

# Image Based Data Mining Workshop

---

Andrew Green

With Material from Eliana Vasquez Osorio

# Image Based Data Mining

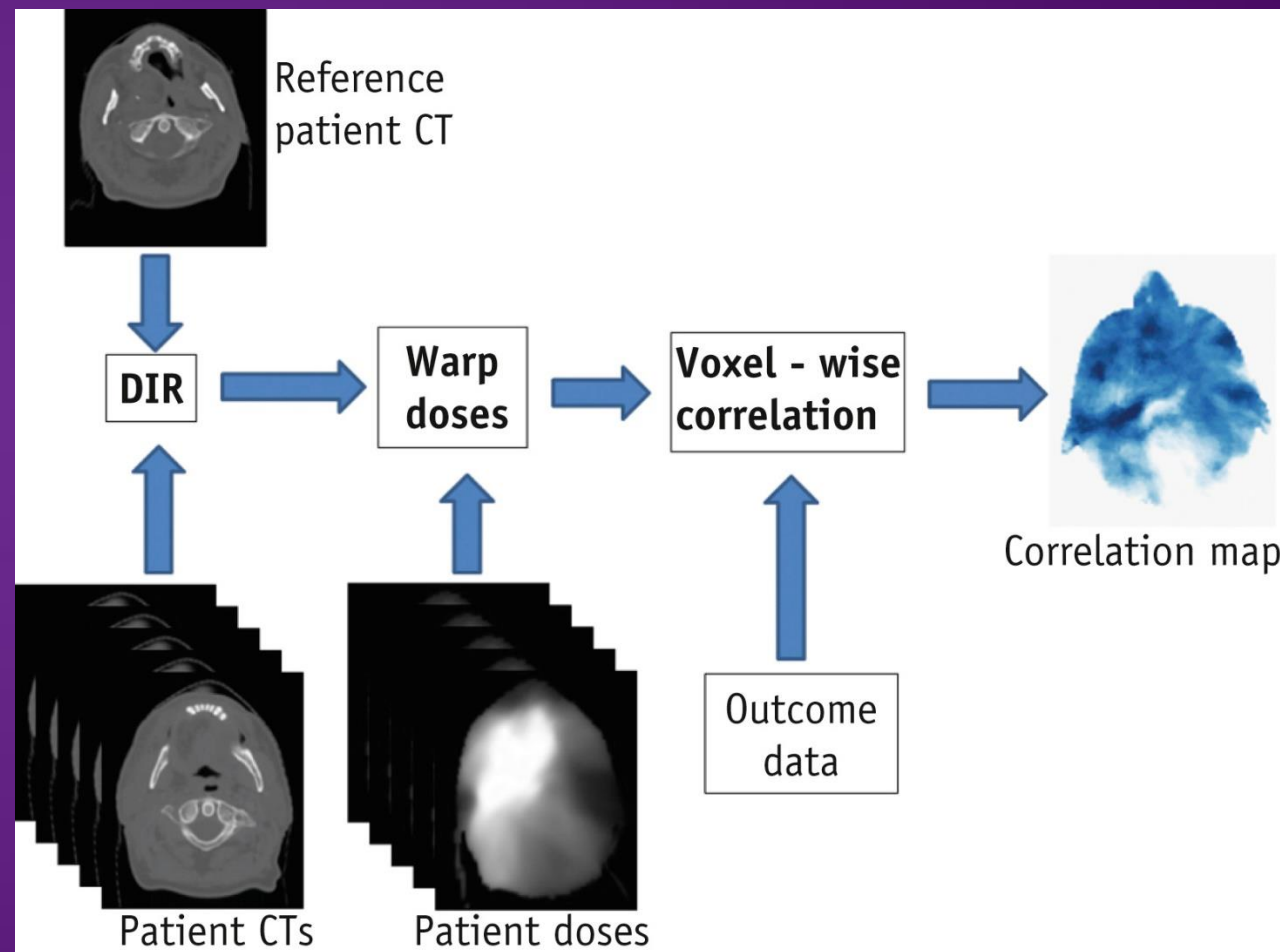
- Technique to relate planned RT dose to observed outcome
  - Makes maximum use of imaging
  - Minimum assumptions (no delineations needed)
- Relatively simple, with a few caveats
  - Multiple comparisons
  - Inter-voxel correlations
- Usually, we would take weeks/months over an IBDM analysis
  - We've got 1 hour

# The plan for this workshop

- Example IBDM workflow in python
  - Three notebooks to have a play with
  - Illustrative example, not necessarily best practise
- Focus on binary IBDM
  - Registration and continuous left as exercise for the reader!
  - Short walkthrough, remainder to play/discuss

## IBDM basics

- Three steps:
  - Data collection
  - Spatial normalization
  - Statistical analysis
- Notebooks do these, though not necessarily in order!



