Recovery of Kidney Function After Acute Kidney Disease - a Multi-Cohort Analysis

Sawhney, Simon; Ball, William; Bell, Samira; Black, Corri; Christiansen, Christian Fynbo; Heide-Jørgensen, Uffe

Published in:
Nephrology Dialysis Transplantation

DOI:
10.1093/ndt/gfad180

Publication date:
2023

Document Version
Other version

Multi-cohort analysis of recovery of kidney function after acute kidney disease

Simon Sawhney [1,2]; William Ball [1]; Samira Bell [3]; Corri Black [1,2]; Christian Fynbo Christiansen [4]; Uffe Heide-Jørgensen [4]; Simon Kok Jensen [4]; Emilie Lambourg [3]; Paul E. Ronksley [5]; Zhi Tan [6]; Marcello Tonelli [5,6] and Matthew T. James [5,6]

Supplementary material	**Page**

Supplementary table S1 – Description of AKD subsets for each cohort when limited only to people presenting after meeting one subset criterion exclusively

	Alberta				Denmark				Grampian				Tayside			
	48hr	7d	90d	365d	48hr	7d	90d	365d	48hr	7d	90d	365d	48hr	7d	90d	365d
N	30832	24785	75674	71385	16378	11299	41504	26656	5256	3988	11771	9549	6893	5096	16703	12776
age median/IQR	75 (62-84)	65 (51-78)	66 (52-79)	61 (43-75)	77 (68-84)	69 (58-79)	72 (61-81)	70 (56-81)	78 (68-85)	70 (57-81)	72 (59-82)	70 (53-81)	80 (71-86)	74 (61-83)	74 (63-83)	73 (59-82)
female %	32.6	60.6	56.4	61.5	32.1	59.9	53.2	56.4	36.1	62.8	57.2	60.4	40.0	64.3	55.8	58.8
inpatient %	95.3	88.7	65.3	53.1	96.5	85.8	59.0	46.4	94.0	85.0	51.8	36.9	95.9	90.4	61.7	47.1
ref eGFR median/IQR	57 (38-79)	105 (92-119)	89 (66-106)	96 (75-114)	55 (37-76)	102 (91-115)	86 (64-100)	89 (67-103)	56 (39-77)	102 (90-115)	88 (65-102)	90 (69-107)	53 (36-74)	99 (88-112)	84 (61-99)	86 (64-102)
ref Cr median/IQR	106 (84-145)	48 (38-62)	72 (56-92)	67 (52-85)	109 (85-148)	49 (38-63)	73 (58-94)	71 (56-90)	105 (84-143)	49 (39-63)	71 (55-91)	68 (53-86)	109 (85-147)	49 (38-65)	74 (58-95)	71 (56-91)
Comorbidities (%)																
diabetes	15.4	11.2	11.5	11.5	19.0	11.9	15.4	11.8	23.8	13.6	17.8	14.6	24.1	14.1	16.2	11.9
cancer	14.5	17.0	14.2	7.8	20.1	27.9	24.5	8.2	18.9	24.1	21.1	8.9	17.5	25.3	19.1	7.1
coronary heart disease	27.0	15.8	13.7	9.2	28.5	14.6	13.9	9.8	38.7	21.1	22.3	17.4	29.6	17.8	18.0	13.0
heart failure	20.2	11.3	9.4	5.6	18.2	8.2	9.2	5.5	21.0	8.1	9.9	6.2	19.6	8.6	9.4	5.5
stroke	15.0	13.3	10.7	7.3	13.8	13.7	10.9	9.0	13.3	12.1	8.4	6.6	11.8	11.7	7.7	6.0
peripheral arterial disease	11.6	7.8	7.0	4.6	12.1	7.7	7.4	5.4	13.8	7.7	7.0	4.6	11.3	7.1	5.4	4.0

Abbreviations: AKD, acute kidney disease; Cr, creatinine; d, day; eGFR, estimated glomerular filtration rate; hr, hour; IQR, inter-quartile range

Supplementary table S2 (part a) – Kidney function over the first year after presentation for each subset of AKD

