

SUPPLEMENTAL MATERIALS

Supplementary Figure S1: Schematic diagram of the study.

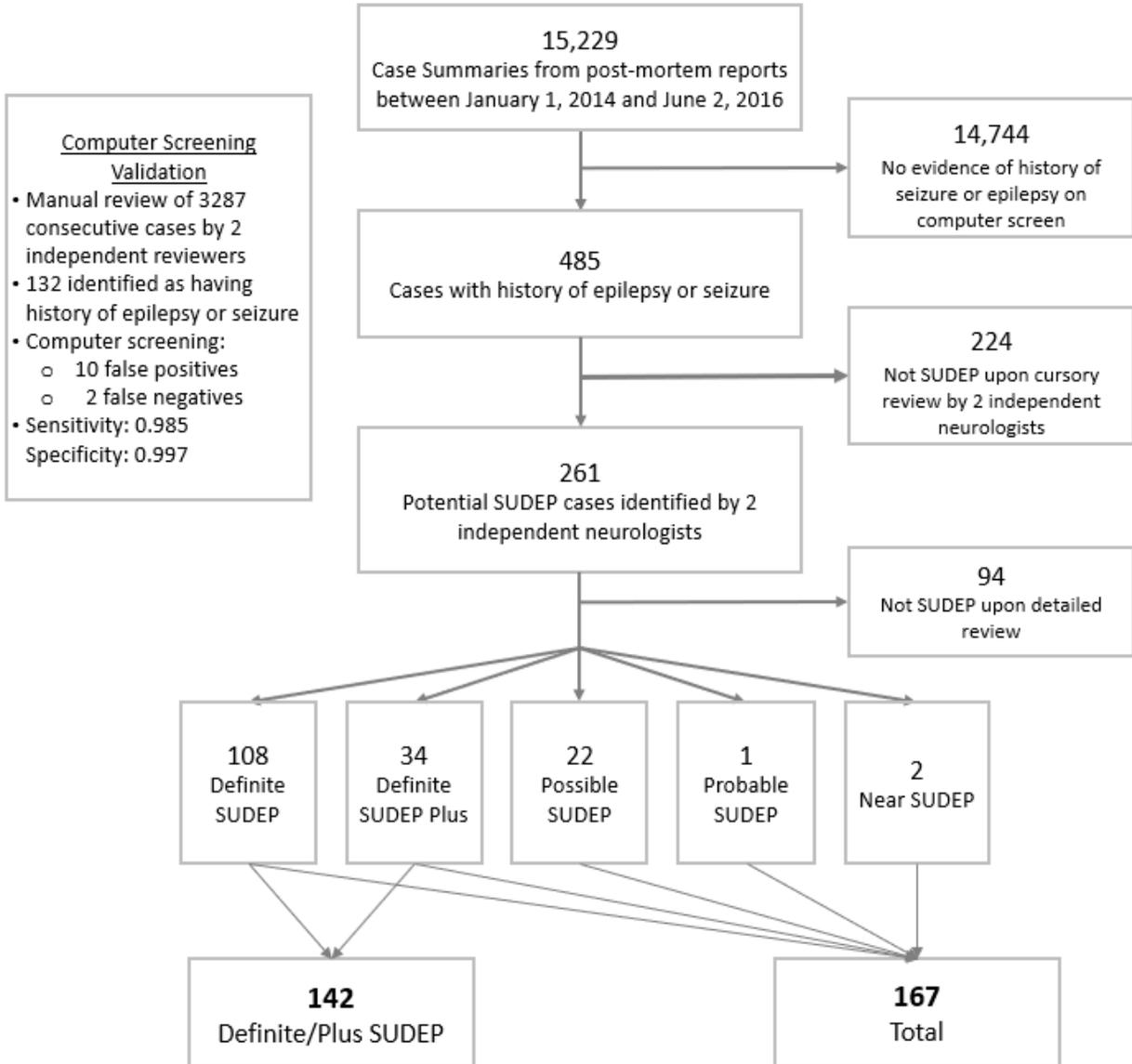


Table S1. Common features of the 62 discordant SUDEP cases, grouped by final consensus adjudication. Categories of features are not mutually exclusive.

	N	Definite	Definite Plus	Possible	Near SUDEP	Not SUDEP
Comorbid cardiac conditions	30	3	12	11		4
Toxicological findings	11	1	4	5		1
Vague seizure/epilepsy history	7	2	1	2		2
Pneumonia	5		3			2
Decomposition	5	3		2		
Other – e.g. sickle cell trait; diabetes; fatty liver; septo-optic dysplasia	9	2	3	2	1	2

Supplementary Methods:

A Python 3 script written to screen the autopsy summaries using natural language processing techniques to create word associations between keywords and other related words. The program used “case history”, “preliminary cause of death”, and “final cause of death” information available for each death to evaluate whether the case had a potential history of epilepsy or seizures.

A list of keywords commonly associated with epilepsy or seizures was created by EJD and JH. This list included “seiz-”, “epilepsy”, “sz”, “sudep”, and a list of commonly prescribed epilepsy medications.

The program was run for six iterations with each iteration including two steps. In the first step, the program processed each case, searching for keywords. All summaries including these keywords were classified as having a possible history of epilepsy or seizure. In the second step, the program formed a word-bank containing every word present in all summaries and gave them a score based on their frequency in all cases and their frequency in epilepsy-related cases. Words with high frequency in all cases (e.g. “the”, “of”) were given a low score, and words with high frequency in cases identified as having a possible history of epilepsy or seizure (e.g. “seized”, “sudden”, etc.) were given a high score. Words that occurred more than 5000 times or less than 5 times in all cases were omitted from the word-bank. Words that appeared in either the preliminary cause of death or final cause of death section were also counted twice to increase their weight. The top 6 scoring words in each iteration would then be placed in the keyword list in the next iteration.

At the end, all summaries were scored based on the number of keywords and the summary length, and all summaries with a score higher than 0 were considered epilepsy-related. The list

of summaries was then printed for manual review by two neurologists for potential SUDEP cases.

All cut-off, weighting, and iteration values described above were generated based on a training set of data, manually classified for evidence of epilepsy or seizures by two reviewers (SAL and JH). The program was also tested using a test set of data, manually classified by two reviewers (AEK and JH), with a sensitivity of 0.985 and a specificity of 0.997.