

Essay – Automated Assistive Deep Brain Stimulation Programming

The increasing complexity and functionality of Deep Brain Stimulation (DBS) devices will likely lead to better treatments for more patients. Implanted pulse generators (IPGs) have evolved considerably: Early constant voltage devices were modified cardiac pacemakers whose train of stimulation pulses was of a single pattern, whereas today's IPGs provide constant current electrical stimulation and a greater range of stimulation parameters, and they allow programmers to program multiple trains of stimulation. Soon, advanced DBS leads with different arrangements of electrical contacts will increase markedly the number of possible electrode configurations. Future devices will provide what is known as closed-loop stimulation, which is dynamic stimulation based on feedback of implanted sensors.

Greater functionality of DBS systems will also greatly increase the challenges already faced by many programmers. Increased functionality, which breeds increased complexity, creates an incentive to forgo use of the latest IPGs' advanced capabilities. The introduction of recent IPGs had made this clear. They offered a choice between constant current stimulation and constant voltage stimulation. Despite the former's clear electrophysiological advantages, many programmers have not moved away from the comfort of the more familiar. Fewer programmers still have utilized interleaved stimulation, in which two stimulation patterns of different electrode configurations and stimulation amplitudes grant them improved control over the shape and distribution of a volume of tissue activation.

The lack of comfort with the latest IPG technology, which perhaps owes to lack of understanding of electronics, electroneurophysiology and regional anatomy, creates the risk that automated assisting devices or software will be received as a panacea that, as such, will invite

uncritical use possibly contrary to a patient's best interest. Such uncritical or uninformed use of assisted automated systems is a significant ethical issue, and it is addressed below.

IPG technology is likely poised to grow exponentially in capabilities. The present writing attempts to anticipate the increasing functionality and offerings of automated assistive DBS programming systems. Though the exact nature of the future systems cannot be accurately predicted, there are nonetheless fundamental principles that are presently known and will be important in the assessment of any future system. Those principles are the primary topics of the present writing.

Automated assistive systems as diagnostic tools

Any automated assistive system, whether it is hardware or software, is a diagnostic tool. Its purpose in DBS is most often that of determining electrode configurations, pulse train architecture, and stimulation parameters that will predict the highest benefit and the fewest adverse effects. Alternatively and also helpful, it may also be used to predict electrode configurations, pulse train architecture, and stimulation parameters that will produce the lowest benefit and the most adverse effects. In this case, the value of such systems rest on the knowledge of those electrode configurations, pulse train architecture, and stimulation parameters to avoid in order to focus resources on electrode configurations, pulse train architecture, and stimulation parameters that will produce the highest benefit with the fewest adverse effects.

The ultimate measure of any diagnostic tool is its positive and negative predictive values. Positive predictive value may be considered as a diagnostic test that indicates the probability of a condition—the most beneficial, least adverse DBS electrode configurations, stimulation parameters, and pulse trains, for example. Negative predictive value, on the other hand, is

instanced by test results that indicate the absence of a condition in a patient in whom that condition is truly absent. The question becomes, then: If an automated assistive system indicates that a particular set of parameters indicate an ineffective response, what is the probability that the patient will respond ineffectively?

There is almost always a tradeoff between positive and negative predictive values; but they are not simply reciprocal. The weight to give either is difficult to determine, because considerations often extend well beyond a patient's symptoms and disabilities. For example, when considering the treatment of a condition where the condition is relatively benign and the treatment has considerable risk, one may place a higher value on the positive predictive value, because she wishes to expose as few patients as possible to the risks of an unbeneficial treatment. As a consequence, many patients who would stand to benefit from the treatment will go unidentified. If a condition is relatively benign, however, a population of patients will experience less harm. The opposite is likely true for relatively benign treatments for serious conditions. In this case there may be less to lose by treating a possibly unresponsive patient than forgoing treatment of a possibly responsive patient. The risks attending use of iodized table salt, for example, is fairly minimal. Its use is nonetheless recommended for the general population, despite the fact that the risk of developing goiter is quite small.

Factors that determine positive and negative predictive values include sensitivity and specificity. Sensitivity indicates the probability that a person who is capable of realizing maximal benefit and experiencing minimal adverse effect will do so with use of an automated assisting system's results. Specificity indicates the probability that an automated assistive system's results indicate that a specific set of DBS parameters would not produce an optimal benefit and that a

person who is capable of an optimal response truly would not benefit from the DBS parameters resulting from the automated assisting system.

Specificity and sensitivity are insufficient measures of diagnostic utility. Rather, they must be combined with prior probabilities to produce positive and negative predictive values. Illustrative of this fact is the example of a test that is 97% specific and 97% sensitive for the diagnosis of Parkinson's disease risk. This means that 97% of those individuals at risk will be identified by a positive test. Also, 97% of those individuals who are not at risk will be identified by a negative test. Now, if it is assumed that the prevalence rate of a person at risk for Parkinson's disease in the general population over 65 years of age is 3%, then essentially every individual at risk would be identified as being at risk. The three persons out of 100 in the population of concern would thus be identified as being at risk for Parkinson's disease. It also means that 3 out of the 97 person who are not at risk would nonetheless be identified as being at risk for Parkinson's disease. Though the positive predictive value would be 100%, the negative predictive value would be 50%. The relationship between these factors is expressed numerically in the equations below:

$$\text{Positive Predictive value} = \frac{\text{sensitivity} \times \text{prior probability}}{\text{sensitivity} \times \text{prior probability} + (1 - \text{specificity}) \times (1 - \text{prior probability})}$$

$$\text{Negative Predictive value} = \frac{\text{specificity} \times (1 - \text{prior probability})}{(1 - \text{sensitivity}) \times \text{prior probability} + (\text{specificity}) \times (1 - \text{prior probability})}$$

The choice of whether the test ought to be used to identify those individuals at risk does not depend on the actual positive and negative predictive values. It depends, rather, on the costs,

in the broadest sense, of falsely treating the three individuals out of 100 who would not need the treatment versus the costs, in the broadest sense, of forgoing treatment of the 3 individuals out of 100 who would have benefited from the treatment.

The exact positive and negative predictive values may be unknowable at the moment the questions are asked. They may be unknowable because the specificities, sensitivities, and prior probabilities are unknown. Estimating those factors may be possible, however, even if the estimate is based solely on clinical experience other than that which arrives from a randomized control trial. At the very least, understanding of the equations above enables one to estimate the effects of specificities, sensitivities, and prior probabilities. In situations in which precise data is unavailable, then, a programmer may not simply do as she pleases. Nor may she become nihilistic and refuse to do anything. The expectation is that she acts reasonably.

The basis of automated assistive systems as diagnostic tools

Positive and negative predictive values as a measure of the diagnostic utility include no presumptions or assumptions regarding the nature of the test being considered, such as automated assistive system for DBS programming. The nature of the test is immaterial; outcomes alone are relevant, at least in an ideal world. For example, the VDRL test used to detect infection with syphilis is based on the presence of antibodies to beef heart tissue. The actual organism that causes syphilis, *Treponema pallidum*, is not directly detected. Yet, the VDRL is routinely used. It has sensitivity of 96% and specificity of 94%. Again, these specificities and sensitivities may only be interpreted according to prior probabilities. According to statistics reported in 2013 by the Center for Disease Control (CDC), there are approximately 10 cases of syphilis per 100,000 individuals in the U.S. state of Georgia (<http://www.cdc.gov/std/stats13/tables/26.htm>). This

means that the VDRL test is likely to detect all 10 cases in a sample of 100,000 individuals. Yet in the same test nearly 6,000 persons without syphilis will test positive for the disease. The VDRL is often applied for screening purposes. If its results are positive a second test, such as a fluorescent Treponema antibody absorbed (FTA-ABS) test, is performed. The second test has a sensitivity of 96% and a specificity of 84%. The prior probability relevant to the application of the FTA-ABS test is not the same as the population for whom the VDRL test was applied. The prior probabilities for the FTA-ABS are just those who tested positive in the VDRL test in the initial population of 100,000. The FTA-ABS test's prior probabilities consist of 10 individuals who actually have syphilis (true positives) and 6,000 individuals who do not have syphilis but tested positive on the VDRL (false positives). Though test will result in a positive in nearly all the 10 persons, only 960 noninfected persons will test positive. By combining both tests, all individuals who have syphilis will be found and only 1% of participating individuals without syphilis will be false positive for the disease.