	Alberta								Denmark							
	48hr		7d		90d		365d		48hr		7d		90d		365d	
Serum Creatinine																
ref median	94	(72-129)	67	(50-91)	73	(57-96)	67	(52-86)	94	(73-130)	70	(53-93)	75	(60-96)	71	(56-89)
presentation median	137	(110-184)	115	(82-165)	130	(97-182)	116	(87-158)	140	(112-187)	123	(89-170)	134	(102-183)	125	(95-170)
peak median	145	(114-202)	120	(85-176)	133	(98-190)	117	(87-162)	150	(117-210)	129	(92-188)	138	(104-195)	127	(96-175)
d14 median	104	(80-140)	84	(64-114)	95	(72-128)	93	(73-123)	103	(78-142)	86	(65-118)	96	(72-129)	96	(73-128)
d90 median	97	(76-127)	80	(63-104)	86	(68-113)	86	(69-111)	96	(75-128)	81	(64-106)	87	(68-114)	88	(69-115)
d365 median	98	(77-130)	81	(65-106)	84	(67-110)	81	(65-104)	98	(77-130)	83	(66-108)	86	(68-112)	84	(67-109)
Ratio vs baseline																
ratio onset median	1.41	(1.29-1.59)	1.63	(1.55-1.80)	1.66	(1.56-1.89)	1.65	(1.56-1.86)	1.42	(1.29-1.62)	1.64	(1.55-1.84)	1.67	(1.56-1.91)	1.66	(1.56-1.89)
ratio 7d median	1.47	(1.33-1.71)	1.69	(1.57-1.96)	1.69	(1.57-1.98)	1.66	(1.56-1.90)	1.50	1.34-1.78)	1.72	(1.58-2.04)	1.71	(1.58-2.04)	1.68	(1.56-1.96)
ratio 14d median	1.10	(0.91-1.33)	1.30	(1.05-1.57)	1.35	(1.04-1.60)	1.53	(1.19-1.67)	1.08	(0.90-1.31)	1.26	(1.02-1.54)	1.29	(1.03-1.58)	1.51	(1.11-1.64)
d90 median	1.05	(0.87-1.26)	1.22	(1.00-1.50)	1.16	(0.98-1.50)	1.40	(1.07-1.60)	1.03	(0.86-1.23)	1.18	(0.99-1.45)	1.14	(0.98-1.41)	1.26	(1.04-1.56)
d365 median	1.07	(0.88-1.29)	1.24	(1.02-1.52)	1.15	(0.98-1.39)	1.23	(1.04-1.52)	1.06	(0.88-1.27)	1.21	(1.01-1.47)	1.14	(0.99-1.36)	1.19	(1.02-1.48)
CKD EPI 2009																
ref median	64	(42-87)	89	(65-108)	84	(60-103)	92	(70-112)	62	(41-86)	86	(63-101)	81	(59-97)	85	(64-101)
presentation median	40	(27-54)	50	(31-75)	43	(28-62)	50	(33-73)	38	(27-52)	45	(30-67)	41	(27-57)	44	(30-62)
peak median	37	(24-52)	47	(29-72)	41	(26-61)	49	(32-72)	35	(23-49)	42	(27-64)	39	(25-56)	43	(28-62)
d14 median	57	(38-79)	72	(49-94)	62	(43-87)	65	(45-89)	56	(37-80)	69	(47-92)	61	(42-85)	61	(41-85)
d90 median	62	(43-84)	77	(54-96)	70	(49-92)	72	(51-93)	61	(42-84)	75	(53-93)	69	(48-09)	68	(47-90)
d365 median	61	42-83	76	(53-95)	73	(51-94)	78	(55-98)	60	(41-82)	73	(52-92)	69	(49-90)	72	(50-93)
CKD EPI 2021																
ref median	68	(45-92)	93	(69-109)	89	(64-106)	96	(75-114)	67	(44-90)	90	(67-104)	86	(63-101)	90	(69-104)
presentation median	42	(29-57)	53	(34-79)	46	(30-66)	53	(35-76)	41	(29-55)	48	(32-71)	43	(29-60)	47	(32-66)
peak median	40	(27-55)	50	(31-76)	44	(28-64)	52	(34-75)	38	(25-52)	45	(29-68)	42	(27-59)	46	(30-65)
d14 median	60	(41-84)	76	(52-98)	66	(46-91)	68	(48-92)	60	(40-85)	74	(50-96)	65	(44-90)	64	(44-90)
d90 median	65	(46-88)	81	(58-100)	74	(53-96)	75	(54-97)	65	(45-88)	79	(56-97)	73	(52-94)	72	(50-94)
d365 median	65	(45-87)	80	(57-99)	77	(54-89)	82	(59-101)	64	(44-87)	77	(55-96)	73	(52-94)	76	(53-97)

Abbreviations: AKD, acute kidney disease; CKD EPI, Chronic Kidney Disease Epidemiology Collaboration Equation; Cr, creatinine; d, day; hr, hour

Supplementary table S2 (part b) – Kidney function over the first year after presentation for each subset of AKD