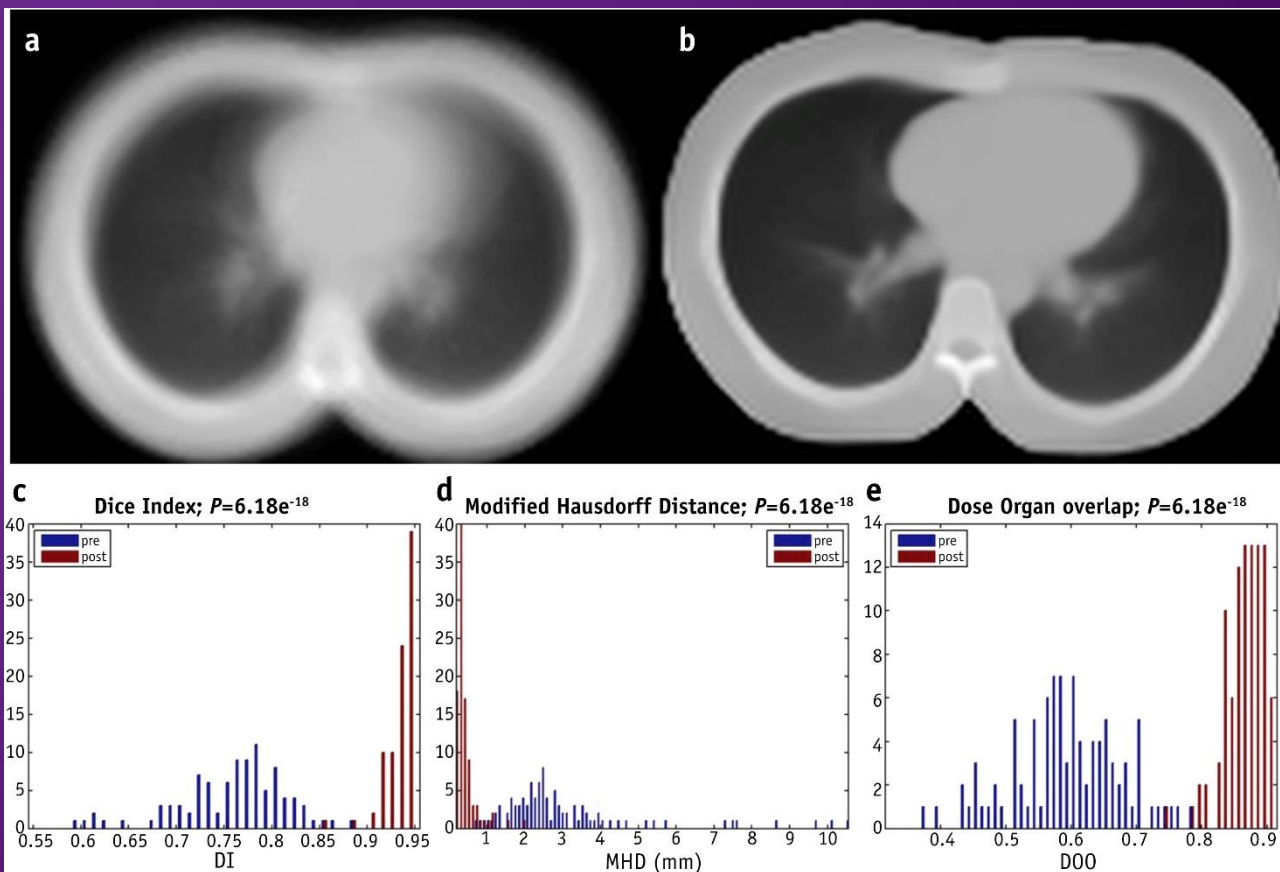
From <https://doi.org/10.1016/j.ijrobp.2018.05.054>

