SUPPLEMENTAL INFORMATION

This document contains information on the distribution of 720 confirmed cancer cases identified in the United Kingdom's Clinical Practice Research Datalink (CPRD) by data source (primary, hospital, and cancer registry data) for the 10 study cancer cases combined and by cancer type. We present the same information in tabular and graphical formats. Additional information can be found in these publications:

- Parent study: Kaye JA, Margulis AV, Fortuny J, McQuay LJ, Plana E, Bartsch JL, Bui CL, Perez-Gutthann S, Arana A. Cancer Incidence after Initiation of Antimuscarinic Medications for Overactive Bladder in the United Kingdom: Evidence for Protopathic Bias. Pharmacotherapy. 2017 Jun;37(6):673-683. doi: 10.1002/phar.1932.
- Validation study: Margulis AV, Fortuny J, Kaye JA, Calingaert B, Reynolds M, Plana E, McQuay LJ, Atsma WJ, Franks B, de Vogel S, Perez-Gutthann S, Arana A. Validation of Cancer Cases Using Primary Care, Cancer Registry, and Hospitalization Data in the United Kingdom. Epidemiology. 2018 Mar;29(2):308-313. doi: 10.1097/EDE.0000000000000786.
- Assessment of value of free-text comments for cancer case validation: Margulis AV, Fortuny J, Kaye JA, Calingaert B, Reynolds M, Plana E, McQuay LJ, Atsma WJ, Franks B, de Vogel S, Perez-Gutthann S, Arana A. Value of Free-text Comments for Validating Cancer Cases Using Primary-care Data in the United Kingdom. Epidemiology. 2018 Sep;29(5):e41-e42. doi: 10.1097/EDE.00000000000000856.

LIST OF TABLES

Table 1. Table 2.	Cancer Cases by Data Source Within CPRD: Primary Care (GOLD), Hospital (HES) and Cancer Registry (NCDR) Data
LIST O	F FIGURES
Figure 1.	Cancer Cases by Data Source Within CPRD: All Cancers Combined114
Figure 2.	Cancer Cases by Data Source Within CPRD: Bladder Cancer
Figure 3.	Cancer Cases by Data Source Within CPRD: Breast Cancer
Figure 4.	Cancer Cases by Data Source Within CPRD: Colorectal Cancer 147
Figure 5.	Cancer Cases by Data Source Within CPRD: Uterine Cancer
Figure 6.	Cancer Cases by Data Source Within CPRD: Kidney Cancer 169
Figure 7.	Cancer Cases by Data Source Within CPRD: Lung and Bronchus
	Cancer <u>1740</u>
Figure 8.	Cancer Cases by Data Source Within CPRD: Non-Hodgkin Lymphoma 1841
Figure 9.	Cancer Cases by Data Source Within CPRD: Pancreatic Cancer
Figure 10.	Cancer Cases by Data Source Within CPRD: Prostate Cancer
Figure 11.	Cancer Cases by Data Source Within CPRD: Skin Melanoma

Table 1. Cancer Cases by Data Source Within CPRD: Primary Care (GOLD), Hospital (HES) and Cancer Registry (NCDR) Data

