Designing a Course in Statistics for a Learning Health Systems Training Program

Gregory P. Samsa[1], Thomas W. LeBlanc, Aimee Zaas,
Lynn Howie and Amy P. Abernethy
Duke University, Durham, NC 27705

Abstract

The core pedagogic problem considered here is how to effectively teach statistics to physicians who are engaged in a "learning health system" (LHS). This is a special case of a broader issue – namely, how to effectively teach statistics to academic physicians for whom research – and thus statistics – is a requirement for professional advancement. A distinguishing feature of these students is the degree of imbalance between high levels of scientific maturity and relatively low levels of training in mathematics and computer programming. Using a constructivist framework, the curriculum is organized around a set of model cases and an explicit conceptual map of how those cases are related. When teaching LHS physicians, the model cases should be different from those used to teach statisticians: they must be simple, clinically relevant, and developed by example. To create such cases, the discipline of statistics must not only be deconstructed but must also then be reconstructed in a framework that is accessible to its students. This is a principle that should also be generally applicable to teaching statistics to non-statisticians from other disciplines.

Keywords: Data science, deconstructing the discipline, learning health care, statistical education.

Narrowly constructed, the core pedagogic problem considered here is how to effectively teach statistics to physicians that are engaged in "learning health systems" (LHS). Learning health systems are defined in many ways, our working definition being: "A LHS leverages new developments in health information technology and a growing health data infrastructure to access and apply evidence in real time, while simultaneously drawing knowledge from real-world patient care delivery to promote health system change and innovation that is rooted in clinical data". (Greene, Reid, & Larson, 2012) This is a special case of a broader issue: namely, how to effectively teach statistics to academic physicians for whom statistics – whether used to assist in research, clinical care, or both – is a requirement for professional advancement.

---

[1] Corresponding author's email: Samsa001@mc.duke.edu.

Presently, physicians at Duke who wish to obtain formal statistical training do so under one of two models: either a physician-focused model or a statistician-focused model. To illustrate the distinction between these two models of instruction, those physicians that wish to obtain masters-level training in statistics can either enroll in the Masters of Biostatistics program or the Clinical Research Training Program. The Masters of Biostatistics program is designed to train statisticians – for example, the "theory" courses are taught in the languages of calculus and linear algebra. The instruction is on the statistician's terms, and the challenge for the occasional physician that enrolls in this program is to get fully up to speed on calculus training that happened at some time in the past.

At the other end of the spectrum is the Clinical Research Training Program. Although students are provided with tools to perform data analyses, the overall emphasis is on interpretation, and the goal is to train investigators to become sophisticated consumers of statistics who can effectively collaborate with statisticians. Mathematical formalism is discouraged, and the instruction is designed to be on the physician's terms. A description of the active-learning-based implementation of the advanced modeling course is provided elsewhere (Samsa, Thomas, Lee, & Neal, 2012).

Both the Masters of Biostatistics and the Clinical Research Training Program models are "extreme" in the sense that their instruction is tightly linked with a single disciplinary perspective. Despite their positive aspects, the single-minded disciplinary perspectives of these programs have raised concerns. Some physicians are concerned that the Masters of Biostatistics program has too much mathematics – especially, that the investment of time required to become proficient in calculus and linear algebra exceeds the benefits of the resulting statistical training. Some statisticians are concerned that the Clinical Research Training Program has too little mathematics – especially, that those students who are unfamiliar with the mathematical derivations of statistical techniques won't understand those techniques in sufficient depth to use them appropriately (or even recognize that they shouldn't be doing the statistical work themselves but rather should call in help and then refocus energy on interpretation).

Context

Presumably, what is needed is something in between the above extremes. The specific context is a new LHS training program. LHS has been variously described (e.g., Etheredge, 2007; Greene et al., 2012), but one way to think of it is as an extension of the traditional framework of evidence-based medicine. In evidence-based medicine, physicians are trained to critically review the medical literature in order to determine how to treat an individual patient in a specific clinical context. LHS, at its most basic, extends evidence-based medicine by empowering physicians to review their own data to better understand how they are actually practicing medicine, and to use the insights gained thereby to improve the processes and outcomes of patient care. Understanding of data can expand beyond one's own practice to understanding aggregate practice at one's own site, or across larger systems of care. The idea that LHS practitioners use aggregate clinical information as evidence was succinctly stated in one of the interviews used to guide the vision-setting

process for the LHS training program at Duke: what most distinguishes LHS practitioners from other physicians is that they "get" data.