One may argue that false diagnosis of 960 individuals as having syphilis in order to find the 10 individuals who do in fact have it seems neither efficient nor fair. Considerations as to its fairness are an ethical matter, resting on risks to society for spread of syphilis as a consequence of failure to diagnosis as weighed against the risks faced by an individual who receives a false diagnosis of syphilis. The risk of syphilis was much higher prior to 1950, any couples interested in marriage then were required to submit to a VDRL test. This requirement reduced the cases of syphilis to such an extent that today such screenings are rarely conducted.

The diagnostic ability of VDRL tests may be further improved if the individual administering it recognizes its basis. It is a test for the presence of an antibody to an antigen that is extracted from beef heart. Antibodies against the spirochete, the causative agent of syphilis,

happen to cross-react with the antigen in beef heart. Yet a number of disorders that produce antibodies also do so, including, as one might expect, a number of autoimmune disorder and infectious diseases. Identification of individuals with disorders that produce falsely positive results in a VDRL test may thus help in realizing optimal test application by removing those patients from the pool of persons without syphilis and thereby increasing the test's negative predictive value.

Direct determination of the specificity and sensitivity requires estimations of true positives, true negatives, false positives and false negatives. The critical question thus becomes: What are these? Ordinarily, the determination requires some measure that is independent of the method tested. In other words, one who seeks to determine whether an automated assistive DBS programming system has high specificity or sensitivity must know whether a patient in question had reached optimal response with its results. The issue is that of knowing whether optimal response had been reached. But what is the definition of optimal response, which would enable one to know whether programming according to the automated method reached it? In other words, what is the standard by which any automated assistive method is to be judged? It cannot simply be the best response obtained by use of the automated method. The only way to know for certain is exhaustively to test every set of DBS parameters, which is not feasible. Alternatively, the standard may be embodied by an expert who always obtains an optimal response using traditional methods. How such an expert is to be discovered, however, is difficult to know.

Determining whether a patient reached the optimal response with automated DBS programming may be unnecessary. It may do simply to determine that the automated system's results were at least as satisfactory as alternative methods, all other things being equal. If the automated system is more costly in terms of resources but clinical responses are sufficiently

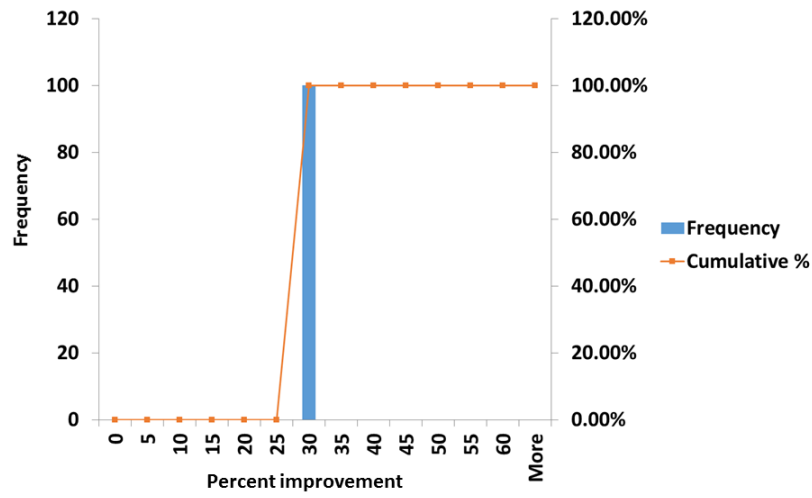
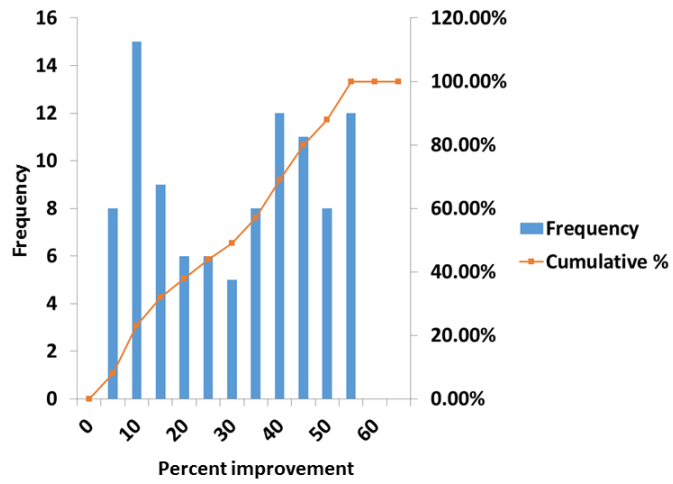
better than the alternatives, the additional resources required for the automated assisted DBD programming may be justified. Conversely, were an automated system's resource cost were small enough compared to alternatives, a lesser clinical response with the automated assisted DBS programming perhaps would be justified. These issues cannot be resolved by statistical tests that either find or fail to find statistically significant differences between the two groups.

The appropriate methods of demonstrating equivalence, superiority, or non-inferiority require prior determination of a meaningful degree of difference followed by a test of a sample size of sufficient size to support reasonable confidence of finding a previously determined clinically significant difference. Yet the sample size is not solely a function of the effect, that is, difference in the clinical status of the two groups, but also the variability of the measure used—variability in the Unified Parkinson Disease Rating Scales in the case of DBS for Parkinson's disease, for example. Illustrative of this fact is a situation in which investigators seek to demonstrate that their automated assistive DBS programming system results in better clinical responses than an alternative method. They randomize each research subject to have their DBS parameters set by either the automated process or by the alternative. Having chosen to use the motor examination part of the Unified Parkinson Disease Scales as their primary outcome measure, they determined ahead of time that a change in the score of at least 5 points would be clinically significant. Also, the investigators are aware that the variability, measured as the standard error, in the scores is approximately 13%. This means that to have an 80% probability of avoiding a type I error (finding a difference that does not truly exist) or type II error (not finding a difference that truly exists) at least 200 subjects must be enrolled; a study with fewer leaves one with no confidence, regardless of the outcome. This would be a serious logistical obstacle to the thorough evaluation of any automated assistive DBS programming method.

Physicians and healthcare professionals often are unable to postpone a decision indefinitely. Neither are they able to await the completion of an exhaustive study. Most physicians and healthcare professionals carefully consider the effect size, that is, the difference in outcomes under two different approaches to management. Decidedly fewer temper that assessment in view of the variance. Variance also depends on sample size. Thus, even were one to know the variance in a small sample, her extrapolating to a larger sample would be quite problematic: She cannot simply assume that variance in a small sample will be lower than variance in a larger sample. Nevertheless, in view of having to make a decision—and in clinical care, failing to decide to treat is a decision not to treat—one may take even the roughest estimations into consideration.

