

Dhaval Adjodah -
HST.S14 Final Project:
The Malawi Dataset -
May 10, 2012 -

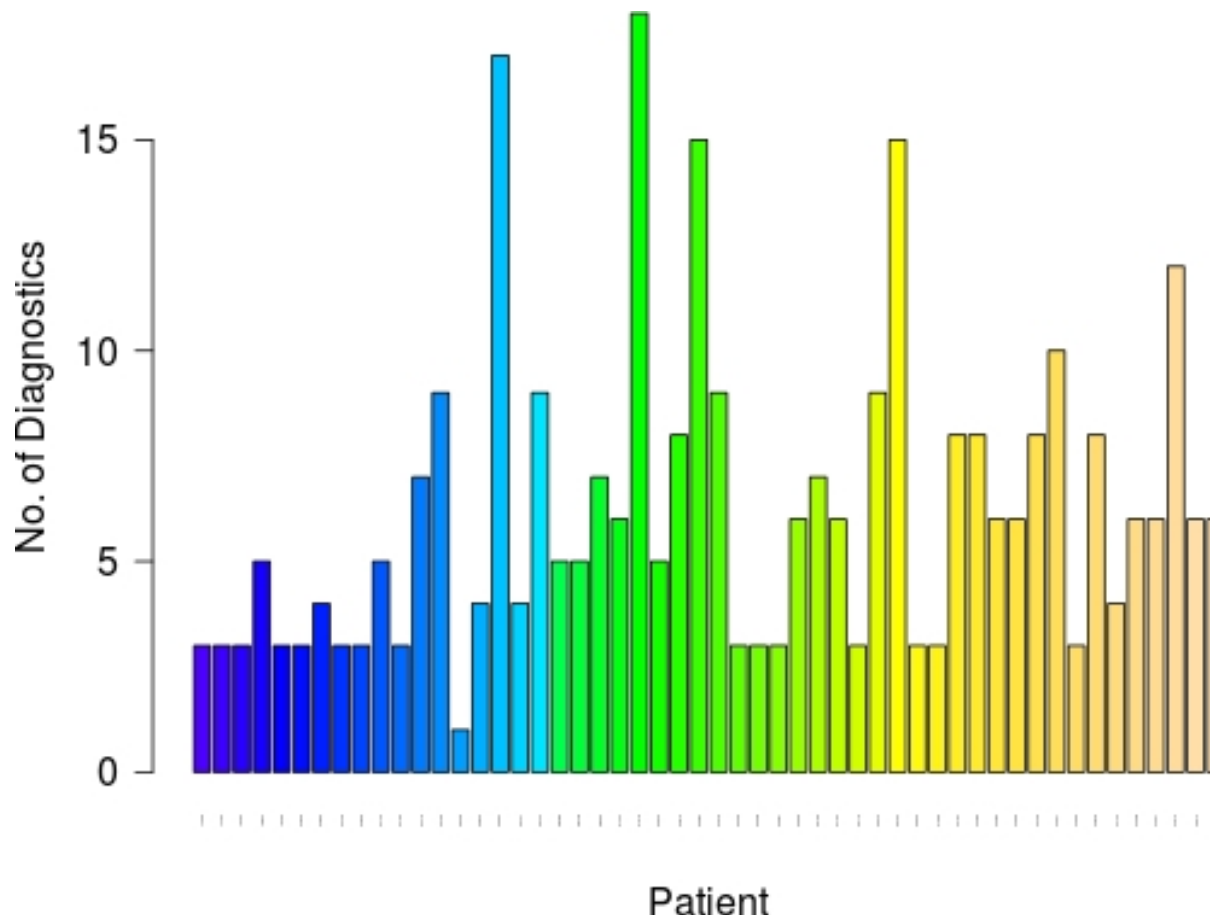
- Unfortunately, I was not able to come up with a statistical prediction model for the diagnostics that lead to somebody having Tuberculosis
 - There was far too little data to reach any such advanced statistical computation
 - However, I found interesting tidbits of information about the structure of the data

- A dataset of patients who have a TB diagnosis at some point or the other was pulled, together with their preceding and subsequent other diagnostics.
- It was found out that 52 patients got TB sooner or later
- They have in total 2213 diagnostics with the PIH Malawi staff of which 197 were unique (a single encounter can yield many diagnostics).
- The idea was to be able to analyze this large set of diagnostics for a trend in diagnostics: perhaps all people who get TB come to the clinic beforehand 2-3 times with a pneumonia diagnostic.

