95	0.95	0.95
Proportion with similarity score > 0.8	0.97	0.98	0.98
Proportion with > 3 words in string (more likely problematic transcriptions)	0.0029	0.0030	0.0030

Accuracy in context

Example: Charles and Cjarles have a similarity of 0.91

Example: Gerry Reid and Henry Reid have a similarity of 0.8

Example: B and Benjamin have a similarity of 0.74

Conclusion

Citizen scientists make more small
(one character) mistakes

In the vast majority of cases highly useable data

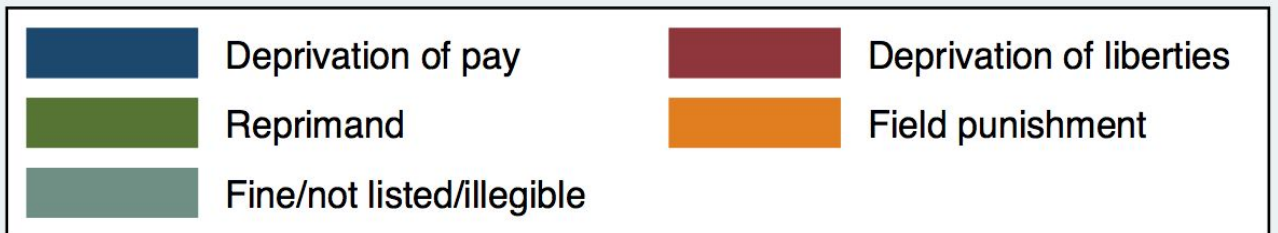
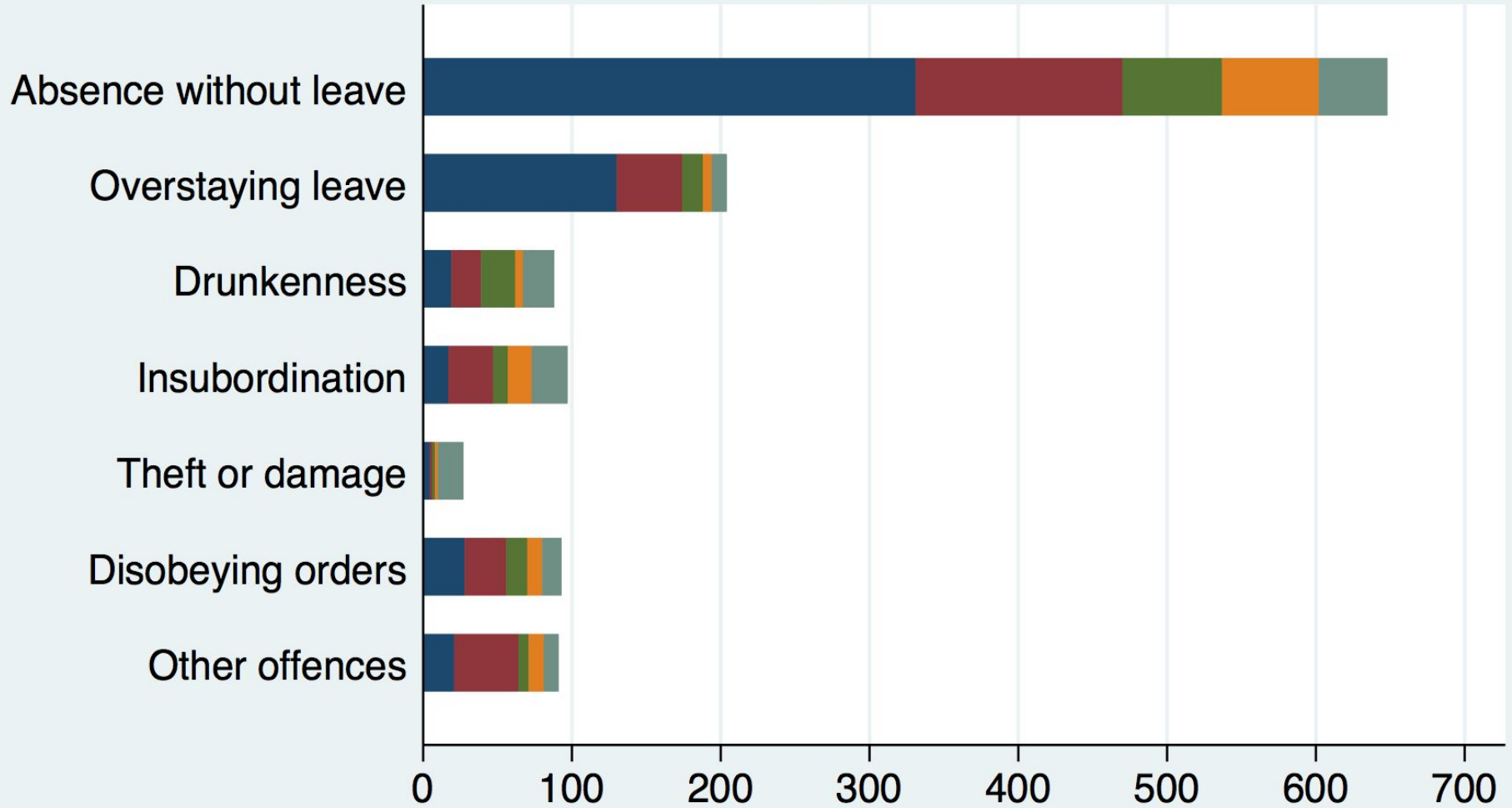
Drunken and disorderly?

CONDUCT-SHEET.

Regiment or Corps.	Place.	Date.	Offence.	Punishment.	Authority for Entry
80 Coy 11 th	Leatherston	5/3/16	Overlapping leave	Forfeited one day pay by R.W.	
33rd Bde 3rd B. G. B. P.	In the Field	23.8.16	On leave out of billet 22.8.16 to 23.8.16	4 days F.P. No 1 Deprived 14 days pay.	A.F.P. 2069 8/16 39, 3
D. Coy	"	20.10.16	abs. from 3 P.M. 20/10/16 to 4 " - 25/10/16 falling out of ranks.	" 7 "	" " " " " "
n. 7 Inf Gen Base Dep	Etaples	5.10.17	Abs w/pl from billet		
		8.10.17	Roll Call 9.30 on 5.10.17 until 2.00 and owing to billeting	28 days F.P. No 2 & 20 days pay with 30 days pay by R.W.	A.F.P. 2069 9.10.17
			7.40. a.m. on morning of 8.10.17.		
			- w. o. a. s. -		
n. 7. R. B.	A'ch.	27.10.18	(a) neglecting to obey the Standing Orders in starting to hand in pass at Police Office	30 days pay by award of 20/10/18	
			(b) neglecting to obey the standing orders in breaking into the mess	(20 days pay gross pay)	A.F.P. 2069 (2177)
			abs. for picquet	Times 10/ + one extra picquet awarded 20-5/15	S.R. D. No 4 P. 11.

23
5
10
9

Offences and punishments in the NZEF



Offences and punishments in the NZEF