Variance in any study is a combination of contributing variances, that is, sources of error. The variability contributing to variance may be intrinsic to a subject, such as different body weights taken on different days, or it may be intrinsic to a measuring device, such as differences in several body weights taken over a relatively short time (reproducibility in the measuring device). In fact, any particular weight is affected by variability of both subject and measure. Determination of the portion of overall variance that is related to the variability of each component is quite problematic, because overall variance is not simply the sum of variances. Nonetheless, it will not be less than the greatest component variability. Variance of any automated assistive DBS programming system will be at least as great as the component with the greatest variance. This means that if the fundamental basis for the system is anatomical, such as a location within a visualized structure, then overall variance will be at least as great as the variance of the optimal stimulation site within an anatomical structure.

The importance of variance (as determined by the variability of the measures) is seen in such instances in which the mean (or average), median (the value where half the observations are less and half the observations greater), and mode (the most frequent value measured) come under consideration. **Figure 1** shows the distributions of the number of subjects who experienced a specific degree of improvement. Each of the distributions has the same mean improvement of 28%. Yet the implications are quite different. **Figure 1** shows that every subject improved by 28%. In comparison, the distribution shown in top graph in **Figure 1** indicates most subjects did not experience a 28% improvement. Indeed, any particular subject included in that distribution could likely have experienced a degree of improvement that ranged from nearly 0% to nearly 60%. The odds of experiencing 0% improvement are the same as experiencing a 58% improvement.



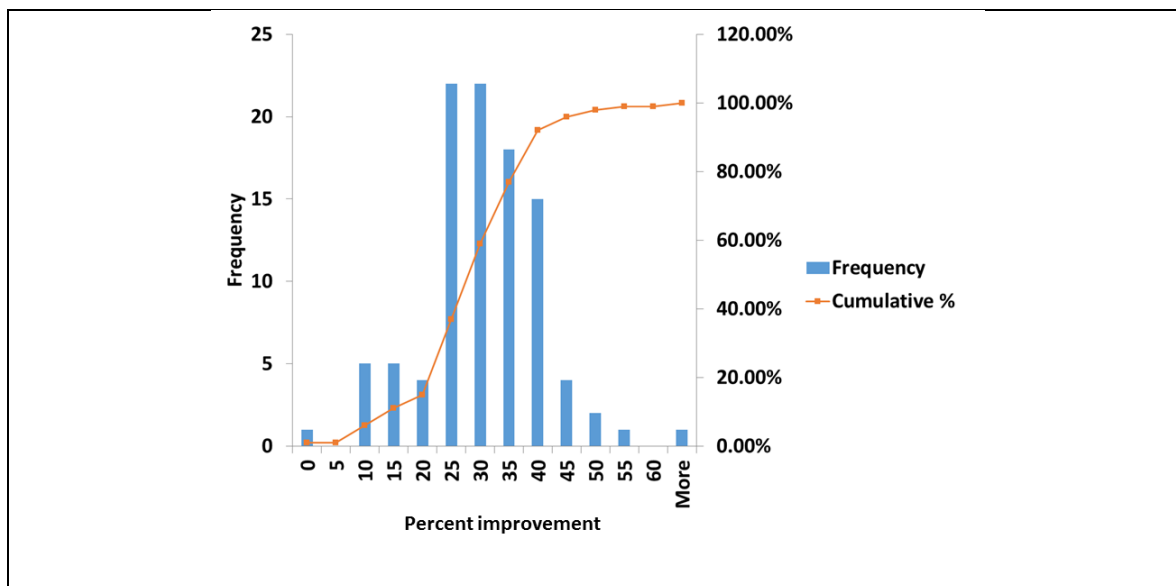


Figure 1. In the top graph, the distribution of the percentage of improvement in a hypothetical case of treatment X for disease Y is shown in blue. The height of each blue bar represents the number of subjects (frequency) whose percentage of improvement fell within the bin, which is 5% wide. Because the distribution is uniform, all percentages of improvement are equally probable. The mean of the percentages of improvement is 28%, and the range is 0% to 58%. Also shown by the orange line is the cumulative percentage function (reflects probability), which may be thought of as the frequency (or probability) of an observation (percentage of improvement) whose value is lower than a target or cutoff value. For example, the frequency or probability of a target or cutoff value being lower than 28% would be 50% (or 0.5). In the middle graph, the mean percentage of improvement for all subjects is 28%. Thus, though all patients would experience a 28% improvement, no patient would experience better than that. In the bottom graph, the distribution is normal (Gaussian). The mean of all percentages of improvement is 28%, the range 0% to 58%, and the standard deviation 10%. Evident is the fact that the

probabilities for different percentages of improvement for any one patient are considerably different in the various distributions.

The overall calculated variance is a function of the distribution of data, such as the results of an automated assistive DBS programming system. The distribution is often represented by a probability density distribution that indicates the probability of gaining a specific result. For example, in **Figure 1** (middle graph), every patient who made use of an automated assistive system would experience a 28% improvement. However, if the distribution is that which is shown in **Figure 1** (top graph), then it is only possible to say that all patients will stand an equal chance of experiencing an improvement ranging from 0 to 58%, which is not much improvement at all.

Anatomy-based automated assistive DBS programming systems

Perhaps the most intuitively appealing approach is based on modeling the stimulation-induced electric field in a construction of the regional anatomy. These models report to describe volume of tissue activation (VTA) in a stimulation field model (SFM) but in reality they describe the volume of the electrical field (Butson and McIntyre 2015). The difference is that the volume of the electrical field is based on the stimulation intensity and the resistivity of the surrounding tissue. To translate the volume of the electrical field into a volume whose neural elements are activated, specifically generate action potentials cannot be accurately predicted just on the modeling of the electrical field. Thus, it must be noted that these two concepts are not synonymous.

The SFM predicts the distribution and gradient of electric charge densities in light of the impedance imposed by the brain substance. However, the particular neural elements activated within the SFM depend on the stimulation parameters, such as pulse width immediately and then frequency components and reentrant percolation through the neural networks. For example, tissue undergoing stimulation contains many small unmyelinated axons for which only a very small pulse width is used. It is compared to the same type of tissue undergoing stimulation with a large pulse width. In both cases the SFM will predict the same distribution and gradient of electric charge densities. Yet in the case of the small pulse width, few if any axons will have action potentials. Those that do would be quite close to the stimulation electrode, and the volume of tissue activation consequently will be much smaller than the SFM. In the case of a wide pulse width, more axons will be activated to produce a larger volume of tissue activation relative to the same size SFM.

The components of the tissue within the SFM determine what regions will form the volume of tissue activation. Fortunately, this can be estimated according to knowledge of the histology in a DBS target, which must include not only structures particular to the nuclei or cortex targeted but also all neuronal elements in the vicinity, even if the latter is encountered only in passing. The level of detail afforded by these models can be extraordinary.

The process in one instantiation of the anatomical approach has been described by Butson and McIntyre (Butson and McIntyre 2015). The first step involved defining the regional anatomy. However, it is critical to refine what is meant by regional anatomy. It may be defined in terms of the histology. An example of a histological-like definition, in the case of MRI-based systems, is the relative distributions of protons among and between structures expected to lie in the region of the DBS lead. This direct approach is problematic for a number of reasons. For one,

it often is difficult to identify accurately the subthalamic nucleus. When the intended target is the ventral intermediate nucleus of the thalamus or globus pallidus interna, the specific DBS target, which is the appropriate homuncular representation, becomes impossible to identify. In DBS lead implantation in the vicinity of the globus pallidus interna, for example, it does the patient with cervical dystonia little good to place the DBS lead in the lower extremity homuncular representation, which is further from the head representation than is the diameter of the volume of tissue activation that is typical of DBS post-operative use.