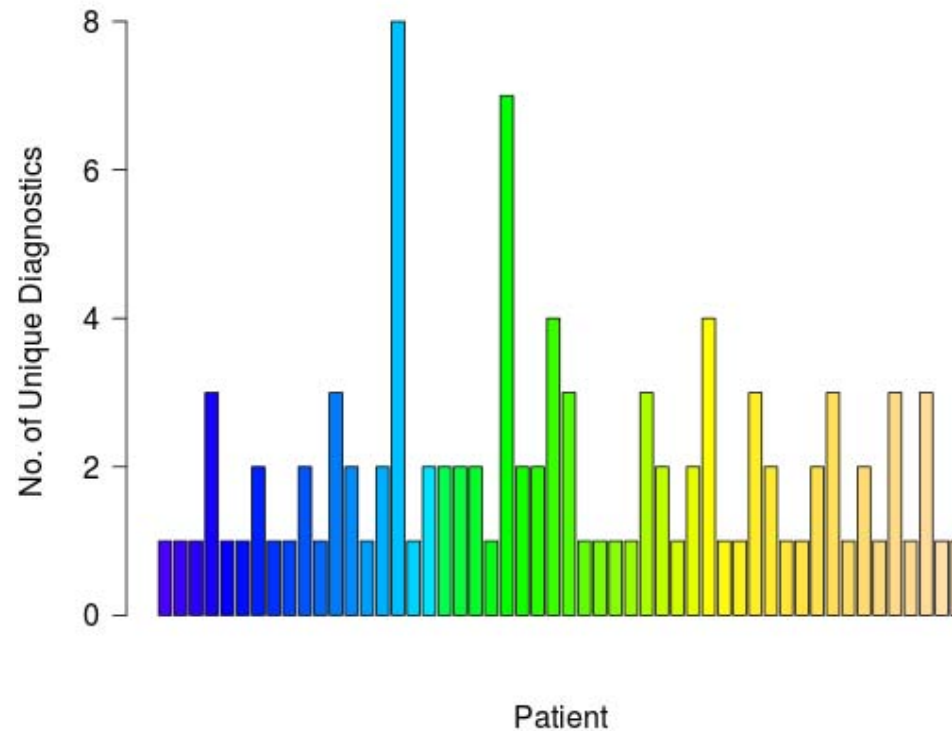
- Upon closer inspection, it was found that a lot of diagnostic rows were duplicated for the same patient id and the same encounter id (and date-time).
- There were many variations of the same ICD-9 diagnostic name: TB had about 6 different names,
 - some of them actually medical variants,
 - some of them different shorthands
- The most extreme example in patient 19027: he shows up on 2009-09-14 at 10:29:34 and receives no less than 63 diagnoses that day.

- Data had to be filtered:
 - remove duplicated diagnoses for the same patient at the same time,
 - to manually find possible variations of the same names that previously appeared to be two different diagnoses (with some help from Evan Waters),
 - to remove diagnoses that appear to be blatant errors in the system such as “PRESUMED” and “CLINICAL” and
 - to remove all diagnoses that are not respiratory-related (with Dr. Fraser's help)

- A filtered data-set was produced:
 - it still contained 52 unique patients,
 - but contained only 325 diagnostics.
 - Each patient had a mean of 6.25 different diagnostics although in most cases, this occurred on the same day.



- The problem is that for each patient who has more than 1 diagnostic.
- Since we are interested in seeing the progression of diagnostics over time, we need the distinct encounters (on different dates) each patient had.
- The mean number of unique encounters by patients is 2.0.

