Supplementary appendix

Benefits and limitations of using individual and different combinations of linked English routine data sources in cancer epidemiology studies

Table 1: Number of patients identified with each cancer site using the gold standard algorithm

Cancer site	Number of patients
Oral Cavity (C00-06)	2097
Oesophageal (C15)	5212
Stomach (C16)	4016
Colorectal (C18-C20)*	22173
Liver (C22)	2230
Pancreas (C25)	5008
Lung (C34)	21978
Malignant melanoma (C43)	7282
Breast (C50)	29297
Cervix (C53)	1503
Uterus (C54-55)	4325
Ovaries (C56)	4157
Prostate (C61)	24888
Kidney (C64)	4086
Bladder (C67)	8871
Brain/CNS (C71-72)	2921
Thyroid (C73)	1314
NHL (C82-85)	6644
Multiple myeloma (C90)	2672
Leukemia (C91-95)	5279
Total	165953

Figure 1: Positive Predictive Value by age

	Oral Cavity (C00-06)	Oesophageal (C15)	Stomach (C16)	Colorectal (C18-C20)*
NCRAS -	•	A	A	^
NCRAS & HES APC -	۸ ۵	A	<u>// / / / / / / / / / / / / / / / / / /</u>	•
ALL SOURCES -	*			▲
CPRD GOLD, HES APC & ONS MORTALITY-	▲ /Δ		۵.	A
CPRD GOLD -	A	A		A
HES APC -		A	A M	_
	Liver (C22)	Pancreas (C25)	Lung (C34)*	Malignant melanoma (C43)
NCRAS -	•	A	A	
NCRAS & HES APC -	A	28	^	A44
ALL SOURCES -		24		*
CPRD GOLD, HES APC & ONS MORTALITY-		<u>/*</u>	A	A A
CPRD GOLD -		A	A	•
HES APC -	•			
	Breast (C50)*	Cervix (C53)	Uterus (C54-55)	Ovaries (C56)
NCRAS -	•	A		
NCRAS & HES APC -	•	A 14		A
ALL SOURCES -	A	▲ ∆		/ ▲
CPRD GOLD, HES APC & ONS MORTALITY-	A	<u>۸</u>		A
CPRD GOLD -	•			
HES APC -				
	Prostate (C61)*	Kidney (C64)	Bladder (C67)	Brain/CNS (C71-72)
NCRAS -	•	A	A	A
NCRAS & HES APC -	**	A	A#	A
ALL SOURCES -	A A		∆ ▲	A
CPRD GOLD, HES APC & ONS MORTALITY-	A A	A		
CPRD GOLD -	*	A.	A A	A
HES APC -	A A	▲ A	/	AA
	Thyroid (C73)	Non-Hodgkin lymphoma (C82-85) Multiple myeloma (C90)	Leukemia (C91-95)
NCRAS -		▲	•	A
NCRAS & HES APC -	*	A	*	A
ALL SOURCES -			A A	A
CPRD GOLD, HES APC & ONS MORTALITY			A	A A
CPRD GOLD -	4	/A		▲ △
HES APC -	A A	^	A A	
	0 20 40 60 80 100	0 20 40 60 80 100	0 20 40 60 80 100	0 20 40 60 80 100
	Δ <=59	🔺 60-79 🔺 a	80-115	