		Data Source, n (%)												
Cancer type	Total	GOLDª	HESª	NCDR ^a	GOLD only ^b	HES only ^b	NCDR only ^b	GOLD and HES only	GOLD and NCDR only	HES and NCRD only	GOLD, HES, and NCDR ^c			
All types	720	492 (68)	581 (81)	608 (84)	22 (3)	56 (8)	54 (8)	34 (5)	63 (9)	118 (16)	373 (52)			
Bladder	113	70 (62)	104 (92)	76 (67)	5 (4)	18 (16)	3 (3)	14 (12)	1 (1)	22 (19)	50 (44)			
Breast	144	130 (90)	120 (83)	131 (91)	4 (3)	3 (2)	4 (3)	6 (4)	16 (11)	7 (5)	104 (72)			
Colorectal	87	56 (64)	76 (87)	77 (89)	1 (1)	8 (9)	7 (8)	1 (1)	3 (3)	16 (18)	51 (59)			
Corpus uteri	28	16 (57)	22 (79)	27 (96)	0 (0)	1 (4)	3 (11)	0 (0)	3 (11)	8 (29)	13 (46)			
Kidney and renal pelvis	28	10 (36)	24 (86)	24 (86)	0 (0)	4 (14)	2 (7)	0 (0)	2 (7)	12 (43)	8 (29)			
Lung and bronchus	91	43 (47)	68 (75)	85 (93)	0 (0)	4 (4)	17 (19)	2 (2)	6 (7)	27 (30)	35 (38)			
Non-Hodgkin lymphoma	29	19 (66)	24 (83)	23 (79)	0 (0)	4 (14)	2 (7)	2 (7)	3 (10)	4 (14)	14 (48)			
Pancreas	21	10 (48)	15 (71)	16 (76)	1 (5)	3 (14)	4 (19)	1 (5)	1 (5)	4 (19)	7 (33)			
Prostate	151	120 (79)	109 (72)	127 (84)	7 (5)	9 (6)	10 (7)	8 (5)	25 (17)	12 (8)	80 (53)			
Skin melanoma	28	18 (64)	19 (68)	22 (79)	4 (14)	2 (7)	2 (7)	0 (0)	3 (11)	6 (21)	11 (39)			

CPRD = Clinical Practice Research Datalink; GOLD = General Practitioner Online Database; HES = Hospital Episode Statistics; NCDR = National Cancer Data Repository.

Formatted: Space Before: 0.4 line, After: 0.4 line

Formatted: Space Before: 0.4 line, After: 0.4 line

Formatted: Space Before: 0.4 line, After: 0.4 line

Formatted: Space Before: 0.4 line, After: 0.4 line

Formatted: Space Before: 0.4 line, After: 0.4 line

^a Cases identifiable in this data source (ie, including cases identifiable in any of the other two data sources).

^b Cases identifiable only in this data source (ie, excluding cases identifiable in either of the other two data sources).

^c Cases identifiable in all the three sources (ie, overlap among the three sources).

Table 2. Characteristics of Study Patients by Source of Cancer Data and Drug at Cohort Entry

		All t	ypes			Blac	lder		Breast					Colorectal			
		In GOLD data		Not in GOLD data		In GOLD data		Not in GOLD data		In GOLD data		ot in D data	In GOLD data			ot in D data	
	n	(%)	n	(%)	n	(%)	n	(%)	n	(%)	n	(%)	n	(%)	n	(%)	
Total	492	(68.3)	228	(31.7)	70	(61.9)	43	(38.1)	130	(90.3)	14	(9.7)	56	(64.4)	31	(35.6)	
Age at outcome																	
18-54	40	(88.9)	5	(11.1)	1	(100)	0	(0.0)	25	(96.2)	1	(3.8)	0	(0.0)	0	(0.0)	
55-64	83	(72.2)	32	(27.8)	9	(56.3)	7	(43.8)	27	(96.4)	1	(3.6)	5	(71.4)	2	(28.6)	
65-74	134	(67.7)	64	(32.3)	18	(66.7)	9	(33.3)	29	(85.3)	5	(14.7)	18	(69.2)	8	(30.8)	
75+	235	(64.9)	127	(35.1)	42	(60.9)	27	(39.1)	49	(87.5)	7	(12.5)	33	(61.1)	21	(38.9)	
Sex																	
Male	230	(68.7)	105	(31.3)	45	(68.2)	21	(31.8)	0	(0.0)	0	(0.0)	21	(63.6)	12	(36.4)	
Female	262	(68.1)	123	(31.9)	25	(53.2)	22	(46.8)	130	(90.3)	14	(9.7)	35	(64.8)	19	(35.2)	
Year of outcome																	
2004 and 2005	49	(72.1)	19	(27.9)	9	(75.0)	3	(25.0)	12	(100)	0	(0.0)	4	(66.7)	2	(33.3)	
2006 to 2008	230	(72.6)	87	(27.4)	33	(67.3)	16	(32.7)	59	(93.7)	4	(6.3)	24	(70.6)	10	(29.4)	
2009 and 2010	213	(63.6)	122	(36.4)	28	(53.8)	24	(46.2)	59	(85.5)	10	(14.5)	28	(59.6)	19	(40.4)	
Increased cardiovascular risk ^a	295	(65.4)	156	(34.6)	48	(63.2)	28	(36.8)	57	(87.7)	8	(12.3)	36	(63.2)	21	(36.8)	