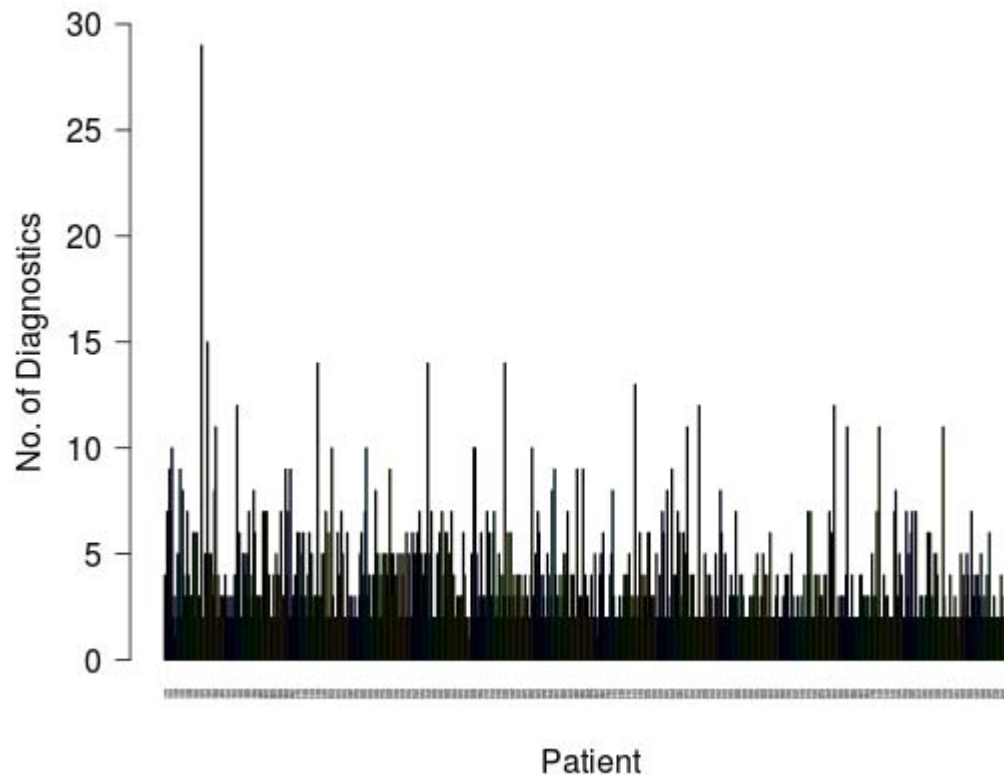


48 57 6	229 261	2010-Feb-17 8:45:00	PERIPHERAL NEUROPATHYEUMONIA NOS			
48 57 6	229 261	2010-Feb-17 8:45:00	PERIPHERAL NEUROPATHY			
48 57 6	229 261	2010-Feb-17 8:45:00	PULMONARY TUBERCULOSIS			
48 57 6	229 261	2010-Feb-17 8:45:00	PERIPHERAL NEUROPATHYEUMONIA, TUBERCULOSIS			
48 57 6	229 261	2010-Feb-17 8:45:00	TUBERCULOSIS PERIPHERAL NEUROPATHYEUMONIA			

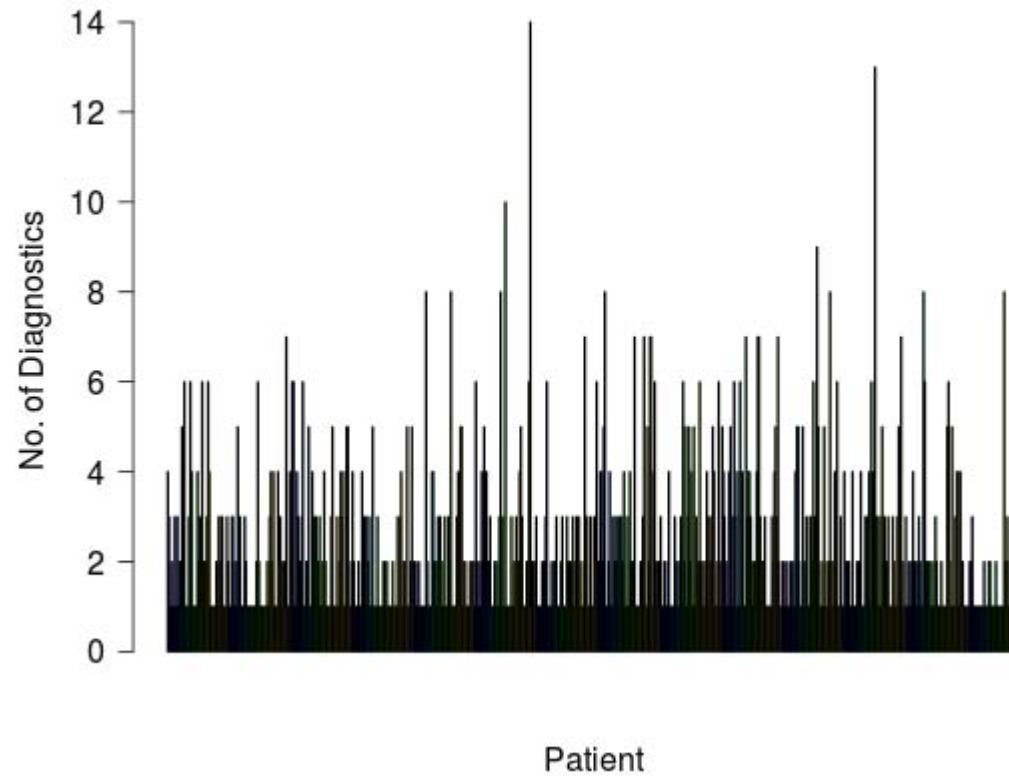
	2008-Dec-02 10:11:00	PERIPHERAL NEUROPATHYEUMONIA NOS			
	2008-Dec-02 10:11:00	PERIPHERAL NEUROPATHY			
	2009-Mar-24 21:01:00	PERIPHERAL NEUROPATHYEUMONIA			
	2009-Mar-24 21:01:00	PERIPHERAL NEUROPATHYEUMONIA NOS			
	2009-Mar-24 21:01:00	PERIPHERAL NEUROPATHY			
	2009-Mar-24 21:01:00	PULMONARY TUBERCULOSIS			
	2009-Mar-24 21:01:00	PERIPHERAL NEUROPATHYEUMONIA, TUBERCULOSIS			
	2009-Mar-24 21:01:00	TUBERCULOSIS PERIPHERAL NEUROPATHYEUMONIA			