	Grampian								Tayside							
	48hr		7d		90d		365d		48hr		7d		90d		365d	
Serum Creatinine																
ref median	88	(68-122)	66	(51-88)	71	(56-92)	68	(53-86)	94	(71-132)	69	(51-92)	74	(58-96)	71	(55-90)
presentation median	134	(108-176)	115	(86-160)	126	(95-173)	120	(90-163)	141	(111-188)	120	(87-169)	134	(100-185)	128	(94-175)
peak median	141	(111-194)	120	(88-173)	129	(97-182)	121	(90-168)	149	(115-206)	126	(90-182)	139	(103-196)	130	(95-181)
d14 median	95	(73-127)	80	(62-106)	87	(67-116)	89	(70-117)	99	(76-134)	82	(63-109)	91	(69-121)	91	(70-122)
d90 median	89	(70-116)	76	(60-97)	80	(63-103)	82	(66-104)	92	(72-120)	77	(61-100)	83	(66-109)	84	(67-110)
d365 median	90	(72-117)	77	(62-97)	79	(63-102)	78	(63-100)	93	(73-123)	79	(63-102)	83	(66-108)	81	(66-106)
Ratio vs baseline																
ratio onset median	1.44	(1.31-1.65)	1.63	(1.55-1.83)	1.66	(1.56-1.89)	1.66	(1.56-1.90)	1.42	(1.30-1.61)	1.65	(1.56-1.86)	1.67	(1.57-1.93)	1.67	(1.57-1.94)
ratio 7d median	1.51	(1.35-1.78)	1.70	(1.57-2.00)	1.69	(1.57-1.99)	1.68	(1.56-1.96)	1.48	(1.33-1.75)	1.72	(1.58-2.02)	1.72	(1.58-2.05)	1.70	(1.58-2.00)
ratio 14d median	1.07	(0.89-1.29)	1.21	(1.00-1.51)	1.21	(1.00-1.54)	1.48	(1.08-1.61)	1.05	(0.87-1.26)	1.20	(1.00-1.51)	1.21	(1.00-1.54)	1.39	(1.06-1.61)
d90 median	1.01	(0.85-1.20)	1.14	(0.96-1.38)	1.10	(0.95-1.34)	1.20	(1.01-1.54)	1.00	(0.83-1.19)	1.14	(0.95-1.39)	1.10	(0.95-1.35)	1.19	(1.00-1.53)
d365 median	1.04	(0.87-1.22)	1.16	(0.98-1.40)	1.10	(0.97-1.30)	1.15	(1.00-1.42)	1.03	(0.84-1.22)	1.17	(0.98-1.43)	1.11	(0.97-1.32)	1.16	(1.00-1.43)
CKD EPI 2009																
ref median	67	(44-88)	87	(65-103)	84	(61-100)	86	(66-106)	60	(39-83)	84	(60-100)	79	(56-96)	82	(61-101)
presentation median	40	(28-53)	48	(31-70)	43	(28-62)	46	(30-67)	37	(25-50)	44	(29-66)	39	(26-57)	42	(28-62)
peak median	37	(25-51)	46	(29-67)	42	(27-60)	45	(29-66)	34	(23-48)	41	(26-63)	37	(24-55)	41	(27-61)
d14 median	61	(42-83)	74	(52-94)	67	(47-90)	65	(46-89)	57	(39-79)	70	(49-91)	62	(43-86)	62	(43-87)
d90 median	66	(47-86)	79	(58-96)	75	(54-94)	73	(53-93)	61	(43-93)	76	(55-93)	69	(50-90)	68	(49-90)
d365 median	66	(46-86)	78	(58-95)	76	(55-94)	77	(55-97)	60	(42-82)	74	(53-91)	70	(49-90)	71	(51-92)
CKD EPI 2021																
ref median	71	(48-93)	91	(70-106)	89	(65-104)	91	(71-108)	64	(42-88)	88	(64-102)	84	(60-100)	87	(66-104)
presentation median	43	(30-57)	51	(34-74)	46	(31-65)	49	(33-71)	39	(27-53)	47	(31-70)	42	(28-60)	44	(30-65)
peak median	40	(27-55)	49	(31-71)	44	(29-64)	48	(31-70)	37	(24-51)	44	(28-67)	40	(26-58)	43	(29-64)
d14 median	65	(45-88)	79	(56-98)	71	(50-94)	69	(49-93)	60	(42-84)	75	(52-95)	66	(46-90)	66	(46-91)
d90 median	70	(50-91)	84	(62-99)	79	(57-98)	77	(56-97)	65	(47-88)	80	(58-97)	74	(53-94)	73	(53-94)
d365 median	70	(50-90)	83	(62-99)	80	(58-98)	81	(58-100)	65	(45-87)	78	(57-95)	74	(53-94)	76	(54-96)

Abbreviations: AKD, acute kidney disease; CKD EPI, Chronic Kidney Disease Epidemiology Collaboration Equation; Cr, creatinine; d, day; hr, hour