	•			_				-				-				
		All ty	/pes			Blad	lder			Bre	ast		Colorectal			
		In GOLD data		ot in O data	In GOLD data		Not in GOLD data		In GOLD data		Not in GOLD data		In GOLD data			ot in D data
	n	(%)	n	(%)	n	(%)	n	(%)	n	(%)	n	(%)	n	(%)	n	(%)
Index of Multiple Deprivation (quintile	es)														
Q1	121	(70.8)	50	(29.2)	19	(61.3)	12	(38.7)	27	(93.1)	2	(6.9)	13	(61.9)	8	(38.1)
Q2	140	(68.0)	66	(32.0)	19	(57.6)	14	(42.4)	41	(89.1)	5	(10.9)	14	(58.3)	10	(41.7)
Q3	96	(70.6)	40	(29.4)	16	(80.0)	4	(20.0)	28	(93.3)	2	(6.7)	11	(68.8)	5	(31.3)
Q4	67	(64.4)	37	(35.6)	8	(72.7)	3	(27.3)	19	(82.6)	4	(17.4)	10	(62.5)	6	(37.5)
Q5	68	(66.0)	35	(34.0)	8	(44.4)	10	(55.6)	15	(93.8)	1	(6.3)	8	(80.0)	2	(20.0)
Smoking																
Current	59	(61.5)	37	(38.5)	7	(46.7)	8	(53.3)	11	(91.7)	1	(8.3)	4	(66.7)	2	(33.3)
Former	248	(67.6)	119	(32.4)	44	(66.7)	22	(33.3)	49	(87.5)	7	(12.5)	28	(62.2)	17	(37.8)
Non-smoker	185	(72.3)	71	(27.7)	19	(59.4)	13	(40.6)	70	(92.1)	6	(7.9)	24	(66.7)	12	(33.3)
Unknown	0	(0.0)	1	(100)	0	(0.0)	0	(0.0)	0	(0.0)	0	(0.0)	0	(0.0)	0	(0.0)
Body mass index																
< 20	15	(65.2)	8	(34.8)	4	(57.1)	3	(42.9)	5	(100)	0	(0.0)	0	(0.0)	1	(100)
20 to < 25	113	(73.4)	41	(26.6)	13	(65.0)	7	(35.0)	23	(95.8)	1	(4.2)	19	(73.1)	7	(26.9)
25 to < 30	139	(70.2)	59	(29.8)	26	(81.3)	6	(18.8)	33	(89.2)	4	(10.8)	14	(60.9)	9	(39.1)
30 to < 40	81	(65.3)	43	(34.7)	9	(45.0)	11	(55.0)	28	(93.3)	2	(6.7)	7	(63.6)	4	(36.4)
40+	9	(60.0)	6	(40.0)	0	(0.0)	0	(0.0)	3	(60.0)	2	(40.0)	0	(0.0)	0	(0.0)
Unknown	135	(65.5)	71	(34.5)	18	(52.9)	16	(47.1)	38	(88.4)	5	(11.6)	16	(61.5)	10	(38.5)
Hypertension (diagnostic codes and/or medications)	467	(67.7)	223	(32.3)	68	(61.8)	42	(38.2)	121	(90.3)	13	(9.7)	55	(64.0)	31	(36.0)

	-	All types				Bladder				Bre		Colorectal				
		In GOLD data		Not in		In GOLD		Not in		In GOLD		ot in D data	In GOLD data			ot in D data
	n			(%)	n	(%)	n	(%)	n	(%)	n	(%)	n	(%)	n	(%)
Diabetes (diagnostic codes and/or medications)	74	74 (62.2)		(37.8)	5	(31.3)	11	(68.8)	14	(93.3)	1	(6.7)	7	(46.7)	8	(53.3)

GOLD = General Practitioner Online Database.

