MATHEMATICAL MODELING IN TEACHING PROBABILITY AND STATISTICS FOR MEDICAL STUDENT THROUGH EXAMPLES

Cao Huu Hoa - Pham Minh Trien
Tra Vinh University
Email: phamminhtrien@tvu.edu.vn.

Abstract: *Nowadays, the teaching of Probability and Statistics in medical education requires a significant shift from a purely theoretical approach to one that is closely connected to real-world medical practice. This article proposes the use of mathematical modeling as both a tool and a goal in teaching probability and statistics. Based on a synthesis of perspectives on modeling and its implementation process, the paper analyzes three practical case studies in the healthcare field to illustrate how modeling can be integrated into teaching. Consequently, several pedagogical recommendations and strategies for developing modeling competencies among medical students are proposed.*

Keywords: *Medical education, Probability and Statistics, mathematical modeling, health sciences; pedagogy.*
Received a manuscript: 19/12/2025; Edited: 22/12/2025; Reviewed: 29/12/2025; Paper accepted: 05/01/2026.

1. Introduction

Probability and Statistics form one of the most essential intellectual foundations for modern medical education. From diagnostic accuracy and treatment comparison to epidemiological surveillance and health economic evaluation, statistical reasoning permeates nearly every dimension of clinical or biomedical practice. In the current era of big medical data, clinicians and researchers must be able not only to understand statistical results but also to interpret their assumptions, judge model appropriateness, and contextualize findings for decision-making. However, instruction in Probability and Statistics for medical undergraduates remains largely theoretical in many institutions. Students often encounter isolated formulas, procedures, and definitions without understanding how statistical reasoning emerges from real-life medical uncertainty. This gap creates a disconnect between classroom learning and professional practice. International literature repeatedly emphasizes the urgency of transforming statistics education toward contextual, practice-oriented, and model-centered approaches. Mathematical modeling offers a promising instructional pathway for bridging this gap. Modeling encourages learners to interpret medical phenomena, conceptualize underlying stochastic mechanisms, and construct mathematical or statistical representations that can generate evidence for clinical decisions. Rather than treating statistics as a set of computational recipes, modeling cultivates a mindset centered on inquiry, explanation, and reasoning.

This study aims to: articulate theoretical perspectives on mathematical modeling relevant to teaching Probability and Statistics for medical students; propose a modeling process adapted from current international frameworks; provide three detailed modeling examples situated in realistic healthcare contexts; and draw pedagogical implications for improving modeling competence in medical statistics education.

2. Perspectives on Modeling in Teaching Probability and Statistics

2.1. Mathematical Modeling

Mathematical modeling is the process of constructing and refining a mathematical model in order to represent and solve real-world problems. Through mathematical modeling, we learn how to select and apply a range of data types, as well as appropriate mathematical methods and tools, to solve real-world problems. Opportunities to work with real data and to use mathematical tools to analyze that data should be an integral part of mathematics learning at all levels.

Mathematical modeling is a complex activity that involves moving back and forth between mathematics and reality in both directions. Therefore, it requires us to possess many other competencies in different areas of mathematics, as well as knowledge related to the real-world situations under consideration. In mathematics education, modeling is often analyzed from various perspectives. When applied to the Probability-Statistics module in Medicine, we can emphasize the following two main viewpoints:

2.2. Modeling as an Educational Tool

Modeling has long been viewed as a powerful means to support conceptual learning. In medical statistics education, modeling as a tool helps students explore how probability distributions, estimation procedures, and inferential frameworks emerge naturally from clinical problems. Lecturers design medical scenarios-such as disease prevalence estimation, treatment efficacy comparison, risk prediction, or diagnostic test evaluation-through which statistical ideas become meaningful.

As Ngoc (2023) argues, contextualized modeling promotes deeper understanding because students continuously move between reality and abstraction. When learners select variables, determine distributions, and justify assumptions, they internalize the meaning behind statistical formulas that would otherwise remain abstract.

2.3. Modeling as an Educational Goal

A more advanced perspective considers modeling competence itself as a major learning outcome. For medical students, modeling competence includes the ability to:

Identify real-world medical problems that require stochastic thinking;

Translate clinical or public health phenomena into mathematical structures;

Perform statistical analysis with appropriate methods;

Evaluate assumptions and interpret results responsibly; and

Communicate findings to both expert and non-expert audiences.

Binh (2023) emphasizes that such competencies underpin evidence-based medicine, where professionals must critically assess uncertainty, variability, and methodological limitations before applying research findings to patient care.

2.4. Integrating Both Perspectives in Instruction

Introductory medical statistics courses should integrate both perspectives: using modeling as a pedagogical tool while gradually cultivating modeling competence as a formal learning objective. Assessment practices must align with this vision. Rubrics may evaluate abilities such as problem identification, abstraction, analytical reasoning, interpretation, and reflection.

