Strategies for Investigating Etiologic Heterogeneity of Cancer

Colin Begg
(and others)

Memorial Sloan-Kettering Cancer Center
New York

Pathological Society of Great Britain and Ireland
Maastricht
June 2018

Incidence of Second Primary Cancers

- Standardized incidence ratio
  - Of cancer type A following cancer type B
  - Numerator
    - Rate of cancer A among people with an existing cancer B
  - Denominator
    - Population rate of cancer type A
- Can be expressed as an O/E ratio
  - If this is close to 1 then we can conclude that the cancers do not share strong risk factors
  - If O/E is high then the cancers do share a common etiology
- Key Point
  - Double primaries tell us a lot about shared etiology

A Selection of Some Key Causes of Cancer

Lung Cancer	Smoking
Breast Cancer	Hormonal events in young women / BRCA1/2 mutations
Colon Cancer	APC mutations / MSI / dietary factors
Mesothelioma	Asbestos
Liver Cancer	Hepatitis B and C
Cervix Cancer	Human papilloma virus
Melanoma	CDKN2A mutations / sun exposure

Data from SEER – Neugut 1999

1st Cancer	2nd Cancer	Observed	Expected	SIR
lung	colorectal	620	592	1.1
colorectal	lung	1948	1790	1.1
breast	colorectal	2253	2202	1.0
colorectal	breast	1461	1249	1.2
lung	head/neck	436	126	3.5
head/neck	lung	1891	432	4.4

Conclusions
- No evidence that lung and colorectal risks are related
- No evidence that breast and colorectal risks are related
- Strong evidence that lung and head/neck risks are related

Key Insight

- If two diseases have similar etiologies they will tend to co-occur in the same patients
- Implication – if someone has disease A the person is more likely to subsequently develop disease B than the person would otherwise
- Technical implication – the standardized incidence ratio will be elevated
- Double primaries tell us something fundamental about etiology without any knowledge of individual risk factors

What is Etiologic Heterogeneity?

The concept that individual site-specific cancers are in fact collections of diseases with distinctive risk profiles

A Thought Experiment

For about 60 years cancer epidemiological studies have been conducted on a site-specific basis

How would research have progressed if cancer had been evaluated as if it were a unitary disease?

Breast Cancer Sub-Types

[from Sorlie, Perou et al. PNAS 2001]

- Luminal A
  - Approximated by ER+ or PR+ and HER2-
- Luminal B
  - Approximated by ER+ or PR+ and HER2+
- HER2 – enhanced
  - Approximated by ER-, PR-, HER2+
- Triple Negative
  - Approximated by ER-, PR-, HER2-
- Are these sub-types etiologically heterogeneous?
- Do they represent the best way to sub-type breast cancer from the perspective of etiology?

Etiologic Heterogeneity

- Breast cancer
  - Luminal A / Luminal B / HER2-enhanced / Triple negative
- If the sub-types have different etiologies then different sub-types will be less likely to co-occur in the same patient
- For example, a person at high risk of triple negative breast cancer is more likely to have a TN breast cancer
  - This is true also for a second primary
- Implication
  - Among people with double primaries (bi-lateral breast cancer), the individual sub-types will tend to be matched to a greater degree than expected
  - The stronger the association – the more distinctive the etiologies

Results Suggest Modified Taxonomy

- Aggregate Luminal A and B – 3 sub-types defined by ER and Her2

	ER+	ER- / HER2 +	ER- / HER2-
ER+	284 / 251	20	37
ER- / HER2 +	16	17 / 4	9
ER- / HER2 -	26	5	28 / 10

Observed / Expected

329 of 442 cases (74%) are concordant – versus 265 (60%) expected
Odds Ratios – 15.1 8.3 10.6

Data from California Cancer Registry 1999-2004

Limited to Metachronous Cases

Bilateral Breast Cancers					
		Second Cancer			
		Luminal A	Luminal B	HER2 Enhanced	Triple Negative
First Cancer	Luminal A	208	28	15	32
	Luminal B	40	8	5	5
	HER2 Enhanced	12	4	17	9
	Triple Negative	23	3	5	28

Odds Ratios			
	Luminal A	Luminal B	HER2 Enhanced
Luminal B	1.5		
HER2 Enhanced	19.6	6.8	
Triple Negative	7.9	14.9	10.6

Looks like Luminal A and B can be combined without loss of information.

