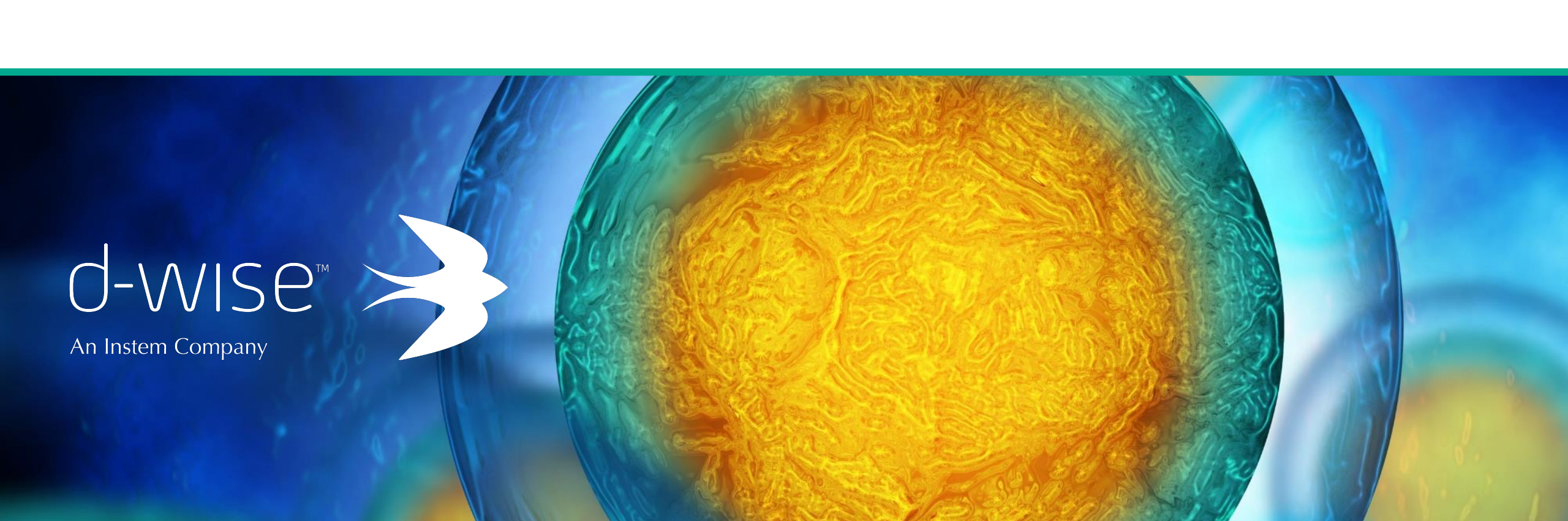




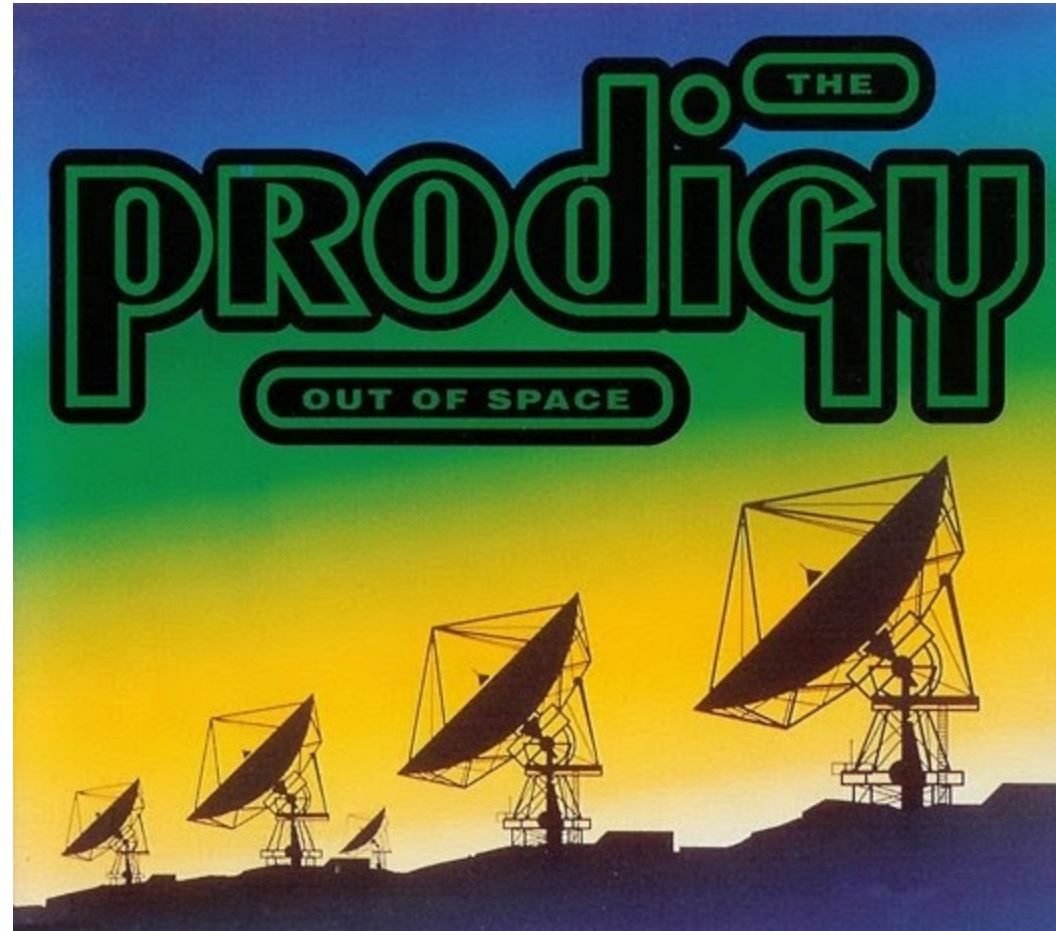
d-wise™

An Instem Company



I'll take your brain to another dimension. ~~Now close the kitchen.~~
Pay close attention.

Cathal Gallagher



1992
Aged 12

[This Photo](#) by Unknown Author is licensed under [CC BY-NC-ND](#)

Disclaimer

The opinions expressed in this presentation and on the following slides are solely those of the presenter and not necessarily those of d-wise/Instem.

d-wise/Instem do not guarantee the accuracy or reliability of the information provided herein.

k-anonymity

k-anonymity is a property possessed by certain anonymized data. The concept of k-anonymity was first introduced by Latanya Sweeney and Pierangela Samarati in a paper published in 1998.

A release of data is said to have the k-anonymity property if the information for each person contained in the release cannot be distinguished from at least $k - 1$ individuals whose information also appear in the release.

	Race	Birth	Gender	ZIP	Problem
t1	Black	1965	m	0214*	short breath
t2	Black	1965	m	0214*	chest pain
t3	Black	1965	f	0213*	hypertension
t4	Black	1965	f	0213*	hypertension
t5	Black	1964	f	0213*	obesity
t6	Black	1964	f	0213*	chest pain
t7	White	1964	m	0213*	chest pain
t8	White	1964	m	0213*	obesity
t9	White	1964	m	0213*	short breath
t10	White	1967	m	0213*	chest pain
t11	White	1967	m	0213*	chest pain

Quasi Identifiers



Sex



Height



Weight



Age



Country

- White
- Black
- Asian
- Pacific Islander

Race

- Hispanic
- Non-Hispanic

Ethnicity

- 17.4
- 19.9
- 27.2
- 35.4

BMI

PHUSE 2020 HoW – Risk measurement

AutoSave On dm1000 Complete.xlsx Search Cathal Gallagher

File Home Insert Page Layout Formulas Data Review View Help Acrobat

Clipboard Font Alignment Number Styles Cells Editing Analysis Sensitivity

AC2 =COUNTIF(AB\$2:AB\$1002,AB2)