3. A Five-Step Modeling Process for Medical Statistics Education

Drawing on Blum (2011), Zbiek and Conner (2006), and recent pedagogical developments, we propose a five-step modeling process tailored for medical contexts.

Step 1: Understanding the Real-World Problem

Students first explore the medical scenario, clarify the research question, determine the target population, and specify the analytical objectives-estimation, prediction, comparison, or optimization. Ethical considerations, patient safety, and data privacy should also be part of this stage.

Step 2: Abstraction and Model Construction

The problem is translated into probabilistic and statistical structures. Students identify variables, parameters, sampling designs, distributions, assumptions, and constraints. This abstraction step is central to modeling competence.

Step 3: Computation and Statistical Analysis

Students select and apply appropriate statistical techniques: confidence intervals, hypothesis testing, regression, survival analysis, simulation, or cost-effectiveness modeling. Software tools such as R, Python, or SPSS (Statistical Package for the Social Sciences) are recommended to enhance analytical accuracy and reproducibility.

Step 4: Interpretation and Model Evaluation

Findings must be interpreted in the original medical context. Students examine uncertainty, effect sizes, model fit, assumption validity, and generalizability. They discuss whether results support clinical decisions or require further investigation.

Step 5: Reflection and Refinement

Students reflect on model limitations, consider additional data needs, modify assumptions, and conduct sensitivity analyses. They articulate how their model contributes to decision-making and identify future improvements.

4. Three Comprehensive Case Studies

In order to illustrate how mathematical modeling can be integrated into the teaching of probability and statistics for medical students, we present three case studies. Each example is analyzed according to the five-step modeling process outlined in Section 3.

4.1. Example 1: Estimating the Prevalence of Human papillomavirus (HPV) Infection Among Women Aged 181 - 25

Step 1 - Understanding the real-world problem

Problem statement: We need to estimate the prevalence of HPV infection among women aged 18 - 25 in order to support the planning of a national HPV vaccination program.

Research question: “What is the true prevalence of HPV infection in the target population?”

Target population: Women aged 18 - 25 years in the community.

Analytical objective: To use representative data while ensuring confidentiality, and to account for potential sources of error that may affect the estimation, such as sampling error, diagnostic test error, and nonresponse bias.

Step 2: Abstraction and model construction

This situation can be modeled as a proportion estimation problem, where p denotes the true prevalence (infection rate) of HPV in the population. A random sample of size (n = 400) women is drawn, and 52 individuals are identified as HPV-positive. Therefore, the sample proportion is .

$$p = 52/400 = 0{,}13$$

Step 3: Computation and statistical analysis

We construct a 95% confidence interval for the true proportion using the normal approximation:

$CI = p \pm z_{0,975}\sqrt{\frac{p(1-p)}{n}}$, where $z_{0,975} = 1,96$

The standard error is $SE = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0,13(1-0,13)}{400}} = 0,0167$

Thus, $CI = 0,13 \pm 1,96 \times (0,0167) = (0,097;\, 0,163)$

Step 4: Interpretation and evaluation

We conclude that the estimated prevalence of HPV among women aged 18- 25 in this city ranges from 9.7% to 16.3% with 95% confidence. This information is critical for determining priorities in HPV vaccination programs, allocating healthcare budgets, and planning preventive health communication strategies. However, the appropriateness of the statistical model also needs to be carefully evaluated:

Is the assumption of random sampling reasonable?

Does the diagnostic test have sufficiently high sensitivity and specificity?

Does the confidence interval adequately reflect the uncertainty inherent in the real-world data?

To what extent can the results be generalized to other regions or population groups?

Step 5: Feedback and improvement

If the confidence interval is excessively wide, a larger sample size may be required. In addition, the model can be improved through the following measures:

adopting a stratified sampling design based on geographic regions or socioeconomic status to reduce sampling error and enhance representativeness;

applying estimates adjusted for test sensitivity and specificity;

extending the model within a Bayesian framework to incorporate prior information;

conducting sensitivity analyses to assess the impact of sampling assumptions;

collecting additional longitudinal data to evaluate trends in HPV infection over time.

4.2. Example 2: Analyzing COVID-19 Mortality Risk by Age Groups

Step 1: Understanding the real-world problem

Problem statement: In the context of the COVID-19 pandemic, hospitals need to assess mortality risk across age groups in order to support resource allocation, prioritize vaccination strategies, and identify patient groups requiring intensive monitoring.

Research question: “Does the risk of death from COVID-19 differ across age groups?”

Target population: Patients hospitalized with COVID-19 within a specified time period.

Analytical objective: To compare mortality risk among age groups (under 50, 50 - 69, and 70+).

Step 2: Abstraction and model construction:

We construct a contingency table of mortality outcomes by age group.