Etiologic Concordance

- Reflects the concept that sub-types of different cancers might have similar etiologies
- If so, among double primaries of cancers of types A and B, etiologically concordant sub-types should occur together more often than expected
- Example – breast and ovarian cancers
- We already know that breast-ovary double primaries occur more often than expected
  - But what about sub-types of these cancers?

Basal-Like Breast Cancer
High Grade Serous Ovarian Cancer

- These sub-types share similar somatic features
  - Very high frequency of *TP53* mutations
  - Recurrent somatic events in *RB1*
  - Mutually exclusive mutations in similar pathway
  - Similar copy number changes
  - Similar expression patterns
- TCGA report on breast cancer concluded from this that these results indicate a "related aetiology". [Nature 2012]
- To test this theory empirically we can examine co-occurrence of these sub-types in breast-ovarian double primaries
- Note: Basal-like tumors have tumors that look similar to basal cells surrounding the mammary duct – positive for cytokeratin 5/6 or EGFR

Summary

- Double primaries are uniquely informative about etiology
  - To distinguish sub-types with different etiologies
  - To identify sub-types (of different cancers) with similar etiologies
- With more data and more elaborate tumor profiling
  - We can search the space of sub-types to discover how best to sub-type tumors
- I will use melanoma data to illustrate the ideas

California Registry – 1999-2013 [Begg BJC 2017]

Breast Sub-Type	Ovarian Sub-Type	Double Primary Cases			95% Confidence
		Observed	Expected	Ratio	
Luminal A	Serous	236	245.1	0.96	0.84-1.09
	Endometrioid	50	64.3	0.78	0.56-0.99
	Clear Cell	27	31.4	0.86	0.54-1.18
	Mucinous	26	37.1	0.70	0.43-0.97
Luminal B	Serous	32	43.1	0.74	0.49-1.00
	Endometrioid	9	11.3	—	—
	Clear Cell	7	5.5	—	—
	Mucinous	7	6.5	—	—
HER2 Enhanced	Serous	25	21.2	1.18	0.72-1.64
	Endometrioid	2	5.6	—	—
	Clear Cell	2	2.7	—	—
	Mucinous	1	3.2	—	—
Triple Negative	Serous	103	43.8	2.35	1.90-2.81
	Endometrioid	13	11.5	1.13	0.52-1.74
	Clear Cell	4	5.6	—	—
	Mucinous	1	6.6	—	—

Investigation of Melanoma Sub-Types
[Mauguen et al. JASA 2017]

- Case-control study of double primary melanoma
- 903 evaluable cases with two independent invasive melanomas
  - Total of 1806 individual melanomas
- Tumor characteristics analyzed (selected from a total of 12)
  - Anatomic site of melanoma – head/neck vs trunk/arms vs legs
  - Histology – lentigo maligna vs superficial spreading
  - Neval remnants – present vs absent
  - Solar elastosis – present vs absent
  - Tumor infiltrating lymphocytes – present versus absent
  - Mutation status – *BRAF* vs *NRAS* vs Wild Type
- Multiple imputation used to account for missing data

Details

- 144 unique combinations of the 6 tumor factors
  - 2.1×10^{85} ways to arrange 144 categories into 4 classes
  - We can't examine them all!
- Focus on the 4 sub-type option
  - Minimal increase in heterogeneity by using 5 sub-types
- Approach
  - k-means clustering to find candidate sub-types
  - Calculate the heterogeneity measure for each
  - Aggregate solutions that have "similar" sets of sub-types
- Note
  - Clustering is accomplished using all 2n tumors
  - Heterogeneity measure calculated based on the pairings

Case-Only Logistic Regression [Based on 2103 Single Primary Melanomas]