	B	C	D	G	H	I	K	L	M	N	O	P	R	S	U	V	X	Y	Z	AA	AB	AC	AD	AE	AF
	SUBJID	Subjid_ne	Subject date offs	Country	Country Ne	sex	race	Race_ne	ethnic	Ethnic Ne	age	Age_ne	Height	Height_ne	Weight	Weight_ne	RFSTOTC	RFSTOTC_ne	RFENDTC	RFENDTC_ne	Unique Share	Freq	In Docu	Risk	Doc Risk
2	1027	2973		23 FRA	Other	F	Asian	Asian	Non-Hispanic		57	Adult	168		58		23/06/2015	16/07/2015	31/08/2015	23/09/2015	Other F Asian Adult	12	15	0.083333	0.083333
3	1170	2830		22 DEU	Other	F	Asian	Asian	Non-Hispanic		37	Adult	169		71		11/06/2015	09/07/2015	26/06/2015	18/07/2015	Other F Asian Adult	12	3		
4	1340	2660		15 ESP	Other	F	Asian	Asian	Non-Hispanic		28	Adult	171		82		07/06/2015	22/06/2015	31/08/2015	15/09/2015	Other F Asian Adult	12	9		
5	1378	2622		18 POR	Other	F	Asian	Asian	Non-Hispanic		93	Adult	159		70		19/06/2015	07/07/2015	17/07/2015	04/08/2015	Other F Asian Adult	12	10		
6	1410	2590		30 ITA	Other	F	Asian	Asian	Non-Hispanic		73	Adult	178		70		02/06/2015	02/07/2015	01/08/2015	31/08/2015	Other F Asian Adult	12	10		
7	1484	2516		8 IRL	Other	F	Asian	Asian	Non-Hispanic		44	Adult	170		84		15/06/2015	23/06/2015	23/08/2015	31/08/2015	Other F Asian Adult	12	5		
8	1524	2476		11 SWE	Other	F	Asian	Asian	Non-Hispanic		69	Adult	166		70		08/06/2015	19/06/2015	22/06/2015	03/07/2015	Other F Asian Adult	12	12		
9	1565	2435		29 CAN	Other	F	Asian	Asian	Non-Hispanic		64	Adult	174		79		25/06/2015	24/07/2015	06/08/2015	04/09/2015	Other F Asian Adult	12	8		
10	1680	2320		5 IRL	Other	F	Asian	Asian	Non-Hispanic		88	Adult	179		68		25/06/2015	30/06/2015	13/08/2015	18/08/2015	Other F Asian Adult	12	11		
11	1734	2266		4 IRL	Other	F	Asian	Asian	Non-Hispanic		79	Adult	158		66		01/07/2015	05/07/2015	02/09/2015	06/09/2015	Other F Asian Adult	12	10		
12	1835	2165		20 GBR	Other	F	Asian	Asian	Non-Hispanic		43	Adult	176		63		18/06/2015	08/07/2015	23/07/2015	12/08/2015	Other F Asian Adult	12	11		
13	1890	2110		4 SWE	Other	F	Asian	Asian	Non-Hispanic		86	Adult	150		82		21/06/2015	25/06/2015	20/08/2015	24/08/2015	Other F Asian Adult	12	1		
14	1022	2978		30 USA	USA	M	Other	Other	Non-Hispanic		81	Adult	159		109		24/06/2015	24/07/2015	24/08/2015	23/09/2015	USA M Other Adult	19	7		
15	1041	2959		23 USA	USA	M	Other	Other	Non-Hispanic		59	Adult	162		85		14/06/2015	07/07/2015	05/08/2015	28/08/2015	USA M Other Adult	19	2		
16	1052	2948		24 ESP	Other	M	Asian	Asian	Non-Hispanic		45	Adult	166		73		25/06/2015	19/07/2015	24/08/2015	17/09/2015	Other M Asian Adult	19	6		
17	1064	2936		12 CAN	Other	M	Asian	Asian	Non-Hispanic		72	Adult	184		101		15/06/2015	27/06/2015	07/09/2015	19/09/2015	Other M Asian Adult	19	8		
18	1133	2867		25 ESP	Other	M	Asian	Asian	Non-Hispanic		81	Adult	192		101		26/06/2015	21/07/2015	03/09/2015	28/09/2015	Other M Asian Adult	19	17		
19	1161	2839		28 USA	USA	M	Other	Other	Non-Hispanic		60	Adult	160		65		06/06/2015	04/07/2015	02/09/2015	30/09/2015	USA M Other Adult	19	1		
20	1184	2816		27 FRA	Other	M	Asian	Asian	Hispanic		31	Adult	179		109		08/06/2015	05/07/2015	29/07/2015	25/08/2015	Other M Asian Adult	19	8		
21	1208	2792		21 SWE	Other	M	Asian	Asian	Non-Hispanic		73	Adult	163		71		07/06/2015	28/06/2015	27/06/2015	18/07/2015	Other M Asian Adult	19	8		
22	1246	2754		29 USA	USA	M	Other	Other	Non-Hispanic		23	Adult	192		98		24/06/2015	23/07/2015	10/09/2015	09/10/2015	USA M Other Adult	19	1		