To help train physicians for the learning health system environment, we established the first Learning Health Systems Training Program (LHSTP). The LHSTP will include core statistical training early in the curriculum, followed by group projects that are similar in appearance to quality improvement initiatives. As an example of a typical project, a LHS trainee might be concerned about the quality of anticoagulation management in an outpatient clinic. The electronic medical record could be queried for patients that are receiving long-term warfarin therapy, the resulting dataset reorganized to estimate the average time in target therapeutic range, and this average compared with national benchmarks. If the time in target therapeutic range is below the desired level the process of care would be redesigned – for example, with eligible patients switched to home-based monitoring – and time in target therapeutic range re-estimated. The final step would be to compare the costs of revising the process of care with the benefits, these latter benefits being estimated through a model that links quality of anticoagulation management with the expected number of clinical events. Beyond traditional so-called "quality improvement" cycles, a LHS system does this in a more seamless and expedited fashion, and as a more active part of routine patients care.

Apart from introductory statistical instruction, additional statistical training will take place on an as-needed basis. Using the anticoagulation example, the advantages and disadvantages of pre-post designs would be discussed, as would various approaches to simplifying the longitudinal patient-level data about anticoagulation levels into (ideally) single summary measures per patient. This, in turn, would provide a practical link between the case example and the statistical thinking underpinning a variety of study designs and analytic approaches. Moreover, careful review of available data and its attributes such as missingness and reliability would serve to reinforce understanding of "what are data" and "what can I do in my own practice to improve the quality of data collected"?

At the time of this writing, the first cadre of LHS trainees is beginning their work. Curriculum development will be cooperative – in particular, the trainees will have a significant voice in the content and delivery of the statistical curriculum. Accordingly, what are described here are not the details of a finished curriculum. Instead, it is a preliminary answer not to the question of "what statistical content should be taught" but, instead, to the question of "how statistical content should be delivered".

Conceptual framework

From the perspective of statistical instruction LHS practitioners must among others be able to:

- Pose clinical questions in a fashion that is amenable to subsequent statistical analysis.
- Select an appropriate study design, which in this context often means to design a sound database query.

- Design a statistical analysis plan.
- Understand how to implement a statistical analysis plan, which in this context often means to understand the flow of data at a sufficient level of detail to be able to diagram those data flows.
- Understand data elements and their attributes (e.g., scale of measurement such as categorical, ordinal and continuous).
- Understand issues that affect data quality.
- Understand how to interpret database queries that utilize statistical modeling – including issues of data quality, validation and limits to proper interpretation.
- Given the results of a statistical analysis, suggest a proper course of action.

One of the distinguishing features of this application is that curriculum development must take into account the unbalanced nature of its students. Among their many positive attributes, the LHS trainees are intelligent, motivated and scientifically sophisticated. On the other hand, relative to other students in graduate-level statistics courses they tend to be quite weak in mathematics and computer programming – so much so that the design of a traditional course that relies on calculus-based derivations and facility with data analysis simply won't work for them. The principles of sound educational pedagogy apply to "unbalanced" students no less than they do for "balanced" ones; nevertheless, the general disgruntlement with statistical training among this community of learners suggests that statistical training for academic physicians is a curriculum development task that is unique and particularly challenging.

In designing the statistical instruction, we applied a constructivist framework similar to that of Fields (Fields, Baxter, & Seawright, 2006). More specifically, we assumed that our physician students will be constructing their understanding of statistics around (a) model cases; (b) a conceptual map of how the principles and techniques illustrated by those model cases fit together; and (c) analogies to assist in applying the model cases and conceptual map to actual problems (Hofstadter & Sander, 2013).