Supplementary table S3 – One year outcomes for each subset of AKD

	Alberta				Denmark				Grampian				Tayside			
	48hr	7d	90d	365d	48hr	7d	90d	365d	48hr	7d	90d	365d	48hr	7d	90d	365d
Outcomes of all people who meet a subset definition either in combination or exclusively																
N	100278	101075	136465	93640	57659	56043	77254	37240	18620	18524	22767	13252	24932	24441	31696	18011
Status at 1year (%) blood tests not carried forward if missing																
dead	33.7	33.4	29.9	15.4	42.0	41.9	37.0	23.8	39.3	39.8	38.1	23.4	44.2	44.9	41.8	27.4
non-recovery	20.0	30.2	24.3	28.9	17.1	25.8	22.4	27.1	14.6	22.3	17.6	23.2	13.4	21.3	17.9	23.5
recovery	36.3	25.4	35.2	33.1	33.6	24.5	33.1	33.3	36.8	27.9	35.6	36.7	34.7	25.3	32.3	34.3
untested at 91-365d	10.0	10.9	10.6	22.6	7.2	7.7	7.5	15.8	9.3	10.0	8.7	16.7	7.8	8.5	7.9	14.7
Status at 1year (%) blood tests carried forward if missing																
dead	33.7	33.4	29.9	15.4	42.0	41.9	37.0	23.8	39.3	39.8	38.1	23.4	44.2	44.9	41.8	27.4
non-recovery	23.6	36.6	30.3	45.0	19.4	29.8	26.0	36.8	17.5	27.0	22.0	33.1	15.7	25.4	21.8	32.0
recovery	42.7	30.0	39.8	39.6	38.6	28.3	37.0	39.3	43.3	33.2	29.9	43.5	40.1	29.6	36.4	40.5
Outcomes limited only to only those presenting with one subset definition exclusively																
N	30832	24785	75674	71385	16378	11299	41504	26656	5256	3988	11771	9549	6893	5096	16703	12776
Status at 1year (%) blood tests not carried forward if missing																
dead	25.4	23.7	23.8	12.4	33.6	33.2	31.6	20.3	31.8	32.3	32.8	20.0	36.2	36.8	36.2	24.4
non-recovery	23.6	45.6	25.9	28.9	21.3	42.2	23.9	27.5	17.3	37.2	19.3	24.4	17.3	35.9	20.0	24.7
recovery	38.7	14.8	37.1	32.8	36.6	13.9	35.6	34.2	40.9	17.8	37.7	36.3	38.1	15.7	34.3	34.1
untested at 91-365d	12.4	15.8	13.3	25.9	8.5	10.8	8.9	18.0	10.0	12.6	10.2	19.3	8.4	11.6	9.5	16.8
Status at 1year (%) blood tests carried forward if missing																
dead	25.4	23.7	23.8	12.4	33.6	33.2	31.6	20.3	31.8	32.3	32.8	20.0	36.2	36.8	36.2	24.4
non-recovery	28.1	57.7	34.1	48.0	23.8	50.0	28.6	39.0	20.0	46.3	25.1	36.6	19.5	43.5	25.0	34.9
recovery	46.6	18.6	42.1	39.6	42.7	16.9	39.8	40.7	48.2	21.4	42.1	43.4	44.4	19.7	38.8	40.7