23	1305	2695		28 DEU	Other	M	Asian	Asian	Non-Hispanic		47	Adult	187		89		30/06/2015	28/07/2015	17/07/2015	14/08/2015	Other M Asian Adult	19	13		
24	1307	2693		17 USA	USA	M	Other	Other	Non-Hispanic		35	Adult	179		89		03/06/2015	20/06/2015	21/07/2015	07/08/2015	USA M Other Adult	19	8		
25	1315	2685		30 CAN	Other	M	Asian	Asian	Non-Hispanic		23	Adult	168		74		25/06/2015	25/07/2015	15/09/2015	15/10/2015	Other M Asian Adult	19	6		
26	1316	2684		10 FRA	Other	M	Asian	Asian	Non-Hispanic		93	Adult	179		65		03/06/2015	13/06/2015	25/07/2015	04/08/2015	Other M Asian Adult	19	13		
27	1346	2654		13 USA	USA	M	Other	Other	Non-Hispanic		31	Adult	175		96		20/06/2015	03/07/2015	17/08/2015	30/08/2015	USA M Other Adult	19	11		
28	1352	2648		22 USA	USA	M	Other	Other	Non-Hispanic		24	Adult	193		105		18/06/2015	10/07/2015	09/08/2015	31/08/2015	USA M Other Adult	19	16		
29	1355	2645		11 USA	USA	M	Other	Other	Non-Hispanic		26	Adult	191		77		03/06/2015	14/06/2015	06/07/2015	17/07/2015	USA M Other Adult	19	4		
30	1458	2542		4 USA	USA	M	Other	Other	Non-Hispanic		41	Adult	185		72		24/06/2015	28/06/2015	17/07/2015	21/07/2015	USA M Other Adult	19	10		
31	1467	2533		8 FRA	Other	M	Asian	Asian	Non-Hispanic		77	Adult	191		71		04/06/2015	12/06/2015	26/07/2015	03/08/2015	Other M Asian Adult	19	7		
32	1508	2492		28 USA	USA	M	Other	Other	Non-Hispanic		35	Adult	194		82		04/06/2015	02/07/2015	13/08/2015	10/09/2015	USA M Other Adult	19	17		
33	1529	2471		14 SWE	Other	M	Asian	Asian	Non-Hispanic		90	Adult	184		63		01/07/2015	15/07/2015	20/09/2015	04/10/2015	Other M Asian Adult	19	14		
34	1538	2462		11 USA	USA	M	Other	Other	Non-Hispanic		45	Adult	160		92		11/06/2015	22/06/2015	29/06/2015	10/07/2015	USA M Other Adult	19	14		
35	1579	2421		16 USA	USA	M	Other	Other	Non-Hispanic		24	Adult	191		99		29/06/2015	15/07/2015	08/08/2015	24/08/2015	USA M Other Adult	19	15		
36	1642	2358		4 IRL	Other	M	Asian	Asian	Non-Hispanic		81	Adult	162		86		28/06/2015	02/07/2015	01/08/2015	05/08/2015	Other M Asian Adult	19	3		
37	1651	2349		14 PLN	Other	M	Asian	Asian	Non-Hispanic		70	Adult	196		75		24/06/2015	08/07/2015	30/08/2015	13/09/2015	Other M Asian Adult	19	18		
38	1653	2347		9 ITA	Other	M	Asian	Asian	Non-Hispanic		69	Adult	177		98		18/06/2015	27/06/2015	10/09/2015	19/09/2015	Other M Asian Adult	19	14		
39	1659	2341		11 USA	USA	M	Other	Other	Non-Hispanic		91	Adult	169		94		23/06/2015	04/07/2015	20/07/2015	31/07/2015	USA M Other Adult	19	7		
40	1693	2307		19 FRA	Other	M	Asian	Asian	Non-Hispanic		51	Adult	196		91		27/06/2015	16/07/2015	05/09/2015	24/09/2015	Other M Asian Adult	19	10		
41	1780	2220		18 USA	USA	M	Other	Other	Non-Hispanic		44	Adult	169		66		21/06/2015	09/07/2015	22/08/2015	09/09/2015	USA M Other Adult	19	5		
42	1783	2217		19 USA	USA	M	Other	Other	Non-Hispanic		46	Adult	159		64		06/06/2015	25/06/2015	01/08/2015	20/08/2015	USA M Other Adult	19	4		
43	1795	2205		26 DEU	Other	M	Asian	Asian	Non-Hispanic		44	Adult	180		107		20/06/2015	16/07/2015	10/07/2015	05/08/2015	Other M Asian Adult	19	19		
44	1834	2166		19 DEU	Other	M	Asian	Asian	Non-Hispanic		42	Adult	158		100		26/06/2015	15/07/2015	23/08/2015	11/09/2015	Other M Asian Adult	19	9		
45	1863	2130		15 FRA	Other	M	Asian	Asian	Non-Hispanic		30	Adult	166		73		30/06/2015	28/06/2015	28/06/2015	22/06/2015	Other M Asian Adult	19	10		