To illustrate the use of conceptual maps in statistics, one way that statisticians typically conceptualize modeling uses scale of measurement. For example, models that use time-to-event as an outcome variable fall within the category of survival analysis, and models that have a dichotomous outcome (e.g., good versus poor) as an outcome variable fall within the category of logistic regression. Models that have a continuous outcome variable and a 2-category predictor fall within the general category of linear models, and the specific category of the t-test (which, in turn, is a special case of the 1-way analysis of variance). Once the appropriate model is selected, other conceptual maps are also utilized. For example, a conceptual map of modeling strategy would include the distinction between an adjustment application (i.e., which answers the question: controlling for Y, does X predict Z?) and variable selection (i.e., which answers the question: which of the set of variables X and Y predicts Z?).

Conceptual maps use model cases as building blocks. Using the t-test as an example, the standard analysis can be reduced to a protocol.

- Create box-plots for both groups to visualize the data.
- Verify that the sample means for both groups are a reasonable summary of their central tendencies.
- Calculate the means and standard errors for both groups.
- Perform a t-test and use the resulting p-value to assess statistical significance.
- Calculate the difference between the group means and generate a confidence interval for that difference.
- Assess the values within the confidence interval for clinical significance.

The model case is an application of this protocol to a memorable problem, and would also include the computer code required to perform the analysis.

When performing a t-test in practice, the statistician would compare the current problem with the model case. For example, if it appeared that the sample means in question were not representative summaries of central tendency the standard protocol would have to be modified. In this case, the statistician might instead apply a non-parametric test by first transforming the data into ranks and then applying a t-test to the ranked data. The relevant analogy (which turns out to be sound in the case of the t-test) is that non-parametric tests are often equivalent to transforming the data into ranks and then proceeding as usual.

As an example of a higher-level conceptual map, the design of a LHS project typically involves a series of steps that culminate in a statistical analysis plan. These steps can be understood as providing answers to the following increasingly specific questions:

- In medical terms, what is study question?
- How can the medical question be translated into study aims?
- What study design can best achieve the study aims?
- Given the study design, how can the study aims become translated into statistical hypotheses amenable to analysis?
- How can the statistical hypotheses be translated into a statistical analysis plan?

At each of these steps, the project design benefits from an explicit review step – for example, to assess whether the study aims are an adequate representation of the essence of the clinical question, to assess how well the study design can meet its aims, to assess how closely the statistical hypotheses match the underlying medical questions, and to assess how well the statistical analysis plan will test the statistical hypotheses. Such a review is more likely to occur if the LHS practitioner is aware of this higher-level conceptual map.

Application of the conceptual framework

Consistent with the constructivist framework, the curriculum design task was conceptualized as involving the identification of those elements of applied statistics that are relevant to LHS, the deconstruction of those elements into model cases, and the reconstruction of those model cases into an explicit conceptual map. Another element of the curriculum design task was the identification of materials that were consistent with this approach.

Fortunately, van Belle's book on statistical rules of thumb (van Belle, 2002) provided an illustration of one possible way to implement the above teaching strategy. The book is organized around rules of thumb, which roughly correspond to model cases. The order of presentation is (a) an introduction; (b) a statement of the rule of thumb; (c) an illustration of the rule; (d) the basis for the rule; and (e) discussion and extensions. Oriented toward statisticians, the text pertaining to the basis of the rule is often explained algebraically. Moreover, the conceptual map that links the rules of thumb is implicit, consisting of the "tacit knowledge" possessed by members of the statistical community. We propose to use the same general structure as van Belle, but with a somewhat different implementation of steps "c" and "d".

Using van Belle's structure as the basis for the model cases, two design questions appeared to be fundamental:

- How (if at all) should the presentation of the model cases differ when the target audience changes from statisticians to physicians?
- Recognizing that the conceptual maps of statistics used by physicians are likely to be rudimentary at best, how can physician students be encouraged to develop more sophisticated conceptual maps of statistics?

The response to these design questions is discussed below.

Response 1: The model cases should be different.

Physicians should be taught statistics using model cases that are simple, clinically relevant, and developed by example. Regarding simplicity, a principle that can be illustrated to a statistician in a single model case might require multiple sub-cases when designed for a physician. (The reason is that the physician has less tacit knowledge of statistics upon which to rely.) Regarding clinical relevance, the ideal is for model cases to build upon one another and use examples that are clinically interpretable. (For example, as far as a statistician is concerned the model case for a t-test is based on scale of measurement – any 2-category predictor and continuously-scaled response will do. A physician, on the other hand, prefers the example to be realistic. Clinical relevance isn't necessarily critical to simple examples that illustrate the technical mechanics of the computations involved with a t-test, but attains greater importance when developing memorable model cases to which the physician can later refer.) Regarding development, the basis for explaining the model cases should be example rather than mathematical derivation. The model cases should include hands-on interaction with the data, including interpretation.