^a Increased cardiovascular risk was defined as having evidence or history of diabetes, acute myocardial infarction, heart failure, coronary heart disease, atrial fibrillation, stroke, transient ischemic attack, peripheral artery disease, or peripheral vascular disease; or meeting more than one of these criteria: being a current smoker, having a diagnosis of dyslipidemia, having a diagnosis of hypertension.

Table 2. Characteristics of Study Patients by Source of Cancer Data and Drug at Cohort Entry (continuation)

		Corpus	s utei	i	Kidney and renal pelvis					ng and	bron	chus	Non-Hodgkin lymphoma				
		In GOLD data		ot in D data		GOLD lata	Not in GOLD data		In GOLD data		Not in GOLD data		In GOLD data			ot in D data	
	n	(%)	n	(%)	n (%)		n	(%)	n	(%)	n	(%)	n	(%)	n	(%)	
Total	16	(57.1)	12	(42.9)	10	(35.7)	18	(64.3)	43	(47.3)	48	(52.7)	19	(65.5)	10	(34.5)	
Age at outcome																	
18-54	2	(66.7)	1	(33.3)	0	(0.0)	1	(100)	4	(80.0)	1	(20.0)	1	(100)	0	(0.0)	
55-64	4	(80.0)	1	(20.0)	4	(57.1)	3	(42.9)	6	(37.5)	10	(62.5)	6	(100)	0	(0.0)	
65-74	5	(41.7)	7	(58.3)	3	(50.0)	3	(50.0)	12	(50.0)	12	(50.0)	5	(41.7)	7	(58.3)	
75+	5	(62.5)	3	(37.5)	3	(21.4)	11	(78.6)	21	(45.7)	25	(54.3)	7	(70.0)	3	(30.0)	
Sex																	
Male	0	(0.0)	0	(0.0)	7	(43.8)	9	(56.3)	24	(55.8)	19	(44.2)	10	(58.8)	7	(41.2)	
Female	16	(57.1)	12	(42.9)	3	(25.0)	9	(75.0)	19	(39.6)	29	(60.4)	9	(75.0)	3	(25.0)	
Year of outcome																	
2004 and 2005	1	(33.3)	2	(66.7)	2	(40.0)	3	(60.0)	2	(28.6)	5	(71.4)	2	(66.7)	1	(33.3)	
2006 to 2008	7	(63.6)	4	(36.4)	6	(54.5)	5	(45.5)	24	(49.0)	25	(51.0)	12	(80.0)	3	(20.0)	
2009 and 2010	8	(57.1)	6	(42.9)	2	(16.7)	10	(83.3)	17	(48.6)	18	(51.4)	5	(45.5)	6	(54.5)	
Increased cardiovascular risk ^a	8	(50.0)	8	(50.0)	7	(33.3)	14	(66.7)	40	(54.1)	34	(45.9)	9	(56.3)	7	(43.8)	
Index of Multiple Deprivation (q	uintil	es)															
Q1	4	(100)	0	(0.0)	2	(25.0)	6	(75.0)	12	(63.2)	7	(36.8)	4	(66.7)	2	(33.3)	
Q2	3	(42.9)	4	(57.1)	4	(57.1)	3	(42.9)	11	(50.0)	11	(50.0)	4	(50.0)	4	(50.0)	
Q3	5	(71.4)	2	(28.6)	3	(50.0)	3	(50.0)	5	(27.8)	13	(72.2)	0	(0.0)	2	(100)	
Q4	2	(33.3)	4	(66.7)	0	(0.0)	3	(100)	8	(61.5)	5	(38.5)	5	(71.4)	2	(28.6)	
Q5	2	(50.0)	2	(50.0)	1	(25.0)	3	(75.0)	7	(36.8)	12	(63.2)	6	(100)	0	(0.0)	