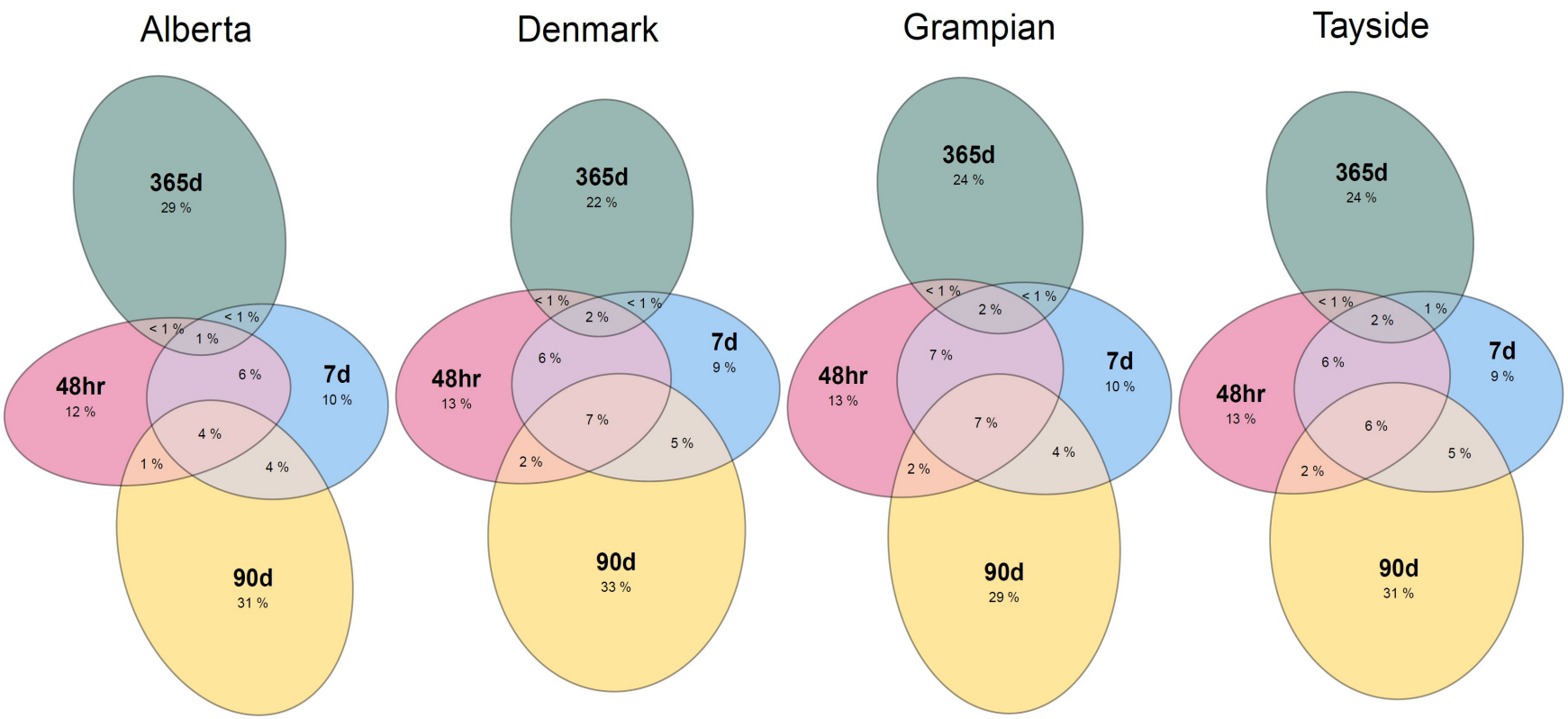
Abbreviations: AKD, acute kidney disease; d, day; hr, hour

Supplementary table S4 – Combinations of subset definitions met during each AKD episode and one year outcomes

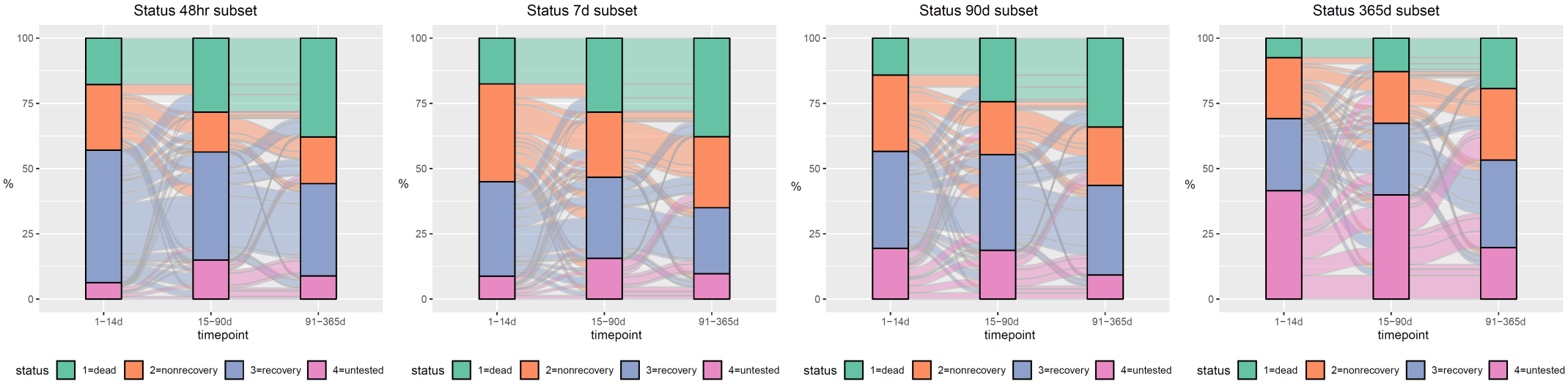
				N	Status at 1 year		
48hr	7d	90d	365d		dead	non-recovery	recovery
1	0	0	0	46804	26.1	25.4	48.5
1	1	0	0	28179	30.0	36.8	33.2
1	0	1	0	13529	42.6	20.5	36.9
1	0	0	1	6821	34.4	25.2	40.4
1	1	1	0	49474	44.4	26.6	29.0
1	1	0	1	10995	36.0	23.6	40.4
0	1	0	0	38311	26.0	54.8	19.2
0	1	1	0	22000	36.6	32.8	30.6
0	1	0	1	3565	22.8	27.6	49.6
0	0	1	0	132262	25.8	31.8	42.4
0	0	0	1	112928	14.3	44.9	40.8
Any combination (overall)				464868	26.6	35.4	38.0