Figure 2: Positive Predictive Value by sex

	Oral Cavity (C00-06)	Oesophageal (C15)	Stomach (C16)	Colorectal (C18-C20)*
NCRAS -	•	A		^
NCRAS & HES APC -				
ALL SOURCES -	A		A	
CPRD GOLD, HES APC & ONS MORTALITY -	*	A	▲	_
CPRD GOLD -	A	A		A
	••			
HES APC -		-	-	-
NCRAS -	Liver (C22)	Pancreas (C25)	Lung (C34)*	Malignant melanoma (C43
NCRAS & HES APC -	••			
	-			-
ALL SOURCES -	*			A
CPRD GOLD, HES APC & ONS MORTALITY-	▲ ∆	▲	A	A
CPRD GOLD -	*	A	A	
HES APC -	▲ ∆	A	A	
	Breast (C50)*	Cervix (C53)	Uterus (C54-55)	Ovaries (C56)
NCRAS -		A	A	
NCRAS & HES APC -		A	A	
ALL SOURCES -	•	A		
CPRD GOLD, HES APC & ONS MORTALITY-	A	A	▲	A
CPRD GOLD -			•	
	-	_	_	-
HES APC -	A	A	•	A
NCRAS -	Prostate (C61)*	Kidney (C64)	Bladder (C67)	Brain/CNS (C71-72)
	4	-	-	-
NCRAS & HES APC -	Δ			*
ALL SOURCES -	Δ	A	A	▲ \
CPRD GOLD, HES APC & ONS MORTALITY-	Δ		A	
CPRD GOLD -	Δ		A	•
HES APC -	۵			
	Thyroid (C73)	Non-Hodgkin lymphoma (C82-85	5) Multiple myelome (C90)	Leukemia (C91-95)
NCRAS -				
NCRAS & HES APC -	A	A		
ALL SOURCES -	*	A	A	A
CPRD GOLD, HES APC & ONS MORTALITY-				
	A	-	-	-
CPRD GOLD -	•	A	A	▲
HES APC -				
	0 20 40 60 80 100	0 20 40 60 80 100	0 20 40 60 80 100	0 20 40 60 80 100

Figure 3: Positive Predictive Value by calendar year

	Oral Cavity (C00-06)	Oesophageal (C15)	Stomach (C16)	Colorectal (C18-C20)*
NCRAS -	•	•	•	•
NCRAS & HES APC -	A	A		^
ALL SOURCES -	1		^	A
CPRD GOLD, HES APC & ONS MORTALITY-	2 *	A		A
CPRD GOLD -	۵.	*	A	A
HES APC -		A		A
	Liver (C22)	Pancreas (C25)	Lung (C34)*	Malignant melanoma (C43)
NCRAS -	•		^	`
NCRAS & HES APC -	A	A	A	
ALL SOURCES -	A	▲	A	A
CPRD GOLD, HES APC & ONS MORTALITY-	▲	A		
CPRD GOLD -	<u>^*</u>	A	A	A
HES APC -	∕≛.			
	Breast (C50)*	Cervix (C53)	Uterus (C54-55)	Ovaries (C56)
NCRAS -		(CCC)	▲	
NCRAS & HES APC -	•	A		A
ALL SOURCES -		<u>^</u>		A
CPRD GOLD, HES APC & ONS MORTALITY-	4	4	2	
CPRD GOLD -				
HES APC -			A.	
HES ARG	_	_	_	_
NCRAS -	Prostate (C61)* ▲	Kidney (C64)	Bladder (C67)	Brain/CNS (C71-72)
NCRAS & HES APC -	A			A
ALL SOURCES -	4			
CPRD GOLD, HES APC & ONS MORTALITY-				
	_	_	_	_
CPRD GOLD -	2	24	A	
HES APC -	•	A		
NCRAS -	Thyroid (C73) No	on-Hodgkin lymphoma (C82-85) Multiple myeloma (C90)	Leukemia (C91-95)
NCRAS & HES APC -				_
	-	-	-	
ALL SOURCES -	A			
CPRD GOLD, HES APC & ONS MORTALITY -	<u>^</u>			A A
CPRD GOLD -	A	<u>^</u>	<u>م</u>	▲
HES APC -	· · · · · · · · ·	_		
	0 20 40 60 80 100	0 20 40 60 80 100	0 20 40 60 80 100	0 20 40 60 80 100
Δ	2000-2004	2005-2009	2010-2014	