This is not only a good general principle to follow in any event, but is consistent with the "see one, do one, teach one" model of medical education with which physicians are already familiar.

Appendices 1 and 2 illustrate what is envisioned, and have been used successfully in teaching physicians in other contexts. Of particular note are (a) the step-by-step development of the cases; and (b) the explicit translation of the principle illustrated by the example into words.

Response 2: Students should be encouraged to describe and talk through their conceptual maps.

Vocalizing, drawing, or otherwise making explicit their conceptual maps can be considered part of the "teach one" component of the usual model of medical education. In particular, if it is discovered that a student's conceptual map of statistics is unsophisticated or inaccurate then this deficiency – now explicitly identified -- can be addressed. During this process, describing our own conceptual maps of statistics (i.e., "deconstructing the discipline") can be extraordinarily helpful (Middendorf & Pace, 2004; Diaz, Middendorf, Pace, & Shopkow, 2008).

As they "do one", it is helpful to ask students to talk through the task. In most cases, technical proficiency (e.g., the ability to follow a standard data analytic protocol) is achieved before higher-level proficiency (e.g., the ability to select a data analysis, the ability to interpret results, the ability to determine which features of the current problem differ from the model case).

A particularly natural application of explicit conceptual maps pertains to the treatment of data. This treatment should be organized around a statistical analysis plan (in essence, the "analyst's story") that diagrams the tables, graphs and other elements of the planned analyses. This analysis plan is then compared with its data requirements – in other words, the structure of the datasets that the plan requires. The LHS practitioner would then work backwards from the data requirements to more detailed information about the data source and the data elements. This is the point where the LHS practitioner would potentially discover mismatches between the actual and desired structure of the data elements – for example, when the analysis plan requires a data element for the presence of absence of chronic atrial fibrillation but the available datasets only contain this information within free text fields. This is also the place where a review of likely data quality would occur.

Discussion

We have attempted to delineate some general principles for teaching statistics to physicians that will be practicing within learning health systems. A central notion is that of first deconstructing the discipline of statistics and then reconstructing it into an annotated set of model cases, with these cases being linked through an explicit conceptual map.

Moreover, this reconstruction is designed to take into account the particular characteristics of its target audience.

The LHSTP is new, and thus formal evaluation data are not yet available. However, some factors do lend encouragement. First, the course development has been theoretically driven, and is consistent with over two decades of experience with a target audience that has a unique set of characteristics (Samsa et al., 2012). Second, the structure of the model cases illustrated in appendices 1 and 2 has proven to be successful in the past. Third, this approach was pilot tested in a short course in statistics for biomedical researchers (principally biologists but also including physicians). The notion of using model cases and an explicit conceptual map was discussed at the start of the course, and was received enthusiastically. Indeed, observation suggests that one of the things that non-statistician students of statistics particularly crave is information about how everything fits together – sometimes stated as "I know how to proceed if you tell me what to do but am not confident that I can decide what to do" – which in essence is a plea for a conceptual map.

Fourth, the proposed methods for statistical instruction are fundamentally consistent with guidelines endorsed by the American Statistical Association (Aliaga *et al*, 2010). These guidelines, originally intended for first undergraduate courses in statistics but applicable more generally, include six overall recommendations. These recommendations, and examples of how they are implemented within the statistical curriculum, are provided in Table 1 below.

Finally, when presented with a draft of this document for comment the LHS trainees and faculty were supportive of the approach proposed here. The current version reflects their comments.

In the biomedical context (among others) statistics is usually practiced within interdisciplinary teams. Non-statistician investigators typically need to (a) be able to perform some basic statistical analyses; and (b) in more complex applications, interpret statistical results and otherwise collaborate with statisticians. These investigators do not need to be exposed to an entire statistics curriculum, nor could they necessarily tolerate one – in other words, they cannot be taught statistics in the same fashion as were their instructors. LHS trainees are an example of such investigators, but are just one example out of many.