		Corpu	s ute	ri	Kidn	ey and	renal	pelvis	Lu	ing and	bron	chus	Non-Hodgkin lymphoma			
	In (GOLD	N	Not in		GOLD	N	Not in		GOLD	Not in		In	GOLD	N	ot in
	d	ata	GOLD data		data		GOLD data		data		GOLD data		data		GOLD data	
	n	(%)	n	(%)	n	(%)	n	(%)	n	(%)	n	(%)	n	(%)	n	(%)
Smoking																
Current	1	(100)	0	(0.0)	1	(16.7)	5	(83.3)	16	(51.6)	15	(48.4)	3	(100)	0	(0.0)
Former	4	(40.0)	6	(60.0)	7	(50.0)	7	(50.0)	23	(45.1)	28	(54.9)	12	(63.2)	7	(36.8)
Non-smoker	11	(64.7)	6	(35.3)	2	(25.0)	6	(75.0)	4	(44.4)	5	(55.6)	4	(57.1)	3	(42.9)
Unknown	0	(0.0)	0	(0.0)	0	(0.0)	0	(0.0)	0	(0.0)	0	(0.0)	0	(0.0)	0	(0.0)
Body mass index																
< 20	0	(0.0)	0	(0.0)	0	(0.0)	0	(0.0)	2	(33.3)	4	(66.7)	0	(0.0)	0	(0.0)
20 to < 25	1	(33.3)	2	(66.7)	0	(0.0)	4	(100)	10	(52.6)	9	(47.4)	4	(80.0)	1	(20.0)
25 to < 30	4	(57.1)	3	(42.9)	3	(33.3)	6	(66.7)	13	(54.2)	11	(45.8)	7	(70.0)	3	(30.0)
30 to < 40	3	(42.9)	4	(57.1)	3	(60.0)	2	(40.0)	4	(28.6)	10	(71.4)	4	(80.0)	1	(20.0)
40+	2	(66.7)	1	(33.3)	0	(0.0)	1	(100)	0	(0.0)	1	(100)	1	(50.0)	1	(50.0)
Unknown	6	(75.0)	2	(25.0)	4	(44.4)	5	(55.6)	14	(51.9)	13	(48.1)	3	(42.9)	4	(57.1)
Hypertension (diagnostic codes and/or medications)	15	(55.6)	12	(44.4)	10	(37.0)	17	(63.0)	41	(46.6)	47	(53.4)	18	(64.3)	10	(35.7)
Diabetes (diagnostic codes and/or medications)	3	(60.0)	2	(40.0)	2	(28.6)	5	(71.4)	13	(59.1)	9	(40.9)	4	(80.0)	1	(20.0)

^a Increased cardiovascular risk was defined as having evidence or history of diabetes, acute myocardial infarction, heart failure, coronary heart disease, atrial fibrillation, stroke, transient ischemic attack, peripheral artery disease, or peripheral vascular disease; or meeting more than one of these criteria: being a current smoker, having a diagnosis of dyslipidemia, having a diagnosis of hypertension.

Table 2. Characteristics of Study Patients by Source of Cancer Data and Drug at Cohort Entry (continuation)