In treating movement disorders, if not other indications, those who use such anatomy-based DBS programming methods are forced to infer the spatial location of the appropriate sensorimotor region of the targeted structure. This necessary act of inference, even if it is quantifiable, introduces another source of variance, which cannot but increase the overall variance. This author's experience has shown that the accuracy of a necessary inference improves only with the realization that an optimal target for DBS lead placement is not anatomically defined, in its final sense. Rather, it is physiologically defined which, owing to current state of the technology, requires that microelectrode recordings of neuronal action potentials be correlated with behavioral activations.

Perhaps in the future imaging methods such as MRI tractography may be useful, obviating the need for a physiological means of target identification. For example, if it can be shown that the homuncular representation in a subcortical structure can be reasonably inferred from identifiable homuncular representations in the cortex by visualizing the axonal projections from the appropriate homuncular representation in the cortex to the appropriate location in the subcortical structure, this may be sufficient for defining the homuncular representation in the sensori-motor regions of the subcortical structures.

One may be tempted to say that issues of accurate placement of the DBS lead are prior to DBS programming and therefore irrelevant to postoperative DBS programming. This is untrue if for no other reason than the location of DBS lead placement determines the relative prior probability that adjustments of electrode configurations, pulse train architecture, and stimulation parameters will be effective. In the abovementioned case, the closer the DBS lead is placed to the lower extremity, the less likely the DBS programming approach will see an optimal outcome in a patient with cervical dystonia.

A DBS lead's actual location constitutes the prior probability, or prevalence, in terms of whether it contains the optimal site for the DBS programming algorithm. The actual placement of the DBS lead will thus affect, independently of the DBS programming algorithm's specificity and sensitivity, the positive and negative predictive value of the DBS programming method. Anyone who studies automated assistive DBS programming, then, would be wise to differentiate the methods used to implant the DBS lead, such as use or nonuse of microelectrode recordings, because they will likely affect the prior probabilities. It may well be that prior probabilities will conduce to acceptable positive and negative predictive value when microelectrode recordings are used, but they will not do so in the absence of such physiological refinement of the DBS lead target.

The regional anatomy may be inferred from stereotactic atlas according to readily visualized co-registering landmarks in the brain and determination of a specific spatial distance to a presumed stimulation target's location. Any number of landmarks may be used, including the following examples: the line connecting the anterior and posterior commissure; in the case of targeting the subthalamic nucleus, spatial relations to the Red nucleus; in such cases in which the globus pallidus interna is targeted, the posterior limb of the internal capsule. (These landmarks

are discussed in detail in Montgomery 2015.) Yet such methods introduce variance that adds to the overall amount.

Generally, some combination of the abovementioned approaches is used post-operatively to a positive effect. The combined use does not change the variance inherent in any one method and clearly does not reduce the overall variance. It does change, however, prior probabilities with each method, which in turn increases prior probability or prevalence for a subsequent method and thus enhances positive predictive value (at the expense of negative predictive value) or, conversely, decreases prior probability or prevalence and thus enhances negative predictive value, albeit at the expense of positive predictive value. Either change in the prior probabilities may be helpful depending on whether the positive predictive value or negative predictive value is of greater clinical consideration.

Though clinical studies of the effectiveness of current instantiations are quite limited for understandable reasons, conclusions have been drawn and recommendations made nonetheless. In some studies, however, practical necessities have further undermined any confidence in the conclusions. What follows does not criticize studies conducted to date. Rather, it outlines problems that will need to be overcome in order to bolster confidence in whatever DBS programming approach is used.

One study of ten patients with Parkinson's disease employed a traditional approach to DBS programming (the approach was not completely defined) and programming based on anatomical modeling of the volume of the electrical field (Frankemolle et al. 2010). The sample size was likely too small to support any confidence that the results owed to neither a type I error nor type II error. For, example, there was no difference in the clinical outcomes which proceeded from use of the part III of the Unified Parkinson Disease Rating Scales — outcomes which may

bespeak of a failure to detect an actually existing difference (type II error). The investigators using the *n*-back test did find a difference in their test of cognitive effects. Yet this finding inspires little confidence, because it may rest on a type I error, that is, a difference that does not truly exist.

Another major problem with the abovementioned study and others like it is the use of historical controls. In the case of the abovementioned study, it is uncertain whether the electrode configurations and stimulation parameters determined according to the traditional approach are comparable to those under the model approach. Differences were only found in the cognitive test rather than the clinically measured motor function. However, this is a bit unfair because traditional approaches target motor functions: Absent a particularly obvious cognitive adverse effect, any cognitive changes would not have been considered. A more appropriate experimental design would include the traditional approach along with targeting the *n*-back test in the DBS programming. Perhaps if the programmers using the traditional approach had that opportunity, the cognitive results would have been comparable. For the outcomes measure realistic to the traditional DBS programming approach, the two methods did not differ in outcome.

The investigators did point out that the model approach required less programming time, a smaller volume of the electrical field and less battery power, all of which would not be insignificant if it were true. The investigators did not consider the ecological validity of their claims of greater efficiency as measured by resources used by the programmer according to the model approach. Yet the investigators failed to amortize the cost of the imaging and modeling prior to the programming. These costs are real.

A similar study was conducted (Pourfar et al. 2015). This study suffered from many of the abovementioned shortcomings. Again, it did not demonstrate any differences in clinical

outcomes involving anatomy-based DBS programming compared to outcomes resulting from traditional methods.

Presumptions of anatomy-based automated assistive DBS programming systems

Requisite to anatomy-based automated DBS programming systems is the presupposition of a specific and restricted region of the anatomy whose stimulation delivers optimal therapeutic effect. Current knowledge holds that it is reasonable to expect that DBS in the vicinity of an appropriate homuncular sensorimotor representation will likelier produce optimal benefit. Yet the presumption is that the representation can be identified anatomically. This presumption relates primarily to clinical improvement in motor function as it relates to movement disorders.

The same issue may be presented in the negative: the anatomy whose stimulation would produce treatment-limiting adverse effects. For example, stimulation of the posterior limb of the internal capsule or optic track is likely to produce treatment-limiting adverse effects. Because these structures may be identified with relative ease and consistency, an anatomical approach may be well-suited to avoiding stimulation that could affect them thereby potentially causing therapy limiting adverse effects.

Though there may well be a different presumption in cases of non-motor disorders, current knowledge is insufficient for judging the analogous presupposition. Though it is focused on movement disorders, the discussion to follow may illuminate ways of approaching non-motor disorders.

Eisenstein and colleagues studied the anatomical locations of effective DBS in the vicinity of the subthalamic nucleus was studied in 51 patients with bilateral DBS (Eisenstein et al. 2014). A representative example of their analyses, which were complicated, is shown in

Figure 2. The different images demonstrate different statistical manipulations demanded by the data's complexity. A very large area is effective in improving motor symptoms in patients with Parkinson's disease. Indeed, effective areas extend beyond the dorsal-most lateral regions of subthalamic nucleus and even outside the subthalamic nucleus. To the critical question as to which specific region ought to be the target of an anatomy-based automated assistive DBS program came the answer that stimulation nearly everywhere in the volumes analyzed was effective.