dm1000

Ready Display Settings 90%

PHUSE 2020 HoW – Individuals - 361

AutoSave

dm1000_age_sex_country.xlsx - Saved

Cathal Gallagher

CG

Share

Comments

FileHomeInsertPage LayoutFormulasDataReviewViewHelpAcrobat

CutCopyFormat Painter

Clipboard

Calibri11A⁺

BBIU

Font

Wrap Text

Alignment

General

Number

Conditional Formatting

Format as Table

NormalBadGoodNeutralCalculationCheck Cell

Styles

InsertDeleteFormat

Cells

AutoSumFillClear

Editing

Sort & FilterFind & Select

Editing

Analyze Data

Analysis

Sensitivity

Sensitivity

A362

1525

	A	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
337	1353	IRL	IRL	M	52	52	IRL M 52	1												
338	1636	IRL	IRL	M	54	54	IRL M 54	1												
339	1462	ITA	ITA	M	41	41	ITA M 41	1												
340	1650	ITA	ITA	M	58	58	ITA M 58	1												
341	1364	ITA	ITA	M	54	54	ITA M 54	1												
342	1702	ITA	ITA	M	50	50	ITA M 50	1												
343	1732	PLN	PLN	M	41	41	PLN M 41	1												
344	1868	PLN	PLN	M	50	50	PLN M 50	1												
345	1388	POR	POR	M	59	59	POR M 59	1												
346	1678	POR	POR	M	49	49	POR M 49	1												
347	1706	POR	POR	M	50	50	POR M 50	1												
348	1463	SWE	SWE	M	42	42	SWE M 42	1												
349	1612	SWE	SWE	M	48	48	SWE M 48	1												
350	1972	SWE	SWE	M	40	40	SWE M 40	1												
351	1215	SWE	SWE	M	58	58	SWE M 58	1												
352	1284	SWE	SWE	M	47	47	SWE M 47	1												
353	1837	USA	USA	M	77	77	USA M 77	1												
354	1746	USA	USA	M	78	78	USA M 78	1												
355	1934	USA	USA	M	61	61	USA M 61	1												
356	1945	USA	USA	M	75	75	USA M 75	1												
357	1573	USA	USA	M	68	68	USA M 68	1												
358	1981	USA	USA	M	74	74	USA M 74	1												
359	1499	USA	USA	F	65	65	USA F 65	1												
360	1134	USA	USA	F	77	77	USA F 77	1												
361	1108	USA	USA	F	47	47	USA F 47	1												
362	1525	USA	USA	F	52	52	USA F 52	1												
363	1090	CAN	CAN	F	89	89	CAN F 89	2												
364	1680	IRL	IRL	F	88	88	IRL F 88	2												
365	1270	USA	USA	M	80	80	USA M 80	2												
366	1642	IRL	IRL	M	81	81	IRL M 81	2												

dm1000

Ready

Average: 407.5 Count: 8 Sum: 1630

Display Settings

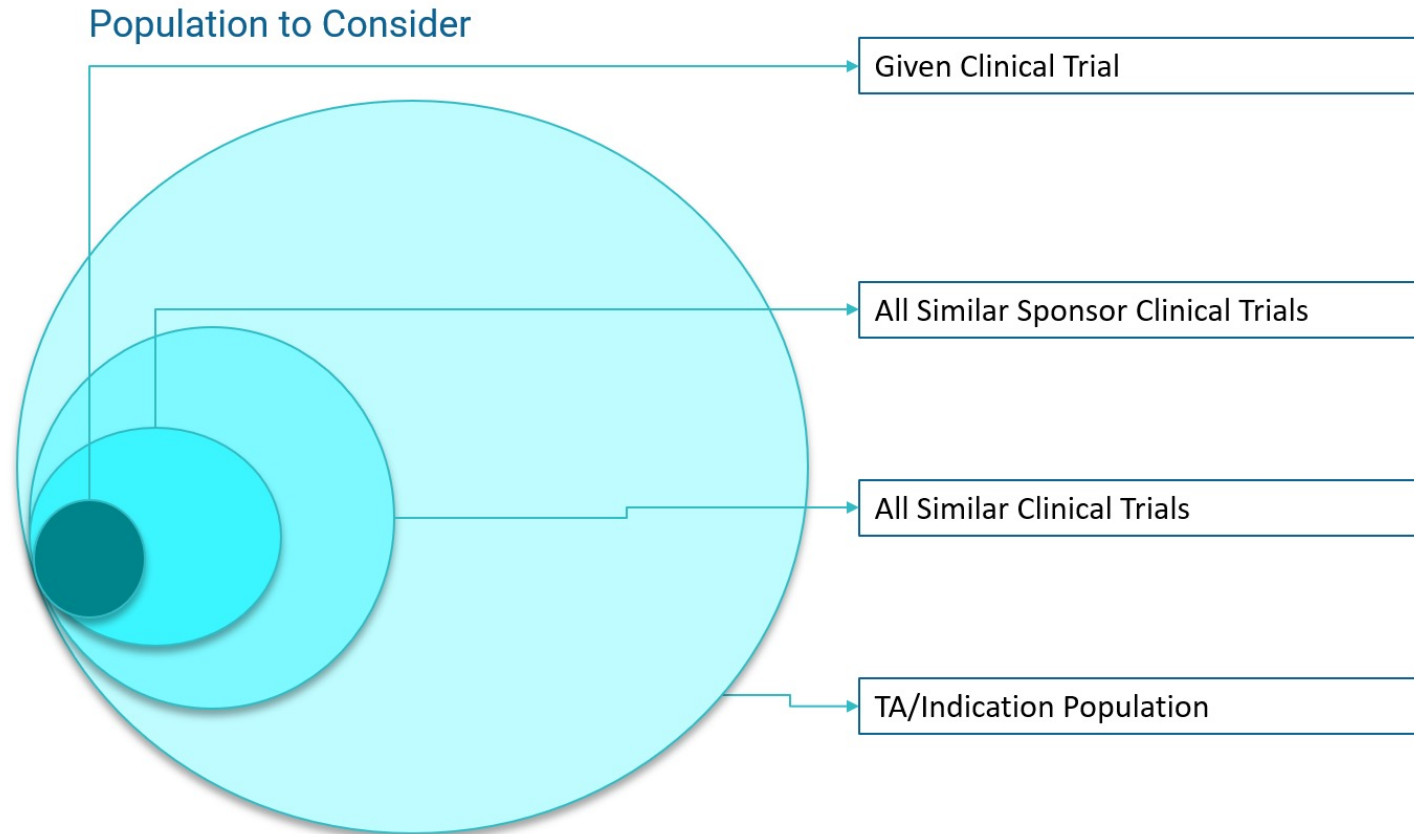
130%

k-anonymity Rules

k=11

A public sharing threshold of 0.09 indicates that a person's information cannot be shared unless cannot be distinguished from at least 10 other individuals within the same release.