Age Group	Survived	Died	Total
<55	180	5	185
50 - 69	140	15	125
70+	60	20	80
Total	380	40	420

Step 3: Computation and statistical analysis

Mortality probability for each age group:

Under 50:

- Under 50: $5/185 \approx 0,027\,(2,7\%)$
- 50 – 69: $15/155 \approx 0,097\,(9,7\%)$
- 70+: $20/80 \approx 0,25\,(2,5\%)$

To examine differences among groups, a chi-square test was applied to the contingency table. The results showed $p < 0,001$, indicating a statistically significant difference in mortality risk across age groups.

Step 4: Interpretation and evaluation

The results indicate that the risk of mortality increases markedly across age groups. A significance level of p < 0.001 demonstrates that age is a factor strongly associated with mortality. When evaluating the model, it is necessary to consider the impact of recording errors,

the assumption of independence of observations;

the adequacy of the contingency table model;

the generalizability of the results to a larger population.

These findings have important practical implications for supporting clinical decision-making and health policy planning.

Step 5: Feedback and improvement

The contingency table model reveals a clear association; however, it does not simultaneously account for other important risk factors such as comorbidities, sex, vaccination status, or disease severity at hospital admission. Therefore, the model can be extended using multivariable logistic regression to estimate adjusted mortality risk while controlling for multiple covariates. Sensitivity analyses or scenario-based simulations may be conducted to assess the robustness of the results. The WHO report (2023) also recommends supplementary analyses incorporating comorbid conditions and vaccination status to achieve a more comprehensive understanding of mortality risk. Collecting longitudinal (cohort) data and periodically updating the model would further enhance its predictive value.

4.3. Example 3: Decision-Making in Treatment Based on Cost-Effectiveness Analysis

Step 1: Understanding the problem

Problem statement: The healthcare system needs to decide whether to select and reimburse a new antihypertensive drug compared to the standard drug.

Research Question: "Does the new drug offer better cost-effectiveness than the standard drug for patients with mild to moderate hypertension?"

Target Population: Adult patients receiving treatment for hypertension.

Purpose of Analysis: To compare cost-effectiveness (Cost-Effectiveness Analysis - CEA). Concurrently, to consider ethical factors (right to access treatment), measurement error in QALYs (Quality-Adjusted Life Years), and data security when collecting patient information.

Step 2: Abstraction and model construction:

Key Variables:

Annual drug cost (USD)

Quality-Adjusted Life Years (QALYs)

Parameters:

Standard Drug: 500 USD/year, 0.80 QALY

New Drug: 1,500 USD/year, 0.90 QALY

Assumptions:

QALYs are estimated from clinical trial data.

No significant difference in severe side effects.

The 1-year time horizon model is representative of long-term outcomes.

Model Structure:

Basic Cost-Effectiveness Model (CEA).

Outcome Measure: Incremental Cost-Effectiveness Ratio (ICER).

Data Design and Distribution:

Cost and QALY data are assumed to follow an approximately normal distribution.

Step 3: Computation and statistical analysis:

Applying the formula

$$\text{ICER}=\frac{(\text{Cost of New Drug}-\text{Cost of Standard Drug})}{(\text{QALY of New Drug}-\text{QALY of Standard Drug})}$$

We obtain the Incremental Cost-Effectiveness Ratio (ICER):

ICER = (1500 - 500) / (0.90 - 0.80) = 1000 / 0.10 = 10,000 USD/QALY

Step 4: Interpretation and evaluation:

The result ICER = 10,000 USD/QALY indicates that the new drug provides an additional 0.10 QALY at an extra cost of 1,000 USD per patient. Assuming the national Willingness-to-Pay (WTP) threshold is 15,000 USD/QALY, then ICER < WTP, meaning the new drug is considered cost-effective.

Step 5: Feedback and improvement:

We should evaluate the limitations:

Confidence intervals for QALYs have not been calculated.

Subgroup analysis (by age, disease severity) has not been performed.

The 1-year model may not reflect long-term impact.

Proposed Improvements:

Univariate and multivariate sensitivity analysis.

Scenario analysis (drug price increase/decrease, change in effectiveness).

Monte Carlo simulation to assess the ICER distribution (Drummond et al., 2023).

5. Pedagogical Recommendations

Based on the analysis of three representative modeling examples in the teaching of Probability and Statistics for medical students, we propose several pedagogical recommendations aimed at enhancing instructional effectiveness, fostering statistical thinking, and strengthening evidence-based decision-making competence in medical practice.

Organizing instruction explicitly around the five-step modeling cycle helps students clearly recognize the role of statistics as a descriptive and analytical tool, understand that statistical results are not "absolute truths" but are contingent on underlying assumptions and data, and cultivate habits of critical appraisal and evaluation of model adequacy.