- IBDM cohort sizes can be 100s-1000s of patients
  - Need as much information as possible
  - Filter data to remove outliers
  - Conditioning on clinical variables later
- Required data:
  - Planning CT
  - Planned Dose
  - Outcome of interest

	T	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	X	Y	
1	ID	Sex	Age	Date of Birth	Day	Site	Histology	Grade	T	N	M	Stage	Date of D	Last Con	Follow up	Follow up	Follow up	Date of D	Survival	Alive or Dead	Cause of	Date of re	Disease-1	Site of re	Cl	
1	HN5CC-01-00 Female	54	17/02/1994	CA soft p	Orophary	SCC	moderate	2	2c	0	0	IVA	#####	#####	3032	8.5169	10.2	#####	99.6	Alive				10107		
2	HN5CC-01-00 Female	56	27/05/1945	NPC	Nasophary	SCC	moderate	4	2	0	0	IVA	#####	#####	1278	3.5014	42.016	#####	42.016	Dead	HN Can	#####	36.553	Local rec		
3	HN5CC-01-00 Male	49	08/01/1953	CA tonsil	Orophary	SCC	moderate	2	2b	0	0	IVA	#####	#####	1261	5.2275	52.73	#####	51.933	Alive	HN Can	#####	62.033	Complete		
4	HN5CC-01-00 Male	65	13/05/1931	CA tonsil	Orophary	SCC	poorly df 4a	2	0	0	0	IVA	#####	#####	409	4.1469	13.787	#####	13.433	Dead	HN Can	#####	9.3	Distant m		
5	HN5CC-01-00 Male	66	08/06/1931	CA tonsil	Orophary	SCC	moderate	2	1	0	0	III	#####	#####	3357	9.4238	13.16	#####	11.27	Alive			27.433	Local rec		
6	HN5CC-01-00 Male	62	11/03/1934	CLUP	CLUP	SCC	moderate	0	2a	0	0	IVA	#####	#####	3621	10.71	122.06	#####	110.933	Dead	Non can	#####	12.07	Complete		
7	HN5CC-01-00 Male	39	26/11/1957	CA BOT	Orophary	SCC	poorly df	2	3	0	0	IVB	#####	#####	3542	9.7041	116.45	#####	116.45	Alive			116.45	Complete		
8	HN5CC-01-00 Male	60	26/07/1936	CA BOT	Orophary	SCC	moderate	2	1	0	0	III	#####	#####	3212	9.0225	108.27	#####	105.5	Alive			107.07	Complete		
9	HN5CC-01-00 Male	57	27/12/1933	CA tonsil	Orophary	SCC	moderate	3	2b	0	0	IVA	#####	#####	3444	9.6472	116.09	#####	113.13	Alive			114.8	Complete		
10	HN5CC-01-00 Male	63	13/06/1933	CA tonsil	Orophary	SCC	moderate	4	3c	0	0	IVA	#####	#####	2855	7.8219	93.863	#####	93.863	Dead	Unknown	#####	54.181	Locoreg		
11	HN5CC-01-00 Male	70	25/08/1927	CA larynx	Glotis	SCC	poorly df	3	0	0	0	III	#####	#####	3223	9.0534	108.64	#####	105.87	Dead	Other ca	#####	107.43	Complete		
12	HN5CC-01-00 Male	51	26/11/1945	CA BOT	Orophary	SCC	moderate	1	2a	0	0	IVA	#####	#####	3159	8.6548	103.86	#####	103.86	Alive			103.86	Complete		
13	HN5CC-01-00 Male	61	08/02/1945	CA BOT	Orophary	SCC	poorly df	2	2b	0	0	IVA	#####	#####	272	0.7564	3.1685	#####	8.9333	Alive			3.0667	Complete		
14	HN5CC-01-00 Male	58	03/03/1941	CA BOT	Orophary	SCC	poorly df	1	2b	0	0	IVA	#####	#####	2619	7.3587	88.281	#####	86.033	Alive			87.3	Complete		
15	HN5CC-01-00 Male	63	27/06/1934	CA BOT	Orophary	SCC	poorly df	1	2b	0	0	IVA	#####	#####	3234	8.8603	106.32	#####	106.32	Alive			106.32	Complete		
16	HN5CC-01-00 Male	54	16/08/1943	CA tonsil	Orophary	SCC	moderate	2	2a	0	0	IVA	#####	#####	2972	8.1425	37.71	#####	37.71	Alive			37.71	Complete		