Factors		Comparisons with Reference Sub-Type A* (Legs) Odds Ratio (95% Confidence)			P-Value
		Sub-Type B* Trunk/Arms SE	Sub-Type C* Trunk/Arms no SE	Sub-Type D* Head/Neck	
Age	50-70 vs <50	1.4 (1.0-1.9)	0.8 (0.5-1.1)	1.6 (1.0-2.4)	$<10^{-4}$
	>=70 vs <50	1.9 (1.3-2.8)	0.9 (0.5-1.4)	4.1 (2.4-6.7)	
Gender	Female	0.2 (0.2-0.3)	0.2 (0.1-0.2)	0.2 (0.1-0.3)	$<10^{-4}$
Center	Rest vs Australia	1.7 (1.2-2.2)	1.4 (1.0-1.9)	1.4 (1.0-2.1)	0.01
Mole Count	5-10 vs 0-4	1.1 (0.8-1.5)	1.2 (0.8-1.8)	0.7 (0.5-1.2)	0.005
	11-25 vs 0-4	1.2 (0.8-1.6)	1.8 (1.2-2.6)	0.8 (0.5-1.2)	
	>25 vs 0-4	1.4 (0.9-2.0)	2.0 (1.3-3.1)	1.2 (0.7-2.0)	
Eye Color	Light	1.5 (1.1-2.1)	1.2 (0.9-1.8)	1.5 (1.0-2.4)	0.02
Hair Color	Light vs Dark	0.9 (0.7-1.2)	1.0 (0.7-1.4)	0.8 (0.5-1.2)	0.34
	Red vs Dark	1.0 (0.6-1.6)	0.8 (0.4-1.4)	0.9 (0.4-1.7)	
Sun Exposure (UVE dose)	2^{nd} quart. vs 1^{st}	1.5 (1.0-2.1)	0.7 (0.5-1.1)	1.1 (0.7-1.8)	$<10^{-4}$
	3^{rd} quart. vs 1^{st}	1.2 (0.8-1.8)	0.6 (0.4-0.9)	1.4 (0.8-2.4)	
	4^{th} quart. vs 1^{st}	1.9 (1.2-3.1)	0.5 (0.3-0.9)	1.2 (0.7-2.3)	
Family Hx.	Yes vs No	1.3 (0.9-1.9)	1.5 (0.9-2.3)	1.2 (0.7-2.1)	0.79

Optimal Solutions

Solution #	Heterogeneity Measure	Approximate Description
1	0.58	A: Legs B: Trunk/Arms + Solar Elastosis C: Trunk/Arms + no Solar Elastosis D: Head/Neck
2	0.56	A: Legs B: Trunk/Arms + Neval Remnants C: Trunk/Arms + no Neval Remnants D: Head/Neck
3	0.48	A: Legs B: Trunk/Arms + Wild Type C: Trunk/Arms + *BRAF/NRAS* D: Head/Neck

Whiteman et al. – divergent pathway hypothesis – JNCI 2003

Why is This Knowledge Relevant?

- *IRF4* (T allele) Genotype
  - The interferon regulatory factor-4 gene (*IRF4*) has been hypothesized to be involved in melanocyte differentiation and proliferation
- *IRF4* rs1223592 has been identified as a low penetrance melanoma risk factor
- But it shows striking differences in risk for different sub-types
  - Body site – Kvaskov et al. Twin Res Hum Genet 2011
  - Presence of solar elastosis versus nevus remnants – Gibbs et al. J Inv Derm 2018
- Heterogeneity of this nature can lead to greater statistical power for discovery of disease-related SNPs – Begg/Zabor Amer J Epidemiol 2012

Data from GEM Study Single Primary Melanomas Based on 2103 Cases

	Comparisons with Reference Sub-Type A* (Legs) Odds Ratio (95% Confidence)		
IRF4 rs12203592*T	Sub-Type B* Trunk/Arms SE	Sub-Type C* Trunk/Arms no SE	Sub-Type D* Head/Neck
0	1.0	1.0	1.0
1	1.3 (1.0-1.6)	0.8 (0.6-1.1)	1.6 (1.2-2.2)
2	1.9 (1.1-3.4)	0.8 (0.4-1.7)	3.8 (2.1-7.1)

Some Technical Limitations / Assumptions

- Logistical
  - Studies of double primaries are hard to do
  - Obtaining tumor tissue from both primaries is challenging
- Both primaries influenced by exactly the same risk milieu
  - Impact of treatment for first primary
  - Impact of age
- Tumors must be biologically independent
  - 2nd primary must not be a mis-diagnosed metastasis
  - Both etiologic heterogeneity and clonality will tend to make somatic profiles of tumors more alike

Messages

- Double primaries are uniquely informative about etiology
- Challenges are to assemble large numbers with tissue specimens from both tumors to allow
  - Extensive genomic profiling
  - Discovery studies to identify and define the sub-types
- Knowledge of etiologically distinct sub-types permits
  - Improved search for genetic risk factors
  - Improved risk prediction
- Knowledge of etiologically concordant sub-types
  - Refined pathologic taxonomy of cancers

Acknowledgements

- GEM Study Team, especially:-
  - Marianne Berwick (University of New Mexico)
  - Nancy Thomas (University of North Carolina)
- MSKCC
  - Audrey Mauguen
  - Venkat Seshan
  - Emily Zabor
- Harvard
  - Shelley Tworoger (Moffitt Cancer Center)