Statistical content should be embedded in authentic and professionally meaningful medical contexts, as such contexts play a central role in activating students' learning motivation and their ability to apply knowledge. Therefore, when teaching Probability and Statistics to medical students, instruction should incorporate scenarios drawn from epidemiology, clinical practice, and health economics; pose questions that arise from real-world decision-making needs (risk estimation, group comparisons, treatment selection); and emphasize the practical interpretation of statistical parameters (prevalence, risk, the incremental cost-effectiveness ratio [ICER], and confidence intervals). This approach enables students to perceive statistics as an integral and inseparable component of evidence-based medical practice.

Developing uncertainty thinking and statistical interpretation competence through confidence intervals, hypothesis testing, and sensitivity analysis. Accordingly, instructors should guide students to understand the meaning and interpretation of confidence intervals and p-values rather than merely performing mechanical calculations, and to clearly distinguish between statistical significance and clinical significance.

Shift the assessment focus toward modeling competence rather than merely evaluating computational techniques. Instructors should employ case-based problems, develop explicit assessment rubrics, and emphasize students' ability to understand the problem, select an appropriate statistical model, interpret results, and critically evaluate model assumptions. In addition, small-scale projects (mini-projects) based on real or simulated medical data should be incorporated. Scaffold scenario complexity, starting from basic estimation and progressing to multivariable modeling and cost-effectiveness analysis.

6. Conclusion

Modeling-based teaching offers a powerful pedagogical framework for transforming the learning of Probability and Statistics in medical education. By anchoring instruction in authentic medical contexts, modeling enhances conceptual understanding, strengthens critical thinking, and promotes evidence-based reasoning.

Through the three case studies, this paper demonstrates how modeling illuminates statistical concepts related to epidemiology, clinical risk assessment, and health economics. Future curriculum development should incorporate systematic modeling instruction, integrate computational tools, and emphasize reflective analysis to cultivate modeling competence among future medical professionals ■

References

[1]. Ngoc, T. H. Đ. (2023). *Mathematical modeling in statistics teaching for healthcare students.* Vietnam Journal of Higher Education, 15 (2), 45 - 58.

[2]. Binh, V. N. (2023). *Developing statistical thinking in medical education through modeling.* Vietnam Journal of Educational Sciences, 18 (1), 12 - 27.

[3]. Trien, M. P. (2024). *Integrating modeling and data analysis in medical statistics courses.* Vietnam Journal of Education and Training, 22 (3), 34 - 49.

[4]. Trien, M. P., & Long, H. N. (2024). *Rubrics for assessing modeling competence in undergraduate statistics.* Vietnam Journal of Faculty and Educational Professionals, 6 (4), 55 - 70.

[5]. World Health Organization (2023). *Global HPV vaccination and prevalence strategies.* WHO Publications.

[6]. Drummond, M. et al. (2023). *Methods for the Economic Evaluation of Health Care Programmes (5th ed.).* Oxford University Press.

[7]. Blum, W. (2011). *Can Modelling Be Taught and Learnt? Some Answers from Empirical Research. In G. Kaiser, W. Blum, R. Borromeo Ferri & G. Stillman (Eds.), Trends in Teaching and Learning of Mathematical Modelling (pp. 15-30).* Dordrecht: Springer.

[8]. Zbiek, R. M., & Conner, A. (2006). *Beyond motivation: Exploring mathematical modeling as a context for deepening students' understandings of curricular mathematics.* Educational Studies in Mathematics, 63(1), 89 - 112.

Mô hình hóa toán học trong giảng dạy xác suất - thống kê Cho sinh viên y khoa thông qua các ví dụ

Cao Hữu Hòa - Phạm Minh Triển
Trường Đại học Trà Vinh
Email: phamminhtrien@tvu.edu.vn.

Tóm tắt: *Hiện nay, việc dạy học Xác suất và Thống kê trong đào tạo sinh viên ngành Y đang đòi hỏi một sự chuyển đổi mạnh mẽ từ cách tiếp cận thuần túy lý thuyết sang các phương pháp gắn kết với thực tiễn y học. Bài viết đề xuất vận dụng tiếp cận mô hình hóa toán học như một công cụ hoặc mục tiêu trong giảng dạy xác suất - thống kê. Trên cơ sở tổng hợp các quan điểm về mô hình hóa và quy trình thực hiện, bài viết phân tích ba ví dụ minh họa mang tính thực tiễn trong lĩnh vực y tế để làm rõ phương pháp triển khai mô hình hóa trong giảng dạy. Từ đó, một số khuyến nghị về phương pháp sư phạm và phát triển năng lực mô hình hóa cho sinh viên ngành Y được đề xuất.*

Từ khóa: *Giáo dục y học, xác suất và Thống kê, mô hình hóa toán học, lĩnh vực sức khỏe, các đề xuất về phương pháp giảng dạy.*