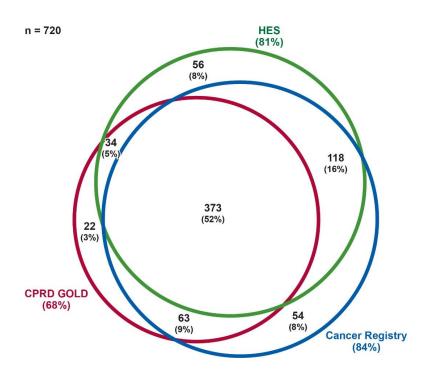
		Panc	reas			Pros	tate		Skin melanoma					
			Not i	n GOLD			Not i	n GOLD			Not i	n GOLD		
	In GO	In GOLD data		lata	In GOLD data		d	ata	In GO	LD data	d	ata		
	n	(%)	n	(%)	n	(%)	n	(%)	n	(%)	n	(%)		
Total	10	(47.6)	11	(52.4)	120	(79.5)	31	(20.5)	18	(64.3)	10	(35.7)		
Age at outcome														
18-54	1	(100)	0	(0.0)	4	(100)	0	(0.0)	2	(66.7)	1	(33.3)		
55-64	2	(66.7)	1	(33.3)	17	(81.0)	4	(19.0)	3	(50.0)	3	(50.0)		
65-74	3	(75.0)	1	(25.0)	36	(80.0)	9	(20.0)	5	(62.5)	3	(37.5)		
75+	4	(30.8)	9	(69.2)	63	(77.8)	18	(22.2)	8	(72.7)	3	(27.3)		
Sex														
Male	0	(0.0)	5	(100)	120	(79.5)	31	(20.5)	3	(75.0)	1	(25.0)		
Female	10	(62.5)	6	(37.5)	0	(0.0)	0	(0.0)	15	(62.5)	9	(37.5)		
Year of outcome														
2004 and 2005	0	(0.0)	2	(100)	15	(93.8)	1	(6.3)	2	(100)	0	(0.0)		
2006 to 2008	5	(45.5)	6	(54.5)	52	(82.5)	11	(17.5)	8	(72.7)	3	(27.3)		
2009 and 2010	5	(62.5)	3	(37.5)	53	(73.6)	19	(26.4)	8	(53.3)	7	(46.7)		
Increased cardiovascular risk ^a	7	(46.7)	8	(53.3)	74	(77.1)	22	(22.9)	9	(60.0)	6	(40.0)		
Index of Multiple Deprivation (quint	iles)													
Q1	2	(40.0)	3	(60.0)	35	(81.4)	8	(18.6)	3	(60.0)	2	(40.0)		
Q2	3	(60.0)	2	(40.0)	37	(78.7)	10	(21.3)	4	(57.1)	3	(42.9)		
Q3	1	(33.3)	2	(66.7)	23	(79.3)	6	(20.7)	4	(80.0)	1	(20.0)		
Q4	2	(33.3)	4	(66.7)	8	(72.7)	3	(27.3)	5	(62.5)	3	(37.5)		
Q5	2	(100)	0	(0.0)	17	(81.0)	4	(19.0)	2	(66.7)	1	(33.3)		

		Panc	reas			Pros	tate		Skin melanoma				
			Not i	n GOLD			Not i	n GOLD			Not i	n GOLD	
	In GO	LD data	c	data		In GOLD data		data		LD data	d	ata	
	n	(%)	n	(%)	n	(%)	n	(%)	n	(%)	n	(%)	
Smoking													
Current	1	(33.3)	2	(66.7)	13	(76.5)	4	(23.5)	2	(100)	0	(0.0)	
Former	4	(40.0)	6	(60.0)	67	(81.7)	15	(18.3)	10	(71.4)	4	(28.6)	
Non-smoker	5	(62.5)	3	(37.5)	40	(78.4)	11	(21.6)	6	(50.0)	6	(50.0)	
Unknown	0	(0.0)	0	(0.0)	0	(0.0)	1	(100)	0	(0.0)	0	(0.0)	
Body mass index													
< 20	0	(0.0)	0	(0.0)	3	(100)	0	(0.0)	1	(100)	0	(0.0)	
20 to < 25	4	(57.1)	3	(42.9)	33	(82.5)	7	(17.5)	6	(100)	0	(0.0)	
25 to < 30	3	(60.0)	2	(40.0)	32	(74.4)	11	(25.6)	4	(50.0)	4	(50.0)	
30 to < 40	1	(50.0)	1	(50.0)	16	(72.7)	6	(27.3)	6	(75.0)	2	(25.0)	
40+	0	(0.0)	0	(0.0)	3	(100)	0	(0.0)	0	(0.0)	0	(0.0)	
Unknown	2	(28.6)	5	(71.4)	33	(82.5)	7	(17.5)	1	(20.0)	4	(80.0)	
Hypertension (diagnostic codes and/or medications)	10	(47.6)	11	(52.4)	114	(79.2)	30	(20.8)	15	(60.0)	10	(40.0)	
Diabetes (diagnostic codes and/or medications)	5	(62.5)	3	(37.5)	21	(87.5)	3	(12.5)	0	(0.0)	2	(100)	

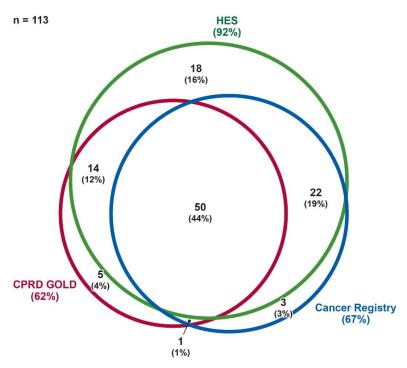
^a Increased cardiovascular risk was defined as having evidence or history of diabetes, acute myocardial infarction, heart failure, coronary heart disease, atrial fibrillation, stroke, transient ischemic attack, peripheral artery disease, or peripheral vascular disease; or meeting more than one of these criteria: being a current smoker, having a diagnosis of dyslipidemia, having a diagnosis of hypertension.