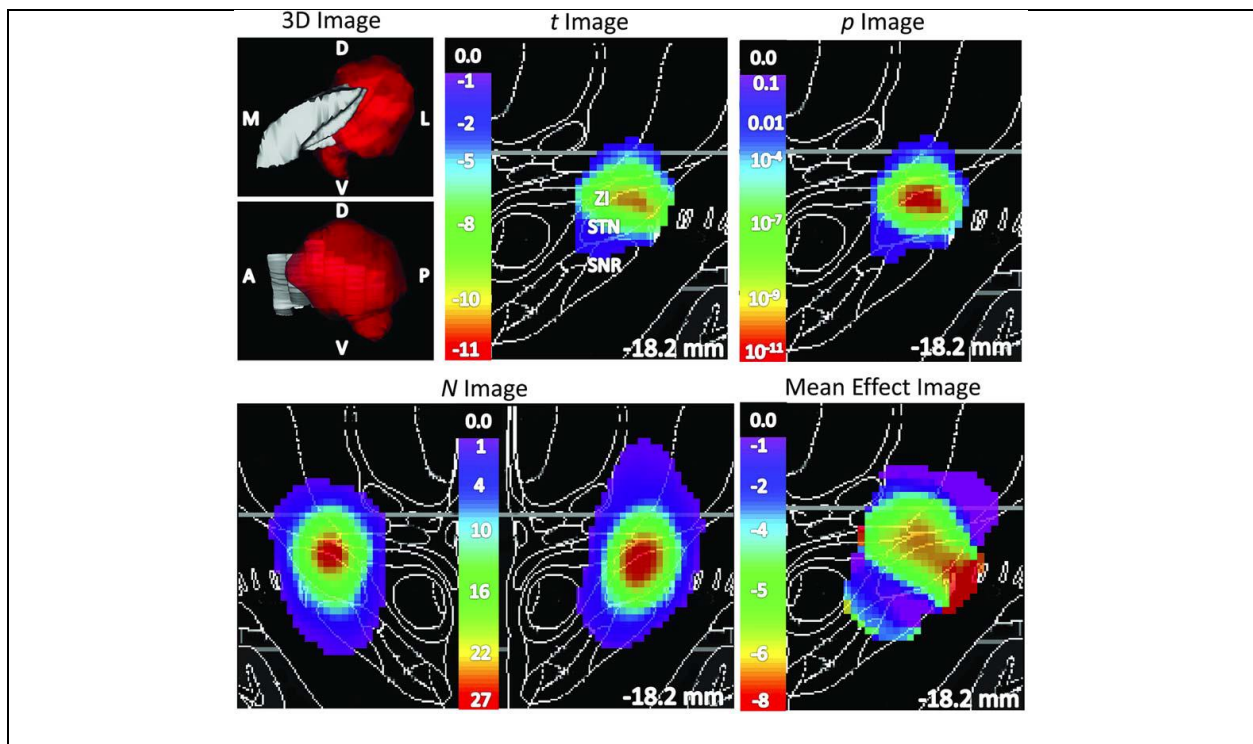


Figure 16.2. Subthalamic nucleus (STN) DBS-induced improvement in right-side Unified Parkinson Disease Rating Scale III motor scores did not differ significantly by stimulation site. The left panel shows three-dimensional (3D) views of the left STN in gray, with color indicating voxels for which $p < 0.05$ by weighted t test (upper image: viewed from a point anterior to STN; lower image: viewed from a point lateral to STN). Statistical images in color,

laid over coronal slices from the atlas, are shown in the top center panel (t image, thresholded at $n \geq 6$), top right panel (p image, thresholded at $p < 0.05$), bottom left panel (N image), and bottom right panel (weighted mean effect image, thresholded at $n \geq 6$). Notably, although the p image includes values as low as 10211, the improvement did not differ significantly by contact location ($p=0.12$ by permutation test; see Results). A = anterior; D = dorsal; L = lateral; M = medial; P = posterior; SNR = substantia nigra; V = ventral; ZI = zona incerta. (verbatim from Eisenstein et al. 2014).

The study by Eisenstein and colleagues does have some important limits beyond those which were actually raised by the authors. One of the problems is that nearly every subject improved — on the total score for the motor examination (part III) of the Unified Parkinson Disease Rating Scales, for example. The relative lack of variability in the outcomes makes it quite difficult to make correlations with the presumed factors underlying the improvement, such as a spatial location of the DBS electrode. Illustrative of this fact is the hypothetical example of three predictors that determine response to a treatment. If the response to treatment was 100% for all subjects, determining which predictor was causal would be impossible. If, however, there was a large range of responses, and, particularly, if the subjects differed in their predictors, then it is much easier to determine the predictor or combination of predictors that determined a response. The more interesting question is: What anatomical differences in active electrode configurations differentiated those who responded from those who did not? Relatively few subjects did not respond. Though this was fortunate for them, it was unfortunate for the investigators who wished to conduct of statistical analyses.

A second problem was that all but two subjects were implanted with the Medtronic 3389 DBS lead, whose 1.5 mm contacts are separated by 0.5 mms. From the perspective of the study, this is problematic, because the volume of effective locations was greater than the contacts' 7.5 mm span. (The more widely spaced DBS lead is preferable for this reason, as well as for its higher available electric charge densities.) The choice of a more narrowly spaced DBS lead was predicated on the assumption that therapeutic effect of DBS in the vicinity of the subthalamic nucleus actually has to do with neurons of the subthalamic nucleus. This presumption may be invalid (*see Chapter 6 – Nervous System Responses to DBS, Montgomery Jr. EB, Intraoperative Neurophysiological Monitoring for Deep Brain Stimulation: Principles, Practice and Cases, Oxford University Press, 2015*). Consequently, it is not entirely clear whether the subthalamic nucleus is indeed the optimal target for any anatomy-based automated assistive DBS programming system.

If it is assumed that the volume of the electrical field has a radius is on the order of 2.5 mm, the volume of the electrical field for the ventral-most contact may overlap with the volume of the electrical field for the dorsal-most contact. In light of this, there may possibly be little effective difference in stimulation among the contacts on a narrowly spaced DBS lead. Then again, differences may exist. To the extent that such differences are possible, careful selection of contacts for efficacy (as distinguished from adverse effects) may be reasonable.

The study by Eisenstein and colleagues is interesting in another respect. They found a difference in the optimal stimulation sites when they analyzed the scores of bradykinesia and tremor. More importantly, these optimal spatial distributions were different. The question for any individual patient is which to target.

Non-anatomy-based automated assistive DBS programming systems

Closed-loop or feedback controlled systems, which are discussed in *Chapter 6 – Nervous System Responses to DBS, Montgomery Jr. EB, Intraoperative Neurophysiological Monitoring for Deep Brain Stimulation: Principles, Practice and Cases, Oxford University Press, 2015*, in the context of knowing when to stimulate, may also be used for determining where and how to stimulate and thus drive automated assistive DBS programming. The critical consideration is whether such approaches will be sufficiently robust, that is, whether their sensitivity and specificity will be sufficiently high in terms of determining when and how to stimulate for therapeutic effect. In other words, when a system is implanted it is important to know ahead of time that the patient being treated stands a reasonably high chance of benefiting. Doing otherwise is unethical. Specificity is more of an issue when it comes to avoiding adverse effects. Importantly, considerations as to what and how to stimulate for symptom resolution must include a definition of efficacy as something other than simply the reciprocal of considerations to prevent adverse effects. For example, a DBS lead placed in such a way as to have an acceptably low probability of adverse effects greatly reduces the demand for specificity as the term is used here. A useful analogy is the use of iodinated table salt for the prevention of goiter. Even though the risk of goiter is low, the use of iodized table salt has such a low probability of adverse effects (though not necessarily zero) that it is administered freely.

