Supplemental Material

A. Explanation of Cohen's Kappa

This section provides an informal explanation of Kohen's Kappa, used in the main text to quantify inter-rater agreement. For further details, see (Carletta 1996).

Two EEGers (M & B) evalutate 100 EEGs, and declare each "normal" or (B) "abnormal". Suppose the data are as follows:

		M	
		Normal	Abnormal
B	Normal	50	30
	Abnormal	10	10

M & B agree on 50+10=60 records, hence the observed % agreement is **a=60/100=60%.** Note that both readers read the majority of records as "normal: B called 80 records (80%) abnormal, and M calls 60 records (60%) abnormal.

Although we may have no reason to believe that these different "base rates" for calling studies abnormal and the resulting 60% agreement result from anything other than each reader's honest assessment, nevertheless it often makes sense to further assess the % agreement in another way, in terms of the following hypothetical worst-case scenario:

Suppose either B or M or both are casting their 80 resp. 60 votes at random, without looking at the data. Some degree of agreement is still expected simply by chance, and the more similar the two readers base rates are, the higher the degree of chance (meaningless) agreement. Cohen's Kappa is a way of discounting for this hypothetical chance agreement.

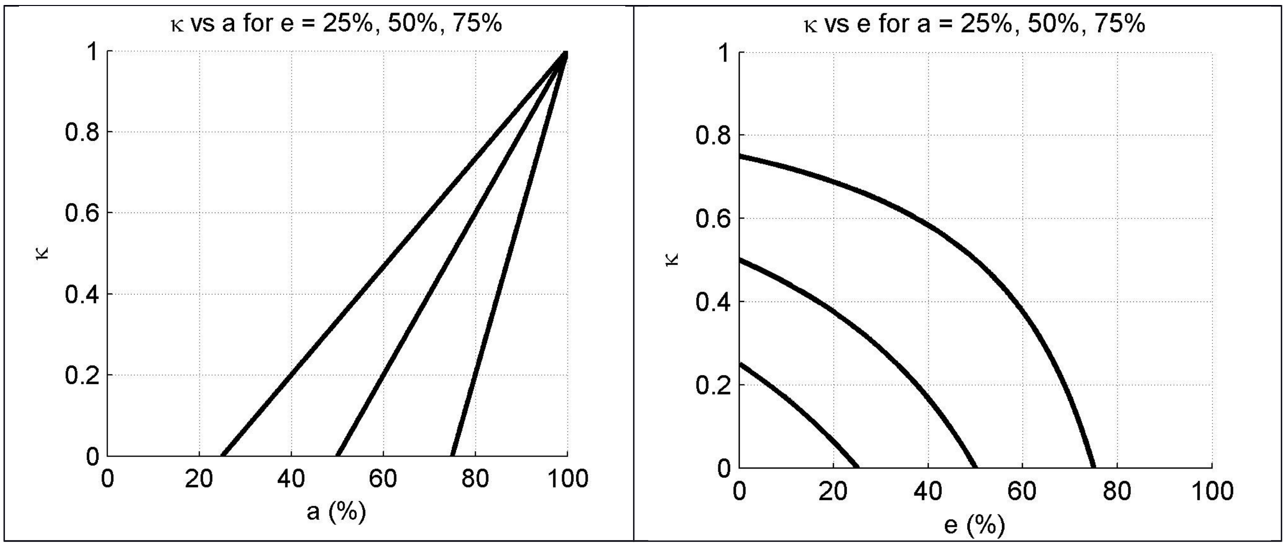
Here is how Cohen's Kappa discounts for the possibility of chance agreement, or, in other words, evaluates the degree of agreement that exists beyond what could be merely due to chance in the worst-case sense given above. If M or B or both were simply guessing independently, then they would be expected to agree "by chance" that a record is normal 0.8x0.6 = 48% of the time, and abnormal 0.2x0.4=8% of the time. Hence the total % expected agreement would be **e=48+8=56%.** Any agreement beyond 56% can be said to have been "real", or "beyond what would be expected by chance". The maximum amount of additional agreement is 100-56 = 44%.

Now consider the question: what fraction of this additional possible 44% agreement did M & B's EEG readings actually achieve? The answer is simply the additional agreement beyond chance actually observed (i.e. difference between the actual vs hypothetical chance agreement), divided by the total amount of possible additional agreement, i.e.

$$\kappa = \frac{a-e}{1-e} = \frac{60-56}{100-56} = \frac{4}{44} = 9.1$$

The dependence of the Kappa statistic on the variables a and e can be better understood by studying the plots in **Fig S1**, showing Kappa vs a holding and Kappa vs e. Note for example that perfect inter-rater agreement always produces a "perfect Kappa" value ($\kappa=1$) (leftmost plot). Conversely, the value of κ can never exceed the actual observed level of agreement (rightmost plot).