PROPORTIONAL VENN DIAGRAMS

Figure 1. Cancer Cases by Data Source Within CPRD: All Cancers Combined







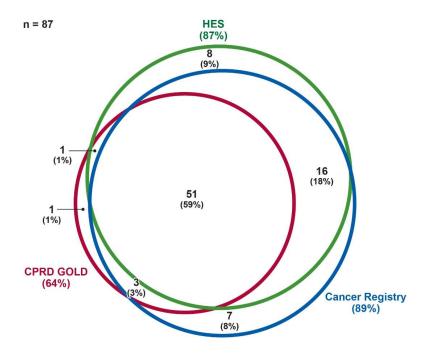
104 (72%)

CPRD GOLD (90%)

Cancer Registry (91%)

Figure 3. Cancer Cases by Data Source Within CPRD: Breast Cancer





HES (79%) n = 28 1 (4%) 8 (29%) 13 (46%) CPRD GOLD (57%) 3 (11%) Cancer Registry (96%)

Figure 5. Cancer Cases by Data Source Within CPRD: Uterine Cancer

Figure 6. Cancer Cases by Data Source Within CPRD: Kidney Cancer

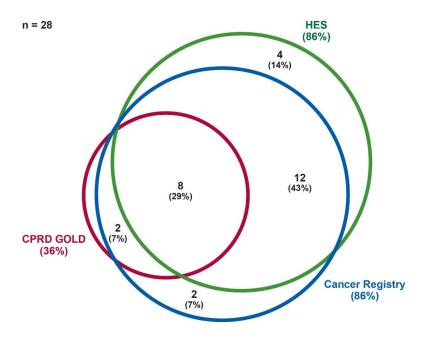


Figure 7. Cancer Cases by Data Source Within CPRD: Lung and Bronchus Cancer

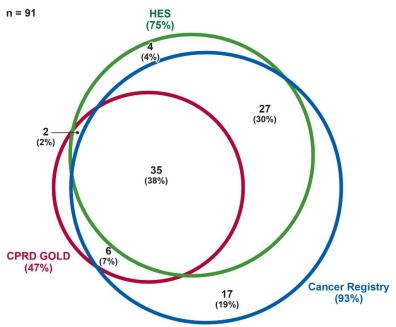
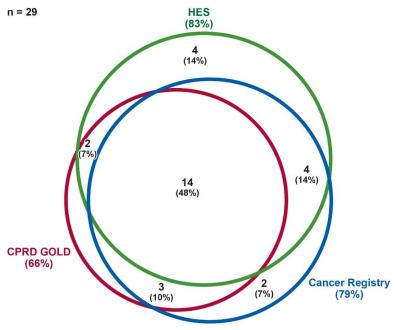
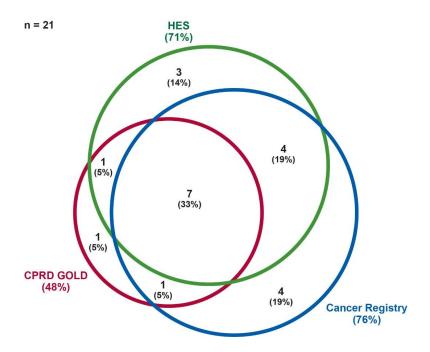


Figure 8. Cancer Cases by Data Source Within CPRD: Non-Hodgkin Lymphoma







n = 151 HES (72%) 9 (6%) 80 (53%) 10 (7%) CPRD GOLD (79%) 25 (17%) Cancer Registry (84%)

Figure 10. Cancer Cases by Data Source Within CPRD: Prostate Cancer