Abbreviations: AKD, acute kidney disease; d, day; hr, hour

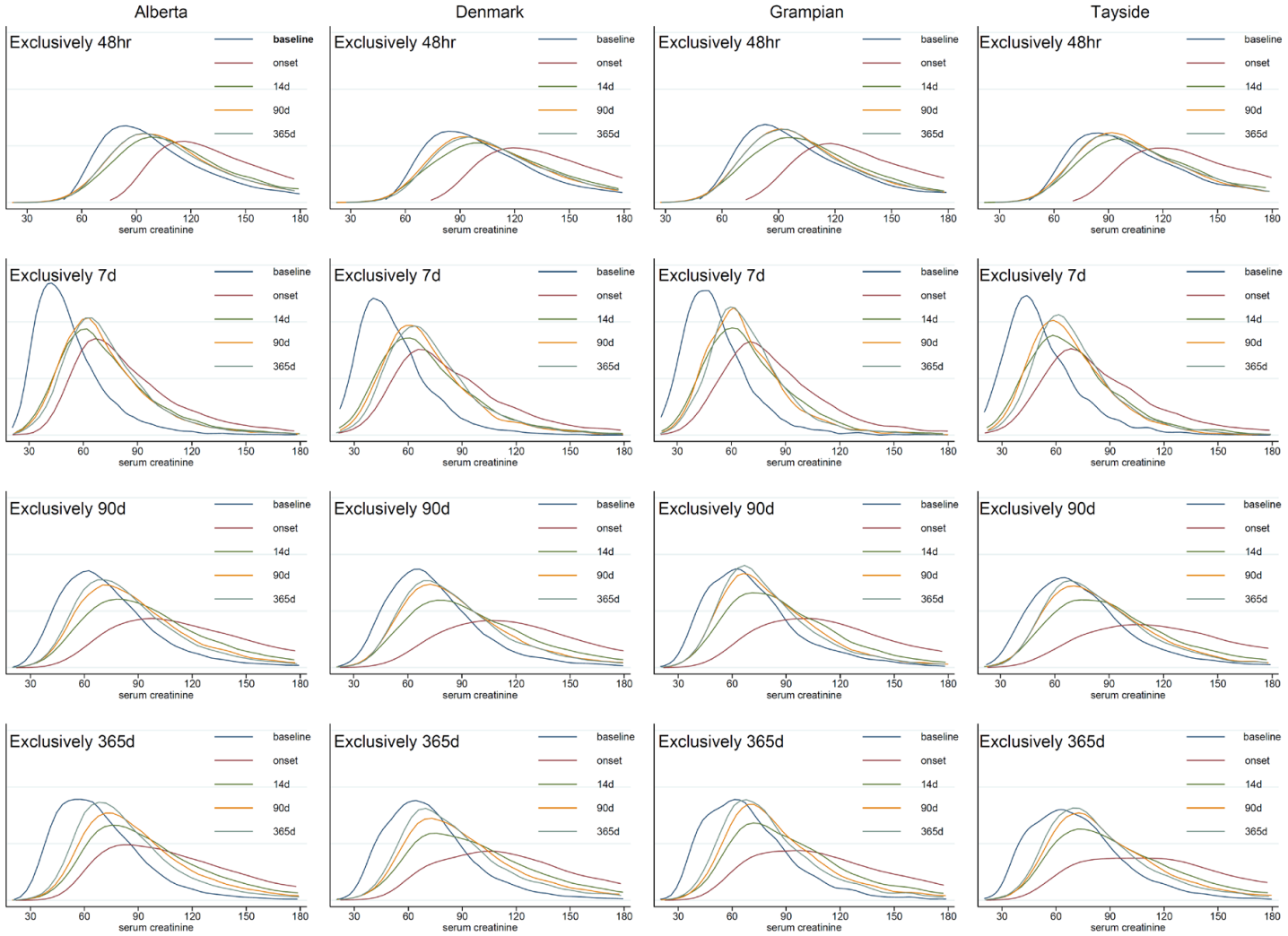
Supplementary figure S1 – Proportions and overlap of people meeting each combination of AKD criteria in each cohort (co-presenting on the same day as first AKD onset)



Supplementary figure S2 – Distribution of creatinine over the course of one year according to each AKD subset definition and cohort when limited only to people presenting after meeting one subset criterion exclusively



Supplementary figure S3 – Proportions of patients with kidney function recovery status over the first year for each AKD subset when those without bloods tests during an interval are included as a separate group



Protocol: Multi-cohort analysis of recovery of kidney function after acute kidney disease

Aim:

To determine the feasibility of comparing AKD epidemiology across geographically different populations, evaluate subsets of AKD criteria, and the consistency of a definition of kidney recovery

Objective:

Evaluate one year kidney function (both absolute and relative to baseline) by cohort population, over time, and definition subset criterion to inform definitions of AKI, AKD, and kidney recovery.