Fig. S1. Plots of Kappa vs the % observed (a) and expected (e) agreements for various values of the expected resp. observed agreement are as follows:



One should keep in mind that the "chance" agreement is a *hypothetical,* worst case-type construction. One drawback of of Cohen's Kappa is that generally the actual base rates *are* influenced by the data, i.e if both M & B call the majority of the records normal, it may be because the majority actually *are* normal, whereas Cohen's Kappa does not give credit for this sort of agreement. On the other hand, if inter-rater agreement looks good when measured by the Kappa statistic, then agreement probably really is good.

B. Burst Suppression Probability (BSP) vs. Burst Suppression Ratio (BSR)

Burst suppression depth is often quantified by the burst suppression ratio (BSR), defined as the fraction of an epoch spent in the suppressed state, i.e.

$$BSR=\frac{Duration\ of\ suppression}{Epoch\ Length}$$

(van den Broek et al. 2006; Hoffman and Edelman 1995; Schwartz, Tuttle, and Poppers 1989; Särkelä et al. 2002; Scheuer and Wilson 2004). The BSR can be recursively computed as follows. For an epoch of length T_e , or equivalently for an epoch of length $n=T_e\cdot F_s$, where F_s is the sampling rate, the number of data samples in the suppressed state within the last n samples is

$$N_t = N_{t-1} + z_t - z_{t-n-1}$$

where z_t is the output of the burst suppression segmentation algorithm described above, with $z_t = 1$ indicating a suppression, and $z_t = 0$, a bursting (non-suppression) period. The BSR at time t is thus

$$BSR_t = \frac{N_t}{n}.$$

The BSR as defined here behaves poorly as a measure of burst suppression depth in that it tends to produce estimates that are either excessively noisy (when the epoch length used in the denominator is small) or sluggish (when the epoch length is chosen long enough to produce a smooth signal). This property of the BSR is illustrated in **Fig. S2** for two epoch lengths within the range of values used in previous literature, spanning 4 seconds (Rampil and Laster 1992) to 5 minutes (Scheuer and Wilson 2004; Hoffman and Edelman 1995; Schwartz, Tuttle, and Poppers 1989).

The burst suppression probability (BSP) is a recently proposed alternative measure of burst suppression depth which, compared with the BSR, provides a superior tradeoff between signal smoothness and responsiveness to changes in the EEG. The BSP represents an estimate of the instantaneous probability that the EEG is in the suppressed state , and can be computed recursively on a sample-by-sample basis in real-time(J. Chemali et al.; J. J. Chemali et al. 2011). Examples of the result of passing the binary data output from the segmentation algorithm through the BSP algorithm with parameters optimized for adult ICU EEG data are presented in Figure 5.

Fig S2. Disadvantages of the burst suppression ratio (BSR) as a measure of burst suppression depth. (A) Binary signal resulting from the automatic segmentation algorithm. (B) BSR computed using a window size of 30 seconds (gray trace) or 2 minutes (black trace).

REFERENCES

1. Van den Broek PLC, van Rijn CM, van Egmond J, Coenen AML, Booij LHDJ. An effective correlation dimension and burst suppression ratio of the EEG in rat. Correlation with sevoflurane induced anaesthetic depth. Eur. J. Anaesthesiol. 2006 May;23(5):391–402.

2. Hoffman WE, Edelman G. Comparison of isoflurane and desflurane anesthetic depth using burst suppression of the electroencephalogram in neurosurgical patients. Anesth. Analg. 1995 Oct;81(4):811–6.

3. Schwartz AE, Tuttle RH, Poppers PJ. Electroencephalographic burst suppression in elderly and young patients anesthetized with isoflurane. Anesth. Analg. 1989 Jan;68(1):9–12.

4. Särkelä M, Mustola S, Seppänen T, Koskinen M, Lepola P, Suominen K, et al. Automatic analysis and monitoring of burst suppression in anesthesia. J. Clin. Monit. Comput. 2002 Feb;17(2):125–34.

5. Scheuer ML, Wilson SB. Data analysis for continuous EEG monitoring in the ICU: seeing the forest and the trees. J. Clin. Neurophysiol. Off. Publ. Am. Electroencephalogr. Soc. 2004 Oct;21(5):353–78.

6. Rampil IJ, Laster MJ. No correlation between quantitative electroencephalographic measurements and movement response to noxious stimuli during isoflurane anesthesia in rats. Anesthesiology. 1992 Nov;77(5):920–5.