Sensors need not detect the cause of the pathophysiological mechanisms of the disorder being treated. The sensors' output simply needs to be highly correlated with a therapeutic effect. An example of this correlation is increased power in the beta frequencies in neuronal activities that are detected by recordings of local field potentials or action potentials. Because many patients with Parkinson's disease do not have it, increased beta power cannot be a sufficient

cause of Parkinsonism. For those patients who do have it, however, localization of beta power could provide important information about stimulation sites. The critical consideration becomes, then, that of prior probabilities of the presence of increased beta power. **Figure 3** shows the power spectral densities (the amount of oscillations at different frequencies) of local field potentials recorded from DBS leads placed in the vicinity of the subthalamic nucleus in patients with Parkinson's disease (Quinn et al. 2015). Two of the 10 subjects did not have appreciable increased power in the beta frequencies. If this result is corroborated, it would mean that any automated assistive DBS programming system that is based on power in beta frequencies would be ineffective in 20% of patients with Parkinson's disease. Whether such a system would be clinically practical depends on positive and negative predictive values, as well as cost in its widest sense, of a failed use the system as weighed against forgoing use of the system that would not have failed.

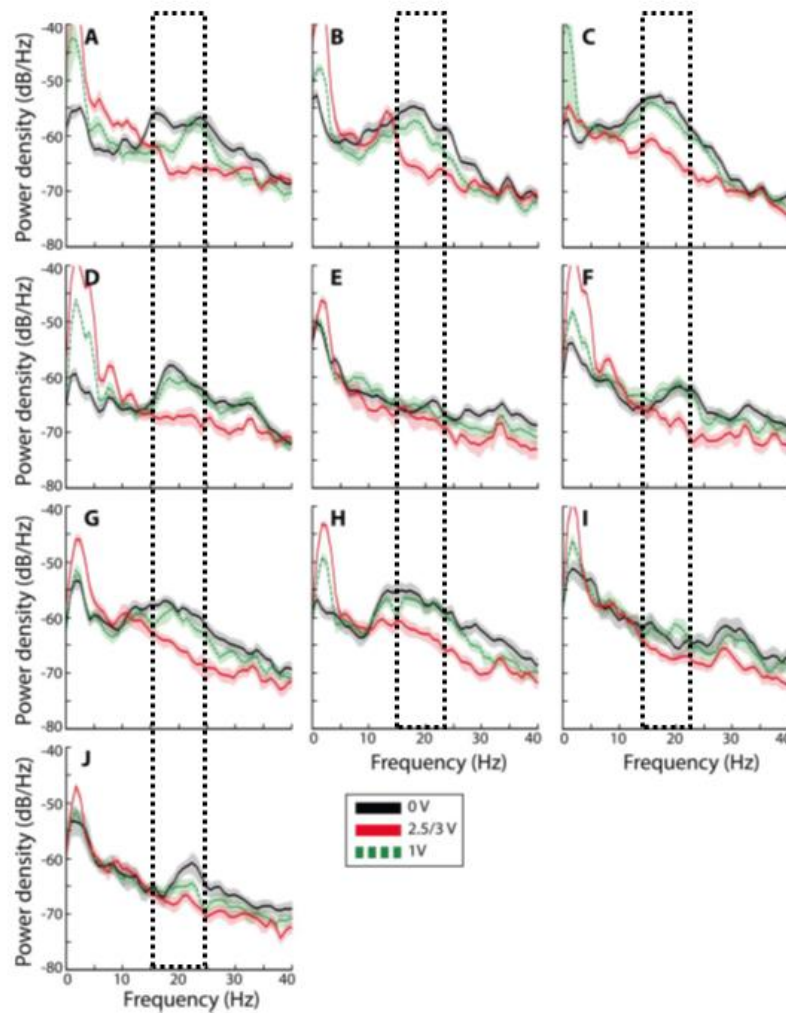


Figure 3. Resting state STN absolute PSDs during randomized presentations of 0 V (baseline), 1 V, and 2.5/3 V 140 Hz STN DBS (n=10, seated). A-D are akinetic rigid (AR); E-J are tremor dominant (TD). Note: STNs D, G, and H received 2.5 V DBS; recording electrode pair 1-3 was used for STNs G and H. STN, subthalamic; PSD, power spectral density; DBS, deep brain stimulation. (Figure legend verbatim and image modified from Quinn et al. 2015). Two of the 10 subjects did not have a discernible peak in the beta power. Use of a beta power to determine electrode configurations, pulse train patterns, or stimulation parameters would consequently be problematic. There are additional subjects in which there may be a peak but

they are broad and not easily discernible from the power at other frequencies (Quinn et al. 2015).

The sensing volume of any sensor must have a resolution that is appropriate to the effectors of such closed loop systems—in this case, the application of negative (cathodal) current to one electrical contact or a combination of contacts on a DBS lead. For example, a sensor whose spatial resolution is on the order of 10 mm prevents one from differentiating among the electrical contacts of most DBS leads whose resolution is on the order of 1.5 mm. This may well be the case for increased beta oscillations in the local field potentials (Verhagen et al. 2015). Under development, however, is technology that could greatly increase the resolution to small submillimeter range despite use of macroelectrode contacts whose size is on the order to millimeters. These technologies, which are applied to electroencephalographic evoked potentials created by appropriate behavioral stimulation, could enable the advent of automated assistive DBS programming that are based on localized generators of submillimeter resolution. Advances in the ability of implanted pulse IPGs to acquire and transmit electrical data may enable such automated systems.

Though it is perhaps more technically demanding of those who are responsible for DBS programming, use of electrical potentials recorded over the scalp during DBS through different electrical contacts may be helpful. For example, it is possible to record evoked potentials from scalp electroencephalographic activity that is driven by the DBS pulse. This has localizing ability. For example, if it can be demonstrated that the distribution of antidromically driven electroencephalographic potentials are able to predict subsequent clinical outcome, then such an automated system may prove practical, technical challenges notwithstanding.

Special cases of long latency to efficacy

A particular challenge to any feedback-based automated assistive DBD programming is the latency between a DBS adjustment and a recognizable effect, which must be brief relative to the time required for optimization of a DBS procedure. For example, one may quickly adjust to, observe, and readjust to tremor responses within seconds of changing DBS parameters. Indeed, when a patient is able to appreciate the feedback measures, whether they are directly tied to clinical benefit, adverse effects or some biomarker thereof, the programming effort may be partly offloaded to her. Spinal cord stimulation for pain is an example. A patient can direct changes in DBS toward paresthesias over the region of pain.

DBS for dystonia and obsessive compulsive disorder stands in contrast to DBS for tremor. With the former, it often takes weeks for patients to experience any improvement, and it take months for them to experience optimal benefit. The nature of the delayed response is critical to DBS programming for dystonia. If the long latencies reflect programming methods, then adequate benefit does not necessarily require such latencies and it may be possible to devised DBS programs that produce optimal benefit within a much shorter period of time. If, however, latencies to benefit require adaptation by the brain—plasticity, for instance—they may be more determined by time rather than DBS programming. The difficulty is that there is currently no way to know whether the former or latter is the case or degree to which the former and latter combine to contribute to the case.