The purpose in summarizing the LHSTP curriculum development efforts to date is to engage the statistical, medical and educational communities in a discussion of how to more effectively teach statistics to physicians. The urgency for doing so, even in the absence of a formal evaluation of the LHSTP, is driven by the premise that something fundamental is wrong with the way that statistics is usually taught to physicians (and also to others outside the discipline of statistics). Although a modest-sized literature exists on teaching statistics outside the discipline and a smaller literature exists on teaching statistics to medical students and physicians (e.g., Freeman, Collier, Staniforth, & Smith, 2008), physicians consistently express dissatisfaction with their statistical training. For example, the orientation session of the Clinical Research Training Program's advanced modeling class

Table 1: GUISE Recommendations and Their Implementation.

Recommendation	Implementation
Emphasize statistical literacy and develop statistical thinking (i.e., understanding the need for data, the importance of data production, the omnipresence of variability, and the quantification and explanation of variability).	Data quality, and also developing a detailed understanding of how the elements within databases are derived, are points of particular emphasis.
Use real data	Just in time statistical instruction will be based upon LHC trainee projects using real data.
Stress conceptual understanding rather than mere knowledge of procedures.	This is facilitated by using explicit conceptual maps, and also by having the LHC trainees translate statistical principles into their own words.
Foster active learning in the classroom.	Among others, the project-based orientation encourages a hands-on approach.
Use technology for developing conceptual understanding and analyzing data.	Data will typically be obtained from electronic medical records and analyzed using statistical software such as R.
Use assessments to improve and evaluate student learning.	Making conceptual maps explicit facilitates assessment of conceptual understanding.

begins with an assessment of its students' working knowledge of statistics and their confidence in that knowledge. These assessments consistently show that physicians are usually taught statistics in a highly protocol-based fashion, have a relatively low level of working knowledge of statistics, and an even lower level of confidence in their ability to apply that knowledge.

We believe that the fundamental error is that when teaching physicians statisticians so often "do the same thing, only less". For example, when designing a course for physicians, a statistician instructor might take a traditional 2-semester course in applied data analysis for statisticians, remove the content from the second semester, remove some of the mathematical proofs, change the examples to biomedical ones, and assume that the result will be effective. In essence, the course retains the same construction of statistics as the original, and has simply presented to the student selected elements of that construction. Instead, what is needed is a more extensive change: namely, a reconstruction that is tailored to the needs of the target audience. This is a principle that should be generally applicable to teaching statistics to non-statisticians from other disciplines.

Although speculative, we anticipate that even in another context – for example, bench scientists – what would stay the same is the benefit of having students describe and talk through their conceptual maps of statistics. What would change is not the principle that

model cases should be adapted to the needs of the target audience, but instead would be the choice and presentation of those model cases. Model cases need to encapsulate the correct statistical content, include examples that resonate with their target audience, and use language that is familiar to that audience. Development of such cases only likely to occur when the statistician: (a) is familiar with the field under study; and (b) is willing to take the initiative to bridge the gap across disciplines.

References

Aliaga, M., Cobb, G., Cuff, C., Garfield, J., Gould, R., Lock, R., Moore, T., Rossman, A., Stephenson, B., Utts, J., Velleman, P., & Witmer, J. (2010). *GAISE College Report*. American Statistical Association.

Diaz, A., Middendorf, J., Pace, D., & Shopkow, L. (2008). The History Learning Project: A department "decodes its students". *Journal of American History*, 94, 1211-1224.

Etheredge, L. M. (2007). Rapid learning: a new approach to diffusing research into practice. *Health Affairs,* 26(2), 107-118.

Fields, P. J., Baxter, A., & Seawright, L. (2006). A constructivist approach to course design in a graduate statistics course. In Rossman, A., & Chance, B. (Eds.) *Proceedings of the Seventh International Conference On Teaching Statistics (ICOTS-7),* Salvador, Brazil. Voorburg, The Netherlands: International Statistical Institute.

Freeman, J. V., Collier, S., Staniforth, D., & Smith, K. J. (2008). Innovations in curriculum design: a multi-disciplinary approach to teaching statistics to undergraduate medical students. *BMC Medical Education,* 8:28, doi:10.1186/1472-6920-8-29.

