JOHNS HOPKINS
U N I V E R S I T Y

Center for Clinical Trials

Department of Biostatistics	Department of Medicine
Department of Epidemiology	Department of Ophthalmology
Department of International Health	Oncology Center

7 January 2014

Memorandum

To: Trialists

Fr: Curtis Meinert

Re: On why *pigs is pigs, data is data*

> *I am fond of pigs. Dogs look up to us. Cats look down on us. Pigs treat us as equals.*
> (Winston Churchill)

People know my propensity for "explaining data" by simply declaring that *pigs is pigs, data is data*. I have found the statement useful because people think it has some profound meaning that escapes them. I herewith pass it onto you but with caveats:

1. Use sparingly; overuse will erode its value as an explanation and the aura of the statement
2. Best reserved for use after a long and rambling discussion of some aspect of study data
3. Best delivered after a pregnant pause in the discussion
4. Most effective if you have been silent throughout the discussion
5. Best delivered standing; remain standing for 30 seconds after delivery to let the profundity of what you said sink in; then sit down and shut up
6. Do not use hogs as a synonym for pigs; do that and you will be recognized as a fraud and your creditability will drop like a rock

Every farm boy knows the difference between hogs and pigs but, since many on my distribution list are city slickers, I feel compelled to expound on the difference.

Hogs are messy. Pigs are fastidious.

Hogs eat like hogs. Pigs do not. Indeed, pigs are the only animals on a farm that do not eat like hogs. Take cows for example. If you leave the gate open and they wander into the alfalfa field they will eat until they bloat and die. The pigs get out and nothing happens.

Hogs smell. Leastwise you doubt what I say just live downwind from a hog farm. In Minnesota there are hog farms but no pig farms. Don't ask me why. It is just the way it is.

The fact that *Pigs is Pigs* was established by Ellis Parker Butler in 1905 with publication of his book by the same title. It is a story about a dispute regarding shipping charges that arose when the father of a young boy came to the freight office to pick up a pair of guinea pigs, a male and a female. The railroad agent wanted to charge the rate for pigs (thirty cents each). The father argued that the charge should be for domestic pets (twenty five cents each). The railroad agent opted to keep the guinea pigs until the dispute could be resolved by higher ups. Months later, without any resolution in sight, the agent relented – after he had hundreds of guinea pigs in his care!

Pigs is pigs
7 January 2014

About sixty years later Jerry Cornfield (biostatistician: 1912-1979) joined pigs and data in *pigs is pigs, data is data* in a talk he gave on research and interpretation of data.

Cornfield was a likelihoodist. He largely spurned p-values and significance tests. He held that all the information from studies was in the data, not in p-values. He used the phrase to underscore the point.

I stole the phrase from Cornfield because I too am an agnostic with regard to p-values and significance testing and to aid my effort in teaching gradual students that "data" is plural for "datum". I naively thought that, as with water dripping on a rock, students would eventually learn the difference, but I was wrong! I knew I was losing the battle when newscasters used the word as a singular, e.g., most recently as in "Card data was stolen" (Target credit card breach): CBS Evening News; 20 Dec 2013.

Now I recognize my efforts at training as having been futile. Who cares if data *are* or data *is*?

What the hell? If pigs *is* pigs then data *is* data. QED!