Population:

Population laboratory dataset constructed from all people in the population with IDMS aligned serum creatinine blood tests between 1st Jan 2009 and 31st Dec 2019.

Exclude from the dataset any blood tests (note not people) in instances where the test was done on a person aged <18, or where the test was done on a person receiving long term renal replacement therapy (i.e. cannot develop AKD as already have established kidney failure).

Exposure:

The first instance of AKD between 2011-2018 (two year run in to avoid the prevalent pool) based on changes in serum creatinine using a subsets of KDIGO based AKI/AKD criteria. This will involve use of the Aberdeen AKI definition that loops through blood tests meeting narrow and broad interpretations of KDIGO criteria and arranges into 90d episodes.

1. 26 micromol/l change in creatinine in 48 hrs
2. 1.5x rise in creatinine compared to lowest in last 7 d.
3. Ascertained reference creatinine from median 8-90 d if available
4. Ascertained reference creatinine from median 91 – 365 if 8-90d value not available.

Each of these subsets will be characterised in combination, separately in parallel, and in mutually exclusive subsets. Event start will be the date on which the respective definition was met, and severity onset based on the peak creatinine within 7d of day of onset vs the reference creatinine determined day of onset.

Characteristics:

Age, sex, blood test pattern, comorbidities (see table), context of presentation (hospitalised or not).

Outcomes:

Kidney function at 14d, 90d and 365d (as creatinine and eGFR), and recovery relative to baseline reference value. Recovery will be determined by a fall in creatinine to within 1.2x baseline for all, and both within 1.2 x baseline and < 26.5micromol/L above baseline for those in definition subset 1.

Analyses:

The analysis will report the total number of people meeting each subset definition, and the total with at least one AKD episode based on any definition. For the first AKD presentation by any definition, the analysis will report the combination of overlapping subset definitions in Euler 4 set diagrams (note a Euler diagram because the 90d and 365d rules cannot co-occur).

For combinations of subset definitions tables will be produced for characteristics, and 1 year outcomes. Characteristics reported for each subset combination will be age, sex, morbidities, kidney function at reference and onset, and context of presentation.

For outcomes within each definition subset, both absolute kidney function and recovery status relative to baseline: recovered, unrecovered, no tests, dead. These will be reported at reference, AKD peak (between days 0-7), 14d, 90d, 365d. A table will provide medians and IQRs, and if feasible (pending disclosure assessment), kdensity plots of absolute function among one year survivors. Overlaid kdensity plots for absolute eGFR distribution at each time point will use "last-value-carried-forward" approach. We note methodological issues with both eGFRs (issue that the 14d result is problematic as non-steady state) or creatinines (less meaningful at 90d and one year). Accordingly, both can be reported.

The results will populate Sankey plots of flow of kidney function recovery status at 14d, 90d, 365d between states of recovered, impaired, dead assuming "last value carried forward" in main analysis, and with/without "untested" as a separate category (sensitivity analysis).

Preparation:

This is a distributed analysis. The same Stata "do" file ("p2_replicationcode.do" attached in supplementary material) applies for each site for the population study. For the do-file to work, a lab data file will need to be prepared and saved as instructed below. A second data file will contain dates of death and date on which the person started kidney replacement therapy for end-stage kidney failure for all people in the population.

Once these files are ready provide update the file-path in the do file to run in Stata and prepare the AKD episodes and characterisation. The analysis will loop through all blood tests, collapse to first presentations, and link in the morbidities. A log output file and table frames will be generated to send back for sense checking. Figures will be generated in R.

File 1:

File name – "labdata_p2.dta"
File path – "C:\AKIstudy\p2\" (if different, you will need to amend the path in the coding file)
Structure – long format (multiple lab entries per individual, each date/lab result on a different row)

Variables and cleaning instructions

studyid	numeric	Pseudonymised ID. Please ensure the index is retained so that morbidity and other event data can be merged in in the future. Note only include results for people who are aged ≥18 years on the date of sample. Note the supplied code will remove observations in people who have already developed kidney failure on the date of the result.
dos	stata dofc date format e.g. 17898 = 01jan2009	Date on which the sample was received by the laboratory. While there may be multiple lab entries on the same day in the dataset you provide, the AKD code will select only the highest creatinine on a given day and drop the rest. Samples should be all creatinines in population 01jan2009-31dec2019 with those samples after a date of long term dialysis excluded as per above.
stcreat	numeric	IDMS aligned serum creatinine. Please check that instances of truncation e.g. "<10" are retained by converting to numeric "10" rather than removing.
age	numeric	Age in years on the date of sample
inpatient	numeric, binary	1 = yes, 0 = no, applies to the location of that particular serum creatinine test result. For consistency across datasets, this includes any acute hospital setting including inpatient wards, admission/triage assessment units, emergency department.
femalesex	numeric, binary	1 = female, 0 = male

File 2:

File name – "patientdata_p2.dta"
File path – "C:\AKIstudy\p2" (must be the same directory as for labdata_p2.dta)
Structure – wide format (one entry per individual, each morbidity date in a different column)

Variables and cleaning instructions

studyid	numeric	Pseudonymised id. Please ensure the index is retained so that morbidity and other event data can be merged in in the future.
dod	stata dofc date format e.g. 17898 = 01jan2009	Date on which the person died. Blank if not dead.