Greene, S. M., Reid, R. J., & Larson, E. B. (2012). Implementing the learning health system: from concept to action. *Annals of Internal Medicine,* 157, 207-210.

Hofstadter, D., & Sander, E. (2013). Surfaces and Essences: Analogy as the Fuel and Fire of Thinking. Basic Books, New York.

Middenddorf, J. & Pace, D. (2004). "Decoding the disciplines: a model for helping students learn disciplinary ways of thinking. *New directions for teaching and learning, 1-12*.

Samsa, G. P., Thomas, L., Lee, L. S., & Neal, E. M. (2012). An active learning approach to teach advanced multi-predictor modeling to clinicians. *Journal of Statistics Education,* 20(1), 1-31.

van Belle, E. (2002). Statistical Rules of Thumb. Wiley Interscience, New York.

Appendix 1 – Illustration of a Model Case: Multiple Testing

Verbal descriptions of the principle:

- The more you look the more you will see, even if nothing is going on.
- The more statistical tests you perform the more statistically significant results you will observe, even if none of them are real.
- If large numbers of tests are performed, be suspicious of statistically significant results.

Model case: (assumes that the tests are independent and p=.05 as a benchmark for declaring statistical significance)

# tests	Probability that all tests are non-significant	Probability that at least one test is significant	Expected number of significant tests
1	.95	.05	.05
2	.90	.10	.10
3	.86	.14	.15
10	.60	.40	.50
20	.36	.64	1
100	.005	.995	5

Derivation:

- The probability that one test is significant =.05 (i.e., because this is the type 1 error rate of the test). *Note: In practice, this isn't necessarily intuitive to the student and requires review of the structure of a hypothesis test.*
- Thus, the probability that one test is non-significant is .95.
- Thus, the probability that two tests are both non-significant is (.95)(.95).
- Thus, the probability that three tests are all non-significant is (.95)(.95)(.95).
- Thus, the probability that K tests are all non-significant is (.95)**K.
- Moreover, the probability that at least one test is significant is one minus the probability that all tests are non-significant.
- Finally, the expected number of significant tests is (.05)*K.

Appendix 2 – Illustration of a Model Case: Testing for a Rare Disease

Verbal description of the principle:

- If the disease is rare, positive test results probably don't indicate disease.
- The operating characteristics of sensitivity and specificity aren't enough to under-stand the performance of a diagnostic test – prevalence also matters.
- For rare diseases, even if the specificity is high large numbers of patients without disease will generate large numbers of false positives – even if the test has perfect sensitivity these false positives will overwhelm the small number of patients with disease.

Note: This illustration assumes that the student is familiar with sensitivity and specificity, and also with the structure of a 2x2 table of disease status versus test result.

Model case with derivation:

Assume: sensitivity=99%, specificity=99%, prevalence=0.01%,
population size=1,000,000.

Population size=1,000,000

	Disease present	Disease absent	Total
Test positive			
Test negative			
total			**1,000,000**

Population size=prevalence=0.01%

	Disease present	Disease absent	Total
Test positive			
Test negative			
total	**100 (i.e., .0001* 1,000,000)**		1,000,000

Number of patients without disease obtained by subtraction

	Disease present	Disease absent	Total
Test positive			
Test negative			
total	100	**999,900**	1,000,000

Sensitivity=99%, specificity=99%

	Disease present	Disease absent	Total
Test positive	**99 (i.e., .99*100)**		
Test negative		**989,901 (i.e., .99*999,900)**	
total	100	999,900	1,000,000

Remainder of the table interior obtained by subtraction

	Disease present	Disease absent	Total
Test positive	99	**9,999**	
Test negative	**1**	989,901	
total	100	999,900	1,000,000

Other totals obtained by addition

	Disease present	Disease absent	Total
Test positive	99	9,999	**10,098**
Test negative	1	989,901	**989,902**
total	100	999,900	1,000,000

Calculate PPV: 99/10,098 = .01 (only 1% of patients with positive tests actually have the disease).

	Disease present	Disease absent	Total
Test positive	**99**	9,999	**10,098**
Test negative	1	989,901	989,902
total	100	999,901	1,000,000