17	HN5CC-01-00 Male	53	21/12/1944	CLUP	CLUP	SCC	well diff.	0	2b	0	0	IVA	#####	#####	3182	8.7178	104.61	#####	104.61	Alive			104.61	Complete		
18	HN5CC-01-00 Female	24	05/12/1973	NPC	Nasophary	SCC	poorly df	4	2	0	0	IVA	#####	#####	3194	8.7507	105.01	#####	105.01	Alive			105.01	Complete		
19	HN5CC-01-00 Male	61	20/07/1936	CA tonsil	Orophary	SCC	moderate 4a	2	3	0	0	IVB	#####	#####	576	1.18	13.46	#####	13.83	Dead	Other Ca	#####	13.2	Complete		
20	HN5CC-01-00 Male	60	23/06/1937	CA BOT	Orophary	SCC	moderate	2	2c	0	0	IVA	#####	#####	3043	8.5646	102.78	#####	100.17	Alive			22.267	Local rec		
21	HN5CC-01-00 Male	56	17/12/1941	CA tonsil	Orophary	SCC	poorly df	2	2b	0	0	IVA	#####	#####	3146	8.6192	103.43	#####	103.43	Alive			103.43	Complete		
22	HN5CC-01-00 Male	64	10/10/1933	CA tonsil	Orophary	SCC	poorly df	2	2b	0	0	IVA	#####	#####	2689	7.8371	88.405	#####	88.405	Dead	Unknown	#####	88.405	Complete		
23	HN5CC-01-00 Male	62	03/08/1938	CA BOT	Orophary	SCC	poorly df	3	2b	0	0	IVA	#####	#####	2843	8.0528	96.034	#####	93.6	Alive			94.967	Complete		
24	HN5CC-01-00 Female	79	26/12/1916	CA soft p	Orophary	SCC	moderate	3	2a	0	0	III	#####	#####	3031	8.516	102.17	#####	93.567	Alive			101.03	Complete		
25	HN5CC-01-00 Male	45	10/07/1953	CA supra Glotis	SCC	poorly df	well diff.	3	0	0	0	III	#####	#####	702	1.2249	26.25	#####	63.267	Alive			82.6	Complete		
26	HN5CC-01-00 Male	59	03/04/1933	CA BOT	Orophary	SCC	poorly df 4a	2	0	0	0	IVA	#####	#####	2641	7.4185	83.022	#####	86.767	Alive			88.033	Complete		
27	HN5CC-01-00 Female	54	27/12/1943	CA oral c	Oral cavi	SCC	moderate	4	1	0	0	IVA	#####	#####	605	1.6834	20.393	#####	13.867	Dead	HN Can	#####	20.167	Residual		
28	HN5CC-01-00 Male	53	15/07/1945	CA pyrifo	Hypophary	SCC	moderate	1	3	0	0	IVB	#####	#####	1043	2.8575	34.29	#####	34.29	Dead	HN Can	#####	10.381	Locoreg		
29	HN5CC-01-00 Male	59	04/07/1939	CA BOT	Orophary	SCC	moderate	1	1	0	0	III	#####	#####	2820	7.9213	95.056	#####	92.633	Alive				Complete		
30	HN5CC-01-00 Male	63	27/05/1935	CA BOT	Orophary	SCC	well diff.	3	2c	0	0	IVA	#####	#####	2970	8.3427	100.11	#####	97.567	Dead	Other Ca	#####	93	Complete		
31	HN5CC-01-00 Male	42	10/05/1956	CA BOT	Orophary	SCC	moderate	1	2b	0	0	IVA	#####	#####	3045	8.3425	100.11	#####	100.11	Alive			100.11	Complete		
32	HN5CC-01-00 Female	70	09/02/1928	CA tonsil	Orophary	SCC	poorly df	2	2a	0	0	IVA	#####	#####	2541	1.1376	85.652	#####	83.467	Alive			84.7	Complete		
33	HN5CC-01-00 Female	44	01/06/1954	CA poste	Hypophary	SCC	poorly df 4a	2	0	0	0	IVA	#####	#####	2799	7.8624	94.348	#####	91.333	Alive			93.3	Complete		
34	HN5CC-01-00 Male	48	01/08/1950	CA retron	Orophary	SCC	poorly df 4a	2c	0	0	0	IVA	#####	#####	3110	8.7336	104.83	#####	102.17	Alive			103.67	Complete		
35	HN5CC-01-00 Male	48	25/02/1950	CA BOT	Orophary	SCC	poorly df 4a	2c	0	0	0	IVA	#####	#####	2443	7.1039	25.247	#####	24.6	Dead	HN Can	#####	10.393	Locoreg		