Figure 4: Sensitivity by age

NCRAS -	Oral Cavity (C00-06)	Oesophageal (C15)	Stomach (C16)	Colorectal (C18-C20)*
NCRAS & HES APC -				^
ALL SOURCES -				
	_			_
CPRD GOLD, HES APC & ONS MORTALITY-	A # A	<u>A</u>		•
CPRD GOLD -	Alb.			
HES APC -		_	A	
NCRAS -	Liver (C22)	Pancreas (C25)	Lung (C34)* ▲	Malignant melanoma (C43
NCRAS & HES APC -		A	A	
ALL SOURCES -				-
CPRD GOLD, HES APC & ONS MORTALITY-	*	ZA	A	-
CPRD GOLD -				
HES APC -	A 45			A
	Breast (C50)*	Cervix (C53)	Uterus (C54-55)	Ovaries (C56)
NCRAS -				*
NCRAS & HES APC -	•	A14	A	A14
ALL SOURCES -			A	
CPRD GOLD, HES APC & ONS MORTALITY-			A	A
CPRD GOLD -				
HES APC -		**		
	Prostate (C61)*	Kidney (C64)	Bladder (C67)	Brain/CNS (C71-72)
NCRAS -				
NCRAS & HES APC -		AM	A	
ALL SOURCES -				
CPRD GOLD, HES APC & ONS MORTALITY-				A
CPRD GOLD -			A	
HES APC -				
	-		_	
NCRAS -	Thyroid (C73) N	on-Hodgkin lymphoma (C82-85 🔺) Multiple myeloma (C90)	Leukemia (C91-95)
NCRAS & HES APC -		A1.	A	
ALL SOURCES -	•			A
CPRD GOLD, HES APC & ONS MORTALITY-		A		
			_	-
CPRD GOLD -	• •			
HES APC -	0 20 40 60 80 100		0 20 40 60 80 100	0 20 40 60 80 100

Figure 5: Sensitivity by sex

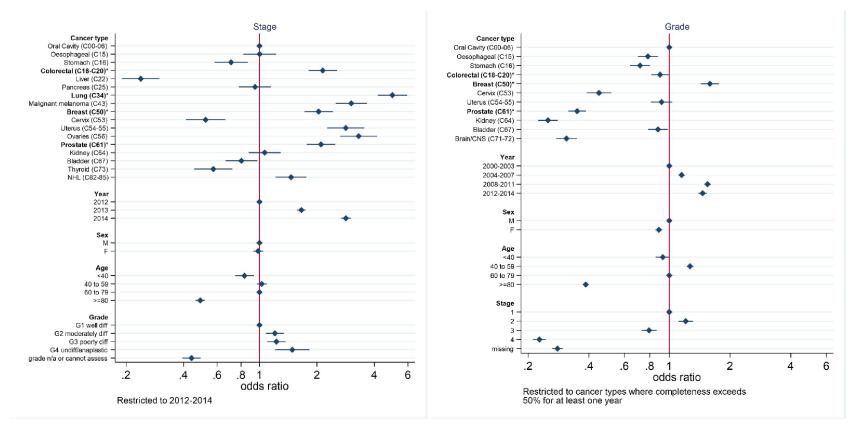
	Oral Cavity (C00-06)	Oesophageal (C15)	Stomach (C16)	Colorectal (C18-C20)*
NCRAS -	A	A		
NCRAS & HES APC -	A	^▲	A	
ALL SOURCES -		^		A
CPRD GOLD, HES APC & ONS MORTALITY-			A	_
CPRD GOLD -			*	*
HES APC -		_		
	_			
NCRAS -	Liver (C22)	Pancreas (C25)	Lung (C34)*	Malignant melanoma (C43)
NCRAS & HES APC -	A A		^	A
ALL SOURCES -	*	A	A	A
CPRD GOLD, HES APC & ONS MORTALITY-				
				_
CPRD GOLD -		-	.	_
HES APC -	*	A	^	
NCRAS -	Breast (C50)*	Cervix (C53)	Uterus (C54-55)	Ovaries (C56)
NCRAS & HES APC -				
			-	_
ALL SOURCES -				
CPRD GOLD, HES APC & ONS MORTALITY-		▲	A	
CPRD GOLD -	A	A		
HES APC -	A		A	A
	Prostate (C61)*	Kidney (C64)	Bladder (C67)	Brain/CNS (C71-72)
NCRAS -	۵	•	A	
NCRAS & HES APC -	Δ	A	A	A
ALL SOURCES -	Δ	4	•	_
CPRD GOLD, HES APC & ONS MORTALITY-	Δ	<u>^</u>	A	
CPRD GOLD -	Δ	A	A	▲
HES APC -	Δ	A	A	A
	Thyroid (C73) N	on-Hodgkin lymphoma (C82-85	 Multiple myeloma (C90) 	Leukemia (C91-95)
NCRAS -	A		A	
NCRAS & HES APC -		A	A	▲
ALL SOURCES -	A	4	A	A
CPRD GOLD, HES APC & ONS MORTALITY-		*		
CPRD GOLD -		A	A	A
HES APC -	A			A
	0 20 40 60 80 100	0 20 40 60 80 100	0 20 40 60 80 100	0 20 40 60 80 100
	▲ fe	male 🛆 male		