What is you Population Size?



k-anonymity Rules

Age

- No Change
- 5 Year Age band
- 10 Year Age Band
- 15 Year Age Band
- 20 Year Age Band
- 25 Year Age Band
- 30 Year Age Band
- Child 10s 65+ Band
- Child Adult 65+ Band
- Drop

Country

- No Change
- Low Frequency 40
- Low Frequency 50
- Low Frequency 60
- Drop

Sex

- No Change
- Drop

100 Possible combinations

Single Dimension Rule Selection

	Age Method	Country Method	Sex Method	PoSA	Precision	PoSA Exceeded
	Numeric Band (Adult Child 65+)	Low Frequency ...	No change	.011	.516	0
	Numeric Band (10 Integer Band)	Drop	Drop	.019	.259	0
	Numeric Band (15 Integer Band)	Drop	Drop	.019	.222	0
	Drop	Low Frequency ...	Drop	.019	.125	0
	Numeric Band (30 Integer Band)	Drop	Drop	.019	.111	0
	Numeric Band (20 Integer Band)	Low Frequency ...	No change	.021	.664	0
	Drop	Low Frequency ...	Drop	.024	.188	0
	Numeric Band (5 integer Band)	Drop	Drop	.026	.296	0
	Numeric Band (25 Integer Band)	Drop	Drop	.026	.148	0
	Numeric Band (Age Child 10s 65...	Drop	Drop	.026	.074	0
	Drop	No change	Drop	.030	.250	0
	Numeric Band (10 Integer Band)	Drop	No change	.038	.676	0
	Numeric Band (15 Integer Band)	Drop	No change	.038	.639	0
	Numeric Band (30 Integer Band)	Drop	No change	.038	.528	0
	Numeric Band (10 Integer Band)	Low Frequency ...	Drop	.040	.322	0
	Numeric Band (15 Integer Band)	Low Frequency ...	Drop	.040	.285	0
	Numeric Band (30 Integer Band)	Low Frequency ...	Drop	.040	.174	0
	Drop	Low Frequency ...	No change	.048	.542	0
	Drop	Low Frequency ...	No change	.053	.604	0

Rules

- Age – 10 year age band
- Country – Drop
- Sex – No change

Risk

- 0.038
- 26 Patients with the same characteristics

Rules applied in Excel Dataset

AutoSave ☐ Off dm1000_age_sex_country.xlsx Search

File Home Insert Page Layout Formulas Data Review View Help Acrobat

Clipboard Font Alignment Number

Calibri 11 A⁺ A⁻ B I U

K2 fx $=1/\text{MIN}(\text{J2:J1002})$

	A	D	E	F	G	H	I	J	K
1	SUBJID	Country	Country Ne	sex	age	Age_nev	Unique Sha	Freq	Risk
2	1052	ESP		M	45	40	M 40	90	0.038462
3	1506	ESP		M	44	40	M 40	90	
4	1195	FRA		M	44	40	M 40	90	
5	1178	ITA		M	45	40	M 40	90	
6	1145	CAN		M	43	40	M 40	90	
7	1902	FRA		M	47	40	M 40	90	
8	1009	GBR		M	40	40	M 40	90	
9	1057	SWE		M	45	40	M 40	90	
10	1808	SWE		M	41	40	M 40	90	
11	1452	CAN		M	45	40	M 40	90	
12	1772	DEU		M	46	40	M 40	90	
13	1171	DEU		M	41	40	M 40	90	

Rules

- Age – 10 year age band
- Country – Drop
- Sex – No change

Sorted to show Max Group Size

- 90 Patients with the same characteristics

Multi Dimension Rule Selection - Primary

	Age Method	Country Method	Sex Method	PoSA	Precision	PoSA Exceeded
	Numeric Band (20 Integer Band)	Low Frequency ...	Drop	.167	.373	106
	Numeric Band (20 Integer Band)	No change	Drop	.200	.435	141
	Numeric Band (Adult Child 65+)	No change	No change	.333	.704	127
	Numeric Band (Adult Child 65+)	Low Frequency ...	No change	.333	.641	95
	Numeric Band (25 Integer Band)	Low Frequency ...	No change	.500	.690	179
	Numeric Band (Age Child 10s 65+)	Low Frequency ...	Drop	.500	.199	95
	Numeric Band (5 integer Band)	No change	No change	1.000	.963	492
	Numeric Band (10 Integer Band)	No change	No change	1.000	.926	472
	Numeric Band (5 integer Band)	Low Frequency ...	No change	1.000	.900	492
	Numeric Band (15 Integer Band)	No change	No change	1.000	.889	461
	Numeric Band (10 Integer Band)	Low Frequency ...	No change	1.000	.863	472
	Numeric Band (20 Integer Band)	No change	No change	1.000	.852	437