RRTdate	stata dofc date format e.g. 17898 = 01jan2009	Date on which the person started RRT for end-stage kidney failure (long term dialysis or kidney transplant). Blank if has not occurred.
CAdate	stata dofc date format e.g. 17898 = 01jan2009	Date on which the person first had a hospital based diagnosis, admin, or claim code for cancer. Blank if has not occurred. ICD-10: C00-C96, except for C44 ICD-09 (Alberta): 140-165, 170-176, 179-208, 2386
CHDdate	stata dofc date format e.g. 17898 = 01jan2009	Date on which the person first had a hospital based diagnosis, admin, or claim code for coronary heart disease. Blank if has not occurred. ICD-10: I20, I21, I22, I23, I24, I25 ICD-09 (Alberta): 414, 410
CHFdate	stata dofc date format e.g. 17898 = 01jan2009	Date on which the person first had a hospital based diagnosis, admin, or claim code for heart failure. Blank if has not occurred. ICD-10: I09.9, I25.5, I42.0, I42.5–I42.9, I43, I50 ICD-09 (Alberta): 39891, 40201, 40211, 40291, 40401, 40403, 40411, 40413, 40491, 40493, 4254, 4255, 4257, 4258, 4259, 428
CVAdate	stata dofc date format e.g. 17898 = 01jan2009	Date on which the person first had a hospital based diagnosis, admin, or claim code for stroke. Blank if has not occurred. ICD-10: G45, G46, H34.0, I6 ICD-09 (Alberta): 36234, 430-438
DMdate	stata dofc date format e.g. 17898 = 01jan2009	Date on which the person first had a hospital based diagnosis, admin, or claim code for diabetes mellitus. Blank if has not occurred. ICD-10: E10-E14 ICD-09 (Alberta): 250

PADdate	stata dofc date format e.g. 17898 = 01jan2009	Date on which the person first had a hospital based diagnosis, admin, or claim code for peripheral arterial disease. Blank if has not occurred. ICD-10: I70, I71, I731, I738, I739, I771, I790, I792, K551, K558, K559, Z958, Z959 ICD-09 (Alberta): 0930, 4373, 440, 441, 4431, 4432, 4438, 4439, 4471, 5571, 5579, V434

Running the code file to generate outputs

File name – "p2_replicationcode.do"
File path – "C:\AKIstudy\p2"

Please open the file in Stata, and check and confirm the file path at the beginning of the document (line that begins "cd… *filepath*"). To run the entire code, select all code (CTRL A) and run the do file (CTRL + D, rather than cutting and pasting into the console). Any potential errors should flag red in the console window, and the analysis should terminate (assume you ran by CTRL+D).

The code file will generate the files listed below. If acceptable with your local disclosure control policy and ethics permissions, please send them to me by email. In populations of ~1 million people, we do not anticipate any counts<5.

Output files

descriptives_p2.xlsx
distributions_p2.csv
euler_p2.xlsx
flow_cf.csv
flow_ncf.csv
function_p2.xlsx

There is also an option at the end of the code of producing the kdensity plots.

Ethics permissions

Waivers of consent were provided by research ethics boards for use of health data for each region. Use of Alberta data was approved by the Conjoint Health Research Ethics Board (CHREB) of the University of Calgary (ID# REB20-0970) – including waiver of consent for use of previously collected health data in accordance with Alberta Health Information Act. Use of data from northern Denmark (the North and Central regions in Denmark) was reported to the Danish Data Protection Agency through registration at Aarhus University (record number 2016-051-000001/812). According to the Danish legislation, no ethical approval was required. For Tayside, data provision and linkage were carried out by the University of Dundee Health Informatics Centre (HIC, https://www.dundee.ac.uk/hic), with analysis of anonymised data performed in an ISO27001 and Scottish Government accredited secure safe haven. HIC Standard Operating Procedures have been reviewed and approved by the NHS East of Scotland Research Ethics Service and consent for this study was obtained from the NHS Fife Caldicott Guardian. Use of Grampian unconsented,

pseudonymised, routinely collected health data were provided by North West Research Ethics Committee (19/NW/0552), Grampian Caldicott guardian, and NHS Research and Development.