36	HN5CC-01-00 Male	66	20/01/1932	CA bucco	Oral cavi	SCC	poorly df	1	2b	0	0	IVA	#####	#####	313	0.8732	10.551	#####	10.267	Dead	HN Can	#####	5.7	Distant m		
37	HN5CC-01-00 Male	52	21/11/1945	CA larynx	Glotis	SCC	poorly df 4a	2c	0	0	0	IVA	#####	#####	2399	8.4242	101.09	#####	98.5	Alive			27.2	Local rec		
38	HN5CC-01-00 Male	72	18/08/1929	CA oroph	Orophary	SCC	poorly df 4a	2b	0	0	0	IVA	#####	#####	1832	5.1611	61.753	#####	60.167	Alive			61.067	Complete		
39	HN5CC-01-00 Male	64	30/11/1934	CA supra Glotis	SCC	well diff.	3	0	0	0	0	III	#####	#####	2425	6.8186	81.742	#####	73.667	Dead	Other ca	#####	80.833	Complete		
40	HN5CC-01-00 Male	76	12/03/1922	CA supra Glotis	SCC	poorly df	3	0	0	0	0	III	#####	#####	2436	6.8427	82.12	#####	80	Dead	Unknown	#####	81.2	Complete		
41	HN5CC-01-00 Male	50	18/10/1948	CA pyrifo	Hypophary	SCC	moderate	3	0	0	0	III	#####	#####	1000	0.9839	37.073	#####	36.133	Dead	Other ca	#####	36.667	Complete		
42	HN5CC-01-00 Female	59	24/03/1939	CA supra Glotis	SCC	moderate	3	0	0	0	0	III	#####	#####	2419	6.7949	81.533	#####	79.467	Dead	Unknown	#####	80.633	Complete		
43	HN5CC-01-00 Male	57	11/01/1941	CA BOT	Orophary	SCC	moderate	3	2b	0	0	IVA	#####	#####	685	1.6888	22.416	#####	21.833	Dead	HN Can	#####	21.4667	Distant m		
44	HN5CC-01-00 Female	54	20/04/1944	CA tonsil	Orophary	SCC	moderate	2	0	0	0	II	#####	#####	2782	7.8146	93.775	#####	91.4	Alive			92.733	Complete		
45	HN5CC-01-00 Female	35	23/01/1963	CA tonsil	Orophary	SCC	moderate	2	2b	0	0	IVA	#####	#####	2358	6.6236	73.463	#####	77.467	Alive			78.6	Complete		
46	HN5CC-01-00 Male	66	24/02/1932	CA tonsil	Orophary	SCC	poorly df	3	2c	0	0	0	IVA	#####	#####	1767	3.9635	93.562	#####	58.033	Dead	Non can	#####	58.9	Complete	
47	HN5CC-01-00 Male	60	06/04/1933	CA tonsil	Orophary	SCC	well diff.	3	2c	0	0	0	IVA	#####	#####	2225	7.2233	75.433	#####	75.433	Dead	Unknown	#####	66	Local rec	
48	HN5CC-01-00 Male	57	15/04/1941	CA tonsil	Orophary	SCC	poorly df	2	2b	0	0	0	IVA	#####	#####	2793	7.9455	94.346	#####	91.733	Alive			93.1	Complete	
49	HN5CC-01-00 Male	77	30/06/1921	CA BOT	Orophary	SCC	poorly df	2	2b	0	0	0	IVA	#####	#####	2611	6.1343	88.01	#####	85.767	Alive			87.033	Complete	
50	HN5CC-01-00 Male	62	25/01/1936	CA BOT	Orophary	SCC	moderate	2	2a	0	0	0	IVA	#####	#####	2378	6.6738	80.157	#####	78.1	Alive			79.267	Complete	
51	HN5CC-01-00 Male	51	07/06/1947	CA tonsil	Orophary	SCC	poorly df	3	2b	0	0	0	IVA	#####	#####	2640	7.157	89.989	#####	86.733	Alive			88	Complete	
52	HN5CC-01-00 Male	43	07/10/1955	CA tonsil	Orophary	SCC	well diff.	2	2b	0	0	0	IVA	#####	#####	2389	6.4707	80.528	#####	78.467	Alive			79.633	Complete	
53	HN5CC-01-00 Male	54	08/03/1944	CA tonsil	Orophary	SCC	moderate	2	2b	0	0	0	IVA	#####	#####	2715	7.6264	91.517	#####	89.167	Alive			90.5	Complete	
54	HN5CC-01-00 Male	50	05/02/1949	CA BOT	Orophary	SCC	poorly df	1	3	0	0	IVB	#####	#####	2682	7.4775	89.73	#####	87.433	Alive			88.733	Complete		

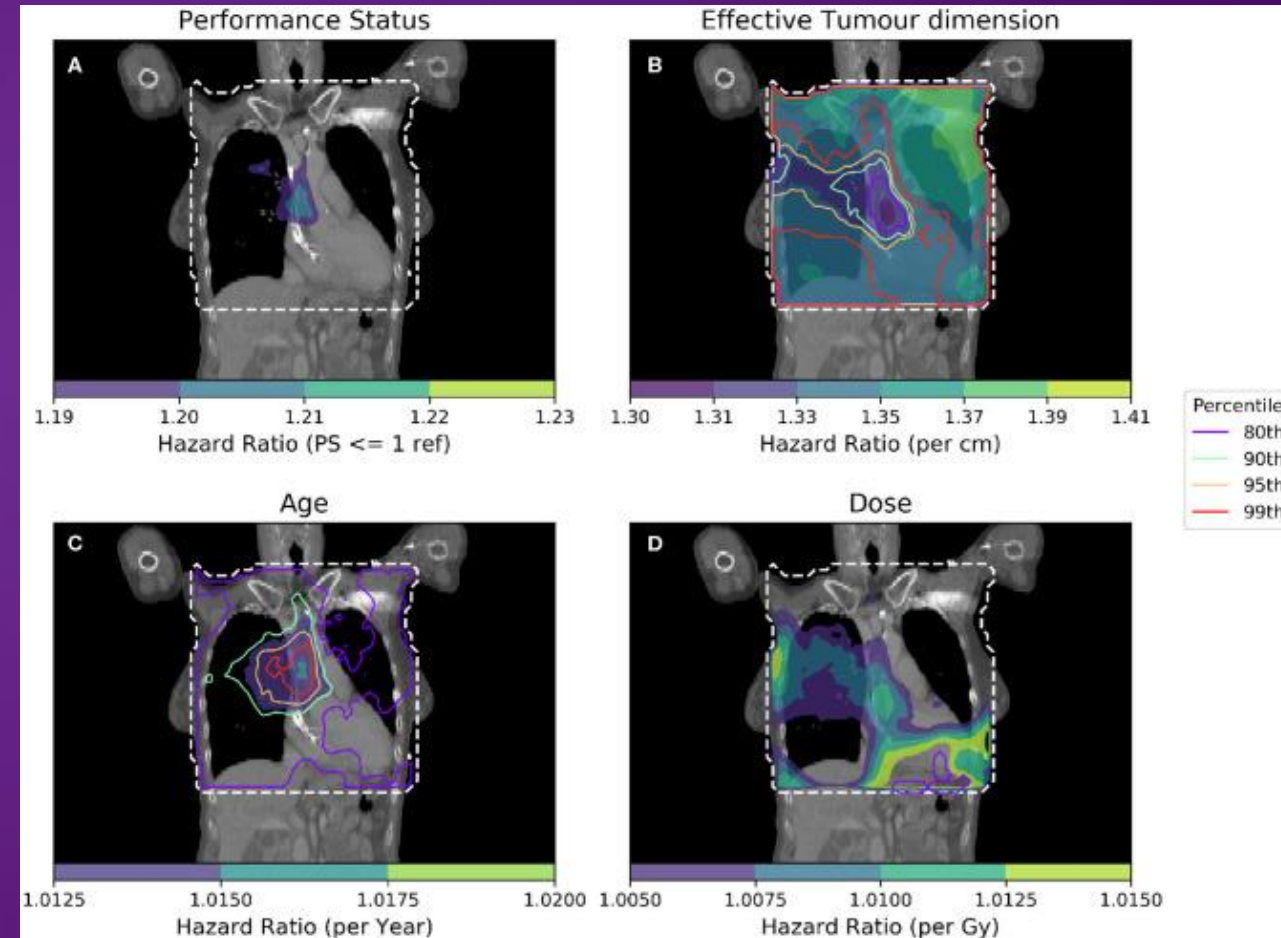
# Spatial Normalisation

- Two stages
  - Rigid/Affine registration
  - Non-rigid registration
- Very important to assess accuracy
  - Use delineations/point identifications
  - Blur dose distribution with local uncertainty



# Statistical Analysis

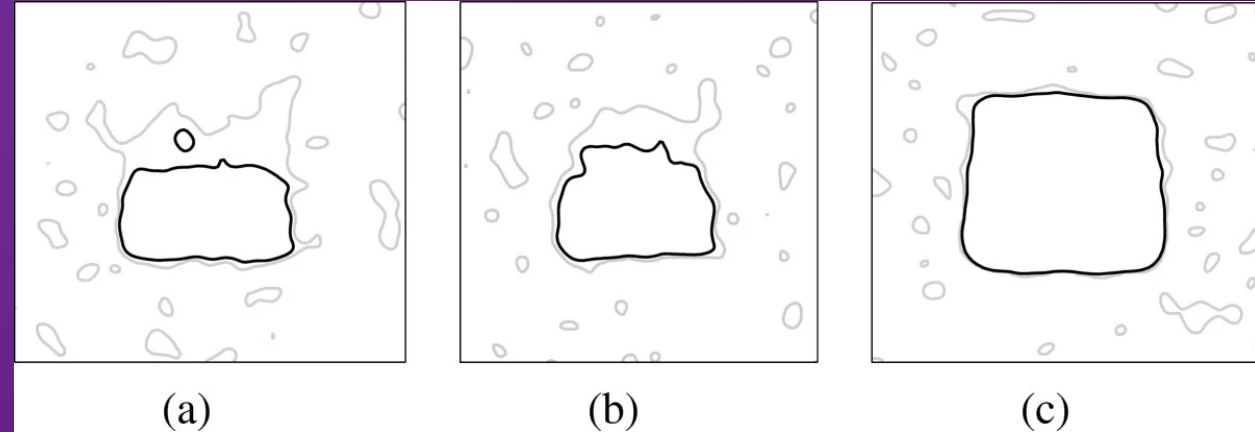
- Increasing complexity:
  - t-test (binary outcome)
  - Spearman/pearson correlation (continuous outcome)
  - Logistic regression (binary outcome)
  - Cox regression (time-to-event)
- Voxels treated independently
  - Huge multiple comparisons problem



From <https://dx.doi.org/10.3389/fonc.2020.01178>

## Multiple comparisons

- Test enough times and you're bound to find an effect
- Classical corrections:
  - Bonferroni
  - FDR
- Easy, nonparametric method:
  - Permutation testing