Rules

- Age – 5 Year age band
- Country – No change
- Sex – No change

Covers 509 out of 1001 patients

Multi Dimension Rule Selection - Secondary

	Age Method	Country Method	Sex Method	PoSA	Precision	PoSA Exceeded
	Numeric Band (30 Integer Band)	Low Frequency (60, LF)	No change	.100	.590	30
	Drop	Low Frequency (50, LF)	No change	.100	.542	20
	Numeric Band (30 Integer Band)	Drop	No change	.100	.528	10
	Numeric Band (Adult Child 65+)	Low Frequency (60, LF)	No change	.100	.516	20
	Drop	Low Frequency (60, LF)	No change	.100	.479	20
	Numeric Band (5 integer Band)	Drop	No change	.111	.713	19
	Numeric Band (5 integer Band)	Low Frequency (60, LF)	No change	.167	.775	45
	Numeric Band (25 Integer Band)	Low Frequency (60, LF)	No change	.167	.627	26
	Numeric Band (Age Child 10s 65+)	Low Frequency (60, LF)	No change	.167	.553	26
	Numeric Band (20 Integer Band)	Low Frequency (40, LF)	Drop	.167	.373	106
	Numeric Band (20 Integer Band)	Low Frequency (50, LF)	Drop	.167	.310	21
	Numeric Band (20 Integer Band)	No change	Drop	.200	.435	141
	Numeric Band (Adult Child 65+)	No change	No change	.333	.704	147
	Numeric Band (Adult Child 65+)	Low Frequency (40, LF)	No change	.333	.641	115
	Numeric Band (Adult Child 65+)	Low Frequency (50, LF)	No change	.333	.579	33
	Numeric Band (25 Integer Band)	Low Frequency (50, LF)	No change	.500	.690	179
	No change	Low Frequency (60, LF)	Drop	.500	.396	441
	Numeric Band (25 Integer Band)	Low Frequency (40, LF)	Drop	.500	.336	18
	No change	Drop	Drop	.500	.333	400