One may tackle the difficulty imposed by long latency of response in DBS by taking one of two different approaches. One approach is to start at maximal DBS stimulation in an attempt to provide optimal clinical benefit in the shortest time. One can conduct, for the purpose of

determining imposed limitations, an initial monopolar survey according adverse effects — tonic contract owing to spread of stimulation to the posterior limb of the internal capsule, for example, or phosphenes, which are suggestive of stimulation of the optic tract — before proceeding to higher values of DBS stimulation settings, such as stimulation current (voltage) and frequencies, in the hope that they are effective and that in a moment the brain will respond appropriately. However, such hopes presume that the relationship between any stimulation parameter and clinical benefit is a monotonically increasing function. There is reason to believe that, at least with respect to DBS frequency in patients with Parkinson’s disease, the relationship is not monotonically increasing (Huang et al. 2014). Thus, it is not necessarily the case that more is better.

Alternatively, one may begin at what may be said to be the average electrode configurations and stimulation parameters, which she may then progressively adjust at two- or three-week intervals. Indeed, a patient or her caregiver may be granted the ability of making those adjustments outside of the clinic. Yet the assumption is again that benefit will be sufficiently manifest by in two or three weeks’ time. In the absence of definitive supporting data, this assumption rests solely on hope. Also, if higher-than-average stimulation settings are required, then the DBS programming process will be drawn out, because each change of DBS parameter may only take place every two to three weeks.

The same issues affect DBS programmers of patients with DBS for obsessive compulsive disorder. There often is a lag phase between onset of DBS and reaching optimal benefit. Many programmers begin with high stimulation settings, as is done in DBS for dystonia (Morishita et al. 2014). Again, it is unclear the extent to which inefficient programming (no fault of the programmer) or brain adaptation explains latencies to benefit.

For dystonia and Obsessive Compulsive Disorder, conditions for which the patient's brain requires time to adapt, automated assistive DBS programming that requires no dynamic feedback—observable changes in the patient's symptoms and signs, for example—may be very helpful. Clearly, anatomy-based systems would be helpful if their positive and negative predictive values were sufficient. Similarly, helpful may be a biomarker that makes its appearance early yet predicts long-term response, such as an evoked potential from scalp electroencephalography or another easily accessible signal generator, or from implanted devices.

The ethics of automated assistive DBS programming systems

Any discussion held 50 years ago on the subject of the ethical responsibility of technology companies to their products' users and customers would have been brief, were it held at all. One major reason for this brevity may have been the technological limitations at the time. Remarkable advances have enabled technology and software particularly to increase in power and capabilities. Thus there is also the potential to do harm in new ways. The critical question is how to assign responsibilities for outcomes, either in the sense of adverse effects as well as the failure to provide benefit that is reasonably expected. The responsibilities that become evident in light of adverse effects are relatively well appreciated. They also enjoy considerable ethical and legal precedent. Adverse effects are governed by the ethical principle of non-maleficence, with maleficence being understood as harm irrespective of blame.

The reasonable responsibilities associated with the failure to provide benefit are much less clear, and they are more problematic from both ethical and legal perspectives. They are governed by the ethical notion of obligation to beneficence and justice (Beauchamp and Childress 2012). The issue of a right to benefit arrives from an interesting direction:

reimbursement by insurers. Reimbursement to physicians, healthcare providers, and their institutions will come to be increasingly tied to achievement of expected outcomes. Expected outcomes according to many metrics are currently the incidence of preventable complications from therapeutic interventions. It is likely these metrics will extend to direct outcomes—degree of blood pressure reduction in patients with hypertension or glucose control in diabetics, for example.

One of the greatest conceptual challenges is how to establish the responsibilities for physicians, healthcare professionals, and industry in the case of an individual patient. This is clearly problematic when it involves extrapolation from sample summary statistics to an individual patient. Yet, this difficulty does not obviate the responsibilities of physicians, healthcare professionals, and industry to the individual patient. One approach to improving the probability of ensuring expected benefit is that of addressing institutional or systems problems that risk failure to provide expected benefit, even if the outcome cannot be directly attributed from an individual patient. What follows is a discussion of the relationship between technology developers, such as those who create automated assistive DBS programming systems, and those who use the systems in caring for patients.

There are many ways in which technology can harm. Most of these may be anticipated by engineers, possibly compensated for, and considered in advance of the some anticipated use. However, there are ways to harm that an engineer may not readily recognize, perhaps because she is unused to considering ethical issues.

As technology has changed and expanded, ethics has also. Indeed, rapid advances in technology have created the need for changes in ethics. Prior to the artificial ventilation and

cardiovascular support, the ethical question of when a person is dead was uncontroversial. It is a very different matter today — indeed, one which keeps ethicists quite busy.

With expanding technology, the occasions for problems associated with technology to bump up against the legal system has increased. Discussions of legal issues below occur strictly in the context of ethics in its broadest sense: how people interact with each other. In many ways, legal decisions related to the use of technology have clarified and changed many of the ethical issues related to the practice of medicine in the face of technology.

Among the changes in ethical principles are seen in the now outmoded notion that a physician is a captain of a ship. This is particularly true in such situations in which a physician or surgeon is not the master of all the technology in use and must therefore depend on non-physicians. Also, there are changes in responsibility, and thus culpability, on such occasions when technology supplements physicians and healthcare professionals rather than augmenting them. The notion of a physician as being a captain of a ship is that she has the final say in the provision of care. As such, she is solely responsible. This notion has not changed; it has simply been modified. For example, most states do not allow non-physician healthcare professionals to act independent of physicians. Those states that do allow non-physician healthcare professionals to act independently mostly do as an expedient in the face of physician shortage.

Unfortunately, the notion of the physician as a captain of a ship has grown to include culpability. The notion that a physician bears full responsibility entails some sort of immunity to those who work subordinately to her. A number of court cases have shown that this is not the case. There have been cases in which anesthesiologists have sought immunity from suits that rest on the claim that surgeons are in charge. Ruling against anesthesiologists, courts have argued

that unique knowledge and skills, which surgeons do not replicate, place a unique responsibility on anesthesiologists and thus renders them culpable.

What it means to be a non-physician healthcare professional subordinate to a physician or surgeon has been clarified. Some believe that a subordinate not employed by the physician changes the culpability. This is a defense that has been used by subordinates as well as physicians. For example, a nurse in an operating room commits malpractice. A surgeon claims immunity, citing as the basis the fact that the nurse is an employee of the hospital. The courts rule against the surgeon, stating that she was in a position to control the nurse, which thereby makes the nurse subordinate to her. The surgeon is culpable by the legal and ethical principle of vicarious liability.

Although this author is unaware of any case precedent in which a technology company was held vicariously liable on the basis of malfeasance on the user's part, could a technology company be held liable for misuse of its product by a physician? The question goes beyond more traditional notions of product liability, which typically pertain to manufacturing defects. The admittedly extreme example of litigation against gun manufactures illuminates many of the ethical and legal issues that bear on the question (Lytton <https://www.press.umich.edu/pdf/0472115103-intro.pdf>). Four arguments are typical: (1) malfunction is typically consequent to the manufacturing; (2) negligence by the manufacturer and seller in exercise proper care in selecting those that would use the device; (3) manufacturers and sellers are liable for adverse consequences for failing to take reasonable precautions in the design, manufacturing, and selling of the device (Interestingly, this subject came up in the hearing by the external review for the FDA consideration of approving DBS of the subthalamic nucleus and globus pallidus interna for patients with Parkinson's disease. The issue was whether

the manufacturer would be required to hold training courses for DBS lead implantation surgery - the issue of postoperative management did not come up - and thereby, exert some level of control over who could use the DBS devices; and (4) sellers are liable when more restrictive sales would have prevented harm.

