PharmacoLearn

MPH203T Computer Aided Drug Delivery System, M.Pharm Semester 2

Unit 1: Computers in Pharmaceutical Research and Development and Quality by Design

Free notes at pharmacolearn.fyi

Computer Aided Drug Delivery System, subject code MPH203T, is a core Semester 2 subject of M.Pharm Pharmaceutics. It carries 4 credits and 60 hours of theory, taught at 4 hours per week, for 100 marks. The computer has moved from the edge of pharmaceutical work to its centre. Almost every decision in modern drug development, which molecule to advance, which formulation to scale up, which dose to take into patients, is now made first on a screen and only then at the bench. This subject follows that idea through the whole research and development process. The first unit starts with how computers entered pharmaceutical research, the statistical modelling that underlies all later units (descriptive and mechanistic models, estimation, confidence regions, sensitivity analysis, optimal design and population modelling), and Quality by Design, the ICH Q8 framework that turned modelling and designed experiments into a regulatory expectation. The second unit models what the body does to the drug: computational prediction of absorption, solubility, intestinal permeation, distribution and excretion, with the active transporters (P-gp, BCRP, hPEPT1, ASBT, OCT, OATP, the nucleoside and blood brain barrier choline transporters) that decide the fate of many molecules. The third unit applies computers to the formulation itself: the concept of optimization, factorial and screening designs, the computer aided development of emulsions and microemulsion carriers, and the legal, ethical and commercial frame (protection of computer based innovation, ethics of computing, market analysis). The fourth unit simulates the dosage form in the body: gastrointestinal absorption simulation with model construction, parameter sensitivity analysis, virtual trials and fed versus fasted prediction, the link to dissolution and IVIVC and biowaivers, simulations in pharmacokinetics and pharmacodynamics from the whole organism down to genes, and computers in clinical development, including clinical data management and the regulation of computer systems. The last unit looks at the newest tools: artificial intelligence, robotics and pharmaceutical automation, and computational fluid dynamics, with their applications, advantages, disadvantages and future directions.

1.1 History of computers in pharmaceutical research and development

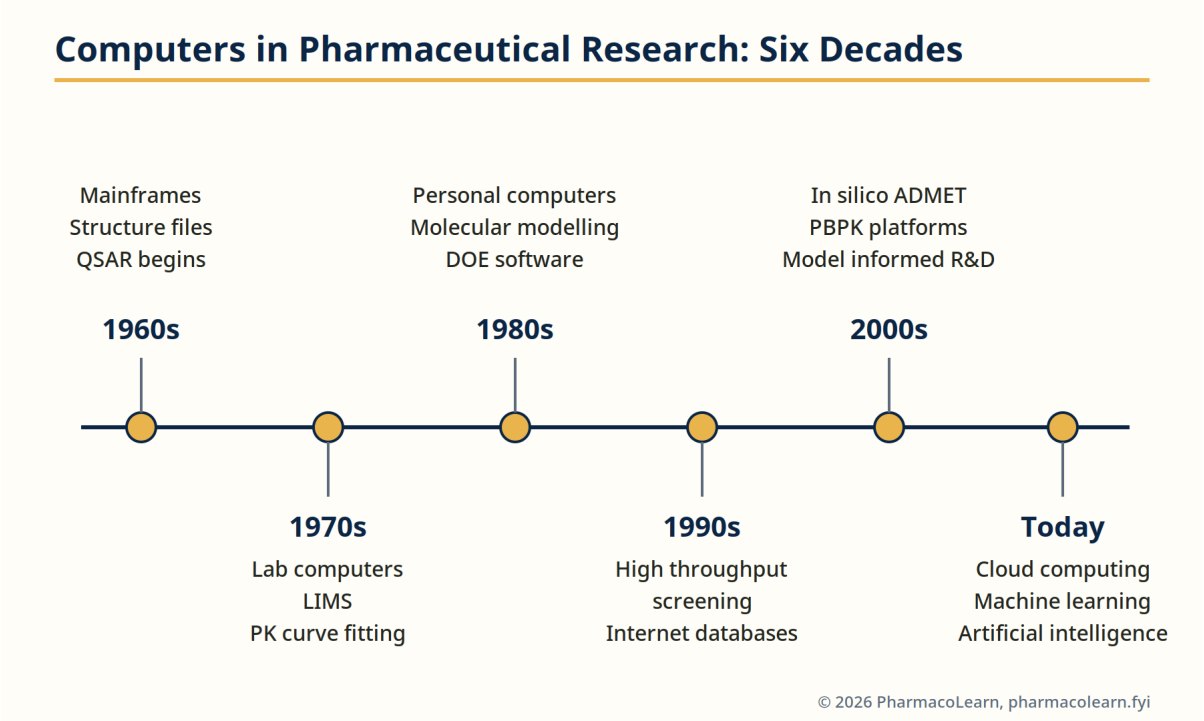

Figure: Sixty years of computing in pharmaceutical research. Each decade first automated a manual task, then made a previously impossible task routine.

Computers entered pharmaceutical research in the 1960s, when mainframe machines were first used to store chemical structure files, search compound collections and process the growing volume of analytical and pharmacological data that manual records could no longer handle. The same decade produced the first genuinely computational science of the field: quantitative structure activity relationships (QSAR), in which Hansch and others showed that biological activity could be regressed against computed physicochemical descriptors such as partition coefficients and electronic parameters. In the 1970s, laboratory minicomputers took over instrument control and data capture, laboratory information management systems (LIMS) began to organise samples and results, and pharmacokinetic curve fitting moved from graph paper to nonlinear regression programs. The 1980s brought the personal computer and the graphics workstation: molecular modelling and docking became visual, interactive activities, spreadsheets entered the formulation laboratory, and statistical packages made designed experiments practical for ordinary development teams. The 1990s were the decade of scale: combinatorial chemistry and high throughput screening generated data faster than scientists could inspect it, the internet connected databases and collaborators, and bioinformatics grew up alongside the sequencing of genomes. In the 2000s the emphasis shifted from generating data to predicting outcomes: in silico ADMET screening, physiologically based pharmacokinetic (PBPK) platforms and the first regulatory acceptance of modelling and simulation as formal evidence, including the United States Food and Drug Administration (US FDA) promoting model informed drug development. The pattern across six decades is constant: each generation of hardware first automated an existing manual task, then made a previously impossible task routine.

1.2 Computers in pharmaceutical research: a general overview

Read as a map, this subject is the tour of a single idea, the computer as a decision tool, applied at every stage of research and development. In discovery, computers screen virtual libraries, predict activity and filter out molecules with hopeless solubility, permeability or toxicity before synthesis. In preclinical development they model drug disposition (Unit 2), so that animal and early human studies are designed around predicted behaviour instead of discovered surprises. In formulation development they replace one factor at a time tinkering with designed experiments and optimization (Unit 3). In biopharmaceutical

evaluation they simulate the dosage form in a virtual gut and a virtual population (Unit 4), linking dissolution to plasma profiles and supporting biowaiver arguments. In clinical development they capture, clean and lock trial data and keep the systems that hold them compliant (Unit 4). Two threads run through all of it: every computer method in the list is a model, a simplified mathematical stand in for a real system, and every model must be built, checked and used within its limits. The statistical modelling of the rest of this unit is the grammar for all of that.

1.3 Statistical modeling in pharmaceutical research: descriptive versus mechanistic modeling