Rules

- Age – 5 year age band
- Country – drop
- Sex – No change

Covers 473 out of 1001 patients

Multi Dimension Rule Selection – Final Dimension

Rules

- Age – Drop
- Country – Drop
- Sex – Drop

Covers 19 out of 1001 patients

Summary

Primary Rules

- Age – 5 year age band
- Country – No change
- Sex – No change
- Covers 509

Secondary Rules

- Age – 5 year age band
- Country – Drop
- Sex – No change
- Covers 473

Final Rules

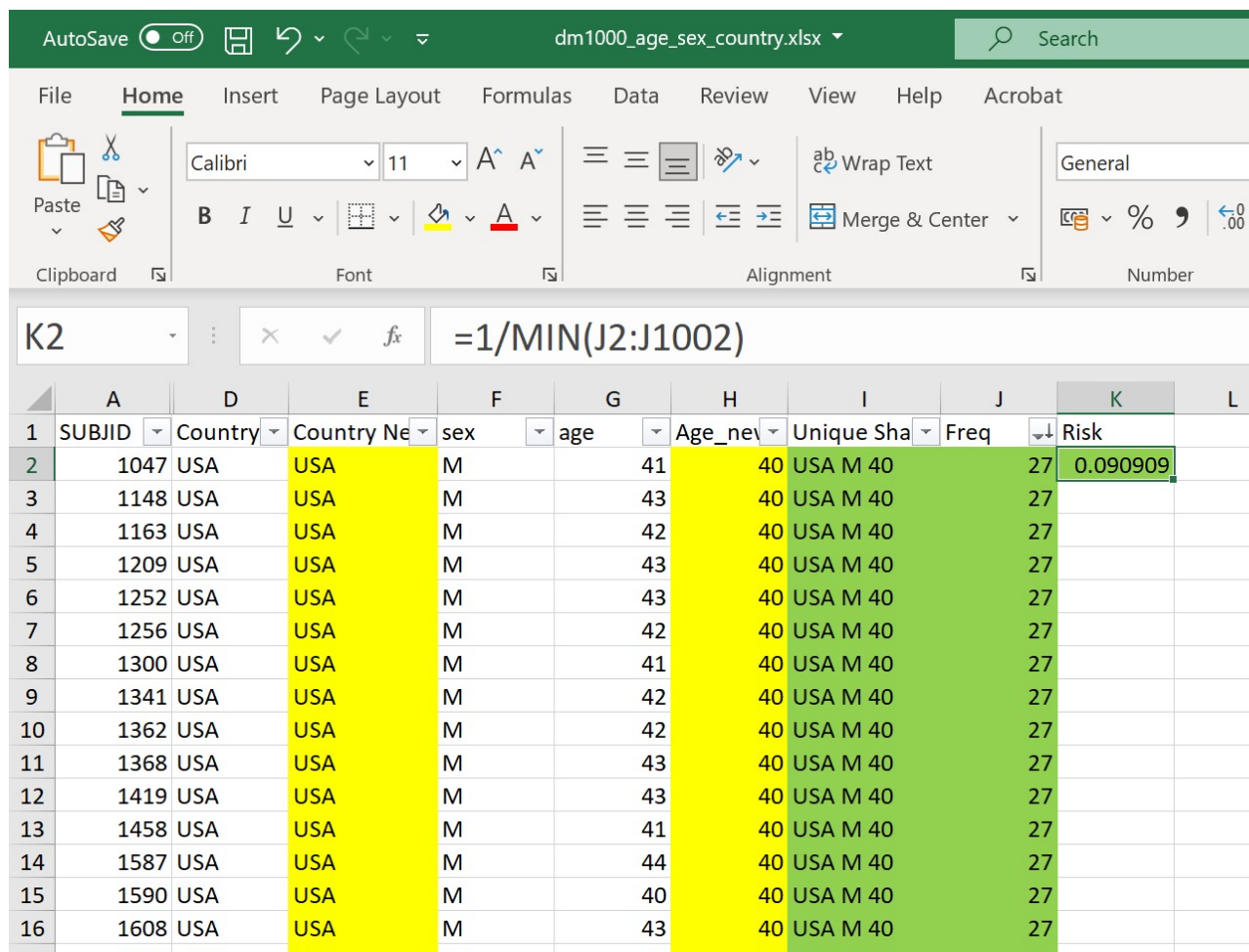
- Age – Drop
- Country – Drop
- Sex – Drop
- Covers 19

Overall Max Risk 0.09. 11 Patients

Single Dimension Rules

- Age – 10 year age band
- Country – Drop
- Sex – No change
- Covers 1001

Multi Dimensional Rules applied in Excel Dataset



	A	D	E	F	G	H	I	J	K	L
	SUBJID	Country	Country Ne	sex	age	Age_nev	Unique Sha	Freq	Risk	
2	1047	USA	USA	M	41	40	USA M 40	27	0.090909	
3	1148	USA	USA	M	43	40	USA M 40	27		
4	1163	USA	USA	M	42	40	USA M 40	27		
5	1209	USA	USA	M	43	40	USA M 40	27		
6	1252	USA	USA	M	43	40	USA M 40	27		
7	1256	USA	USA	M	42	40	USA M 40	27		
8	1300	USA	USA	M	41	40	USA M 40	27		
9	1341	USA	USA	M	42	40	USA M 40	27		
10	1362	USA	USA	M	42	40	USA M 40	27		
11	1368	USA	USA	M	43	40	USA M 40	27		
12	1419	USA	USA	M	43	40	USA M 40	27		
13	1458	USA	USA	M	41	40	USA M 40	27		
14	1587	USA	USA	M	44	40	USA M 40	27		
15	1590	USA	USA	M	40	40	USA M 40	27		
16	1608	USA	USA	M	43	40	USA M 40	27		

Primary Rules

- Age – 5 year age band
- Country – No change
- Sex – No change
- Covers 509

Secondary Rules

- Age – 5 year age band
- Country – Drop
- Sex – No change
- Covers 473

Final Rules

- Age – Drop
- Country – Drop
- Sex – Drop
- Covers 19

Sorted to show Max Group Size

- 27 Patients with the same characteristics

Multi Dimensional Problems

- Referential integrity
- False negatives in Docs
- Tracking in General
- Description in Anonymisation reports
- QC



Conclusion

Multidimensional models will usually produce higher quality results and will always produce which are at least as good as a single dimensional model from a data utility perspective. E.g. less redactions in the overall dataset and less perturbations.

However, multidimensional models can be difficult to understand and even more difficult to track through multiple tables or even through documents.

References

Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression.

<https://dataprivacylab.org/dataprivacy/projects/kanonymity/paper3.pdf>

Mondrian Multidimensional K-Anonymity

<http://pages.cs.wisc.edu/~lefevre/MultiDim.pdf>

Contact me

PPD
PPD
PPD
PPD

PPD

PPD
PPD

PPD



Contact me

Cathal Gallagher
Delivery Lead, Transparency
C: (+44)7786.967.185
Cathal.Gallagher@d-wise.com



d-wise an Instem Company
www.d-wise.com