In litigation against the gun industry, only the first argument has enjoyed any success. The implication is that the onus is on the person purchasing and using the gun. In the case of automated assistive DBS programming systems, the extrapolation is that the onus is on the user. However, use of guns is quite different from use of DBS for many reasons beyond that of the gun industry's lobbying efforts. Nevertheless, the arguments offered in gun industry litigation have important implications for automated assistive DBS programming systems. An important difference is the complexity of the device. Guns are relatively simple devices to operate, notwithstanding the fact that the consequences of its operation are sometimes unpredictable. This relative simplicity makes it at least appear reasonable to assign of the responsibility to a gun user. The point that is taken as credible is that the user should know better. The question in DBS is: Can a user of automated assistive DBS programming systems know better in such instances as an outcome is adverse or a reasonably expected benefit fail to occur?

Responsibility for the technology is clear in certain circumstances where governmental agencies require a manufacturer to provide specific training. In those cases, the technology companies transfer to a surgeon or physician the necessary and sufficient knowledge to independently judge the outcomes of the technology's use. Assuming that the training was appropriate, the presumption is that the technology company, which is assumed to have delivered appropriate instruction, is not culpable for its product's misuse by a surgeon or physician.

However, this approach will be increasingly problematic as technology supplants rather than augments various roles previously performed by a physician and surgeon.

One necessary distinction is between “supplant” and “augment.” By “augment” is meant the function of technology as rendering a physician’s or surgeon’s actions more efficient. Implied is retention on the part of physicians and surgeons the fundamental knowledge and skills inherent in the technology. It is important to note that augmentation only makes those who already provide care do so more efficiently, thereby improving reimbursement and extending care to more patients. (This assumes that the rate patients who enter the queue is greater than the rate of patients who currently receive care.) By “supplant” is meant the offload of various analyses from physicians and surgeons to the technology in question. In this physicians and surgeons no longer have complete knowledge. Rather, they depend on the technology company in the same way as they depend on anesthesiologists.

There is the tendency or temptation to blur the distinction between augmenting physicians and surgeons and supplanting them. No amount of warning or signing off of a technology’s actions by a physician or surgeon change the fact that the technology is supplanting them, with the associated culpability this entails, rather than simply augmenting them. Illustrative of this fact is the effort to automate image-based DBS lead targeting. Certainly, there have been remarkable technical advances, ranging from increasing magnetic field strength in MRI scans, improved identification of regions of interest, novel sequences in MRIs, and nonlinear morphing to a common anatomical atlas. Most of these involve a number of specific steps or processes, each of which may lie beyond the knowledge or experience of most surgeons. Asking a physician or surgeon to sign off on each of these steps is disingenuous unless the technology company can be reasonably sure that she has sufficient knowledge for determining

whether the processes ran correctly. Though to sign off without having such knowledge is to act unethically, how this affects the manufacturer's culpability remains unclear.

Technology companies may have naïve notion of their products' benefits. "It's simply information," they may tend to reply: "What could be wrong with giving the physician or surgeon more information." The fact is that information has a drawback or risk. Illustrative of this fact is the controversy over mammograms in women under the age of 50 who have neither a family history of breast cancer nor BRCA 1 or 2 genes. Prior probabilities suggest that the risk in younger women is relatively low and the risk of false positives relatively high. This could lead to further and more invasive diagnostic measures, which may ultimately result in a greater harm than benefit. The recommendation of professional organizations is to forgo obtaining mammograms for such women. The argument may be made that a mammogram is simply information that a physician is free to consider or ignore. However, such an argument would be disingenuous as it is highly unlikely that any physician would ignore an abnormality found on a mammogram.

Suggesting that the decision to use or ignore information is the prerogative of the physician or surgeon and thus, provides immunity to the technology company relative to the adverse consequences of false positives or negatives becomes more problematic in the context of advertising by technology companies. Certainly, such advertising is not in itself ethically suspect. But as the goal of advertising to affect the mode of thinking of physicians and surgeons, the technology company's becomes more clearly seen as influencing the physicians' and surgeons' choices.

References:

1. Beauchamp and Childress, *Principles of Biomedical Ethics*, Oxford University Press 2012.

2. Butson CR, McIntyre CC, The Use of Stimulation Field Models for Deep Brain Stimulation Programming, *Brain Stimulation* 8 (2015) 974e991.
3. Eisenstein SA, Koller JM, Black KD, Campbell MC, Lugar HM, Ushe M, Tabbal SD, Karimi M, Hershey T, Perlmutter JS, Black KJ. Functional anatomy of subthalamic nucleus stimulation in Parkinson disease. *Ann Neurol*. 2014 Aug;76(2):279-95. doi: 10.1002/ana.24204. Epub 2014 Jul 2. PubMed PMID: 24953991; PubMed Central PMCID: PMC4172323.
4. Frankemolle AM, Wu J, Noecker AM, Voelcker-Rehage C, Ho JC, Vitek JL, McIntyre CC, Alberts JL. Reversing cognitive-motor impairments in Parkinson's disease patients using a computational modelling approach to deep brain stimulation programming. *Brain*. 2010 Mar;133(Pt 3):746-61. doi: 10.1093/brain/awp315. Epub 2010 Jan 7. PubMed PMID: 20061324; PubMed Central PMCID: PMC2842509.
5. Huang H, Watts RL, Montgomery EB Jr. Effects of deep brain stimulation frequency on bradykinesia of Parkinson's disease. *Mov Disord*. 2014 Feb;29(2):203-6. doi: 10.1002/mds.25773. Epub 2014 Jan 2. PubMed PMID: 24395752.
6. Montgomery Jr., E.B. *Intraoperative Neurophysiological Monitoring for Deep Brain Stimulation: Principles, Practice and Cases*, Oxford University Press, 2015).
7. Morishita T, Fayad SM, Goodman WK, Foote KD, Chen D, Peace DA, Rhoton AL Jr, Okun MS. Surgical neuroanatomy and programming in deep brain stimulation for obsessive compulsive disorder. *Neuromodulation*. 2014 Jun;17(4):312-9; discussion 319. doi: 10.1111/ner.12141. Epub 2013 Dec 17. Review. PubMed PMID: 24345303; PubMed Central PMCID: PMC4061279.

8. Pourfar MH, Mogilner AY, Farris S, Giroux M, Gillego M, Zhao Y, Blum D, Bokil H, Pierre MC. Model-Based Deep Brain Stimulation Programming for Parkinson's Disease: The GUIDE Pilot Study. *Stereotact Funct Neurosurg.* 2015;93(4):231-9. doi: 10.1159/000375172. Epub 2015 May 14. PubMed PMID: 25998447.
9. Quinn EJ, Blumenfeld Z, Velisar A, Koop MM, Shreve LA, Trager MH, Hill BC, Kilbane C, Henderson JM, Brontë-Stewart H. Beta oscillations in freely moving Parkinson's subjects are attenuated during deep brain stimulation. *Mov Disord.* 2015 Nov;30(13):1750-8. doi: 10.1002/mds.26376. Epub 2015 Sep 11. PubMed PMID: 26360123.
10. Verhagen R, Zwartjes DG, Heida T, Wiegers EC, Contarino MF, de Bie RM, van den Munckhof P, Schuurman PR, Veltink PH, Bour LJ. Advanced target identification in STN-DBS with beta power of combined local field potentials and spiking activity. *J Neurosci Methods.* 2015 Sep 30;253:116-25. doi: 10.1016/j.jneumeth.2015.06.006. Epub 2015 Jun 12. PubMed PMID: 26079495.