Figure 6: Sensitivity by calendar year

NCRAS -	Oral Cavity (C00-06)	Oesophageal (C15)	Stomach (C16)	Colorectal (C18-C20)*
NCRAS & HES APC -	_			
		-	-	-
ALL SOURCES -	^	A	^	
CPRD GOLD, HES APC & ONS MORTALITY-	A	A	······ · · · · · · · · · · · · · · · ·	A
CPRD GOLD -	A	۵ 🛦	▲ ▲	
HES APC -	A	A	∆ ≜	A
	Liver (C22)	Pancreas (C25)	Lung (C34)*	Malignant melanoma (C43)
NCRAS -	A	A		
NCRAS & HES APC -		A	▲	A
ALL SOURCES -	*	A	A	▲
CPRD GOLD, HES APC & ONS MORTALITY-		A		/A
CPRD GOLD -	AIA	Δ 🔺	△▲	<u>کھ</u>
HES APC -	A	A	A	
	Breast (C50)*	Cervix (C53)	Uterus (C54-55)	Ovaries (C56)
NCRAS -	A	A	•	▲
NCRAS & HES APC -				A
ALL SOURCES -	_	A	A	A
CPRD GOLD, HES APC & ONS MORTALITY-	•	A	A	
CPRD GOLD -	<u>^</u>	۵ 🔺		A
HES APC -			A	
	Prostate (C61)*	Kidney (C64)	Bladder (C67)	Brain/CNS (C71-72)
NCRAS -				
NCRAS & HES APC -	•	A		
ALL SOURCES -	•	/ * ^	_ _	A
CPRD GOLD, HES APC & ONS MORTALITY-		A A		A
CPRD GOLD -	A	A	۵.	
HES APC -	A 10			
NCRAS -	Thyroid (C73) No	n-Hodgkin lymphoma (C82-85)) Multiple myeloma (C90)	Leukemia (C91-95)
NCRAS & HES APC -	•			
ALL SOURCES -	· · · · · · · · · · · · · · · · · · ·	A	_	A
CPRD GOLD, HES APC & ONS MORTALITY-				
CPRD GOLD -	۵.			
		-	-	-
HES APC -	0 20 40 60 80 100	0 20 40 60 80 100	0 20 40 60 80 100	
*				
Δ	2000-2004	2005-2009	2010-2014	

Figure 7: Output from logistic regression models with completeness of stage and grade as the dependent variables

Created using coefplot command in Stata http://repec.sowi.unibe.ch/stata/coefplot/



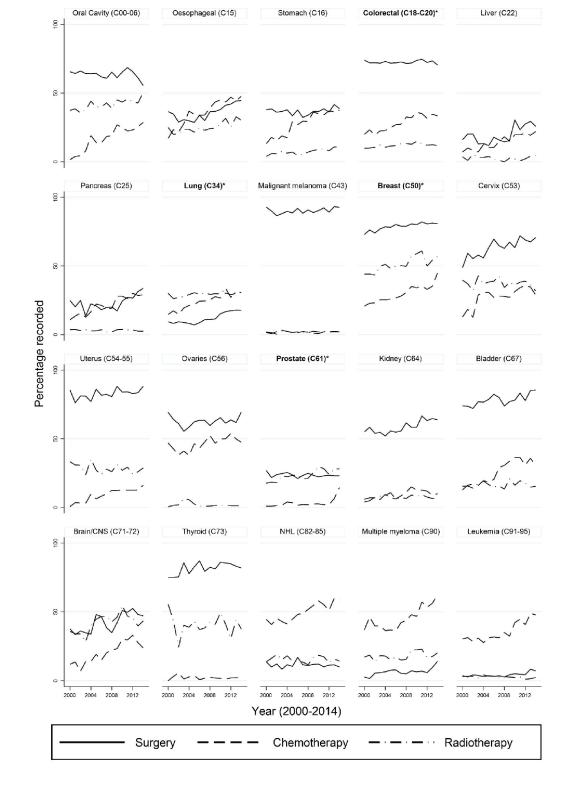


Figure 8: Recording of treatment modalities for patients identified using NCRAS data only