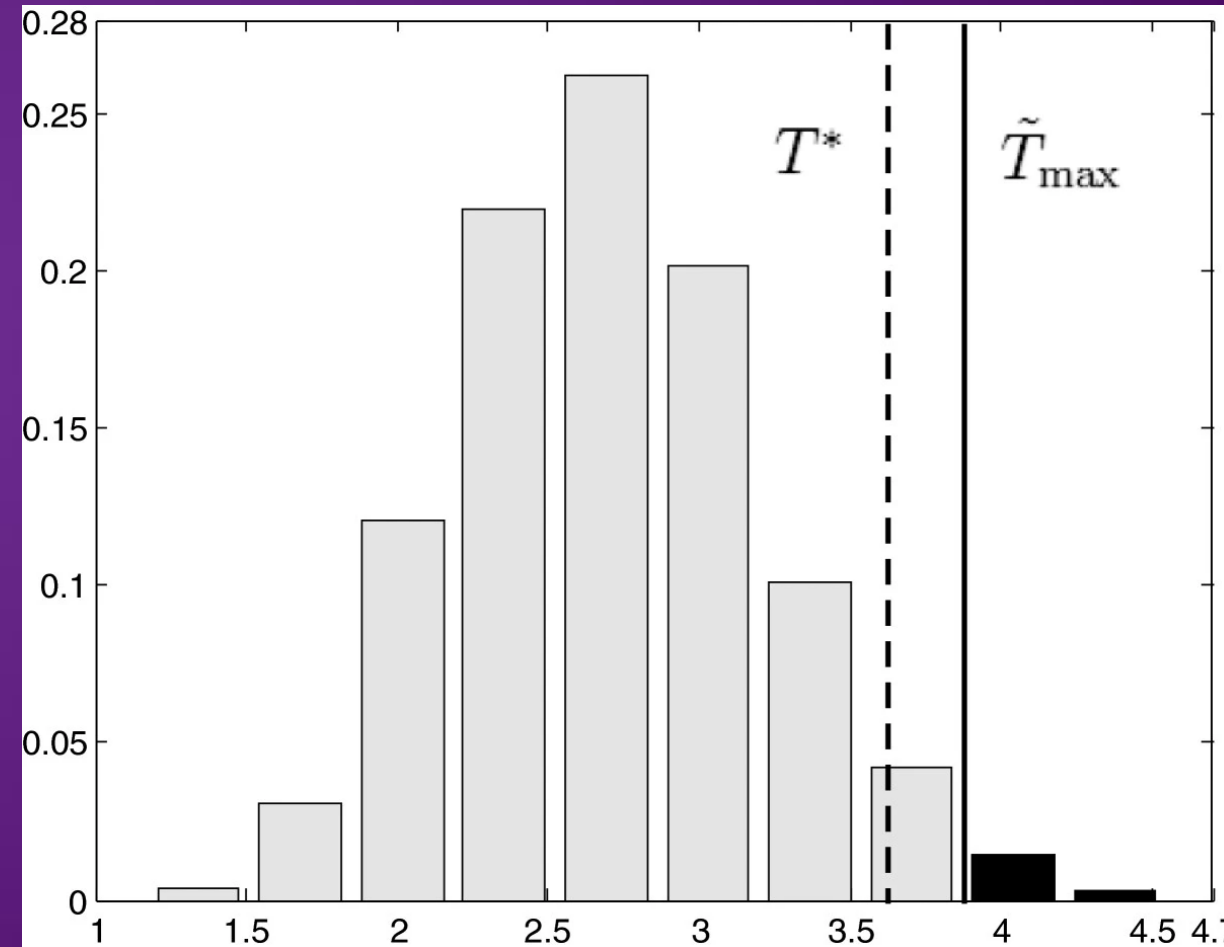


From <https://doi.org/10.1186/1748-717X-8-293>



# Permutation testing

- Null hypothesis: There is no relationship between outcome and dose
  - In which case, it shouldn't matter if we mess up the outcome-dose correspondence
- Test significance by recalculating stats in messed up correspondences
  - Threshold at e.g. 95<sup>th</sup> percentile of null distribution



From <https://doi.org/10.1186/1748-717X-8-293>

# Putting it all together

- The three notebooks do most of the steps here
  - Skipped a few things (time)
  - We will start from binary IBDM to look at the steps
- I'm ignoring what we might do next
  - Finding significant region and building a model
  - See e.g. Alan McWilliam's slides

**TO COLAB!**

## Link to colab notebook

- <https://github.com/rrr-uom-projects/IBDMWorkshop>
- To open in colab, select the github tab, then cancel the authorisation and search for rrr-uom-projects
- Select the IBDMWorkshop repo