2009-Feb-17 14:10:00	PERIPHERAL NEUROPATHYEUMONIA NOS			
2009-Feb-17 14:10:00	PERIPHERAL NEUROPATHY			
2009-Mar-23 21:28:00	TUBERCULOSIS			
2009-Mar-23 21:28:00	ACTIVE TUBERCULOSIS			
2009-Mar-23 21:28:00	TUBERCULOSIS NOS			
2009-Mar-31 14:13:00	TUBERCULOSIS			
2009-Mar-31 14:13:00	ACTIVE TUBERCULOSIS			
2009-Mar-31 14:13:00	TUBERCULOSIS NOS			

- A similar analysis was done on patients with pneumonia.
 - 721 unique patients were found who had a diagnosis of pneumonia sooner or later.
 - These 721 patients generated 2,622 encounters, with a mean number of encounters per patient of 3.6.



- Plotting only unique encounters, we find that the average unique encounters per patient is 2.3.



- Plotting only unique encounters, we find that the average unique encounters per patient is 2.3.

3/24/2009 21:01	PNEUMONIA
3/24/2009 21:01	PNEUMONIA NOS
3/24/2009 21:01	PNEUMONIA, TUBERCULOSIS
3/24/2009 21:01	TUBERCULOSIS PNEUMONIA
3/24/2009 21:01	PULMONARY TUBERCULOSIS

2/4/2009 14:29	PNEUMONIA NOS
3/17/2009 20:55	PNEUMONIA
3/17/2009 20:55	PNEUMONIA NOS
5/12/2009 17:59	PNEUMONIA
5/12/2009 17:59	PNEUMONIA NOS
5/12/2009 17:59	PNEUMONIA, TUBERCULOSIS
5/12/2009 17:59	TUBERCULOSIS PNEUMONIA
5/12/2009 17:59	PULMONARY TUBERCULOSIS
6/3/2009 21:04	URESPIRATORY TRACT INFECTION
3/17/2009 20:55	RECURRENT OR CHRONIC RESPIRATORY TRACT INFECTIONS (SINUSITIS, OTORRHOEA, TONSILLITIS, OTITIS MEDIA)
3/17/2009 20:55	RECURRENT OR CHRONIC RESPIRATORY TRACT INFECTIONS

2/4/2010 11:51	PNEUMONIA NOS
2/4/2010 11:51	PNEUMONIA, TUBERCULOSIS
2/4/2010 11:51	TUBERCULOSIS PNEUMONIA
2/8/2010 10:12	PNEUMONIA
2/8/2010 10:12	PNEUMONIA NOS
2/8/2010 10:12	PNEUMONIA, TUBERCULOSIS
2/8/2010 10:12	TUBERCULOSIS PNEUMONIA
2/10/2010 10:34	LOBAR PNEUMONIA
2/17/2010 9:03	PLEURISY
2/4/2010 11:51	PULMONARY TUBERCULOSIS
2/8/2010 10:12	PULMONARY TUBERCULOSIS

Questions?

MIT OpenCourseWare
<http://ocw.mit.edu>

HST.S14 Health Information Systems to Improve Quality of Care in Resource-Poor Settings
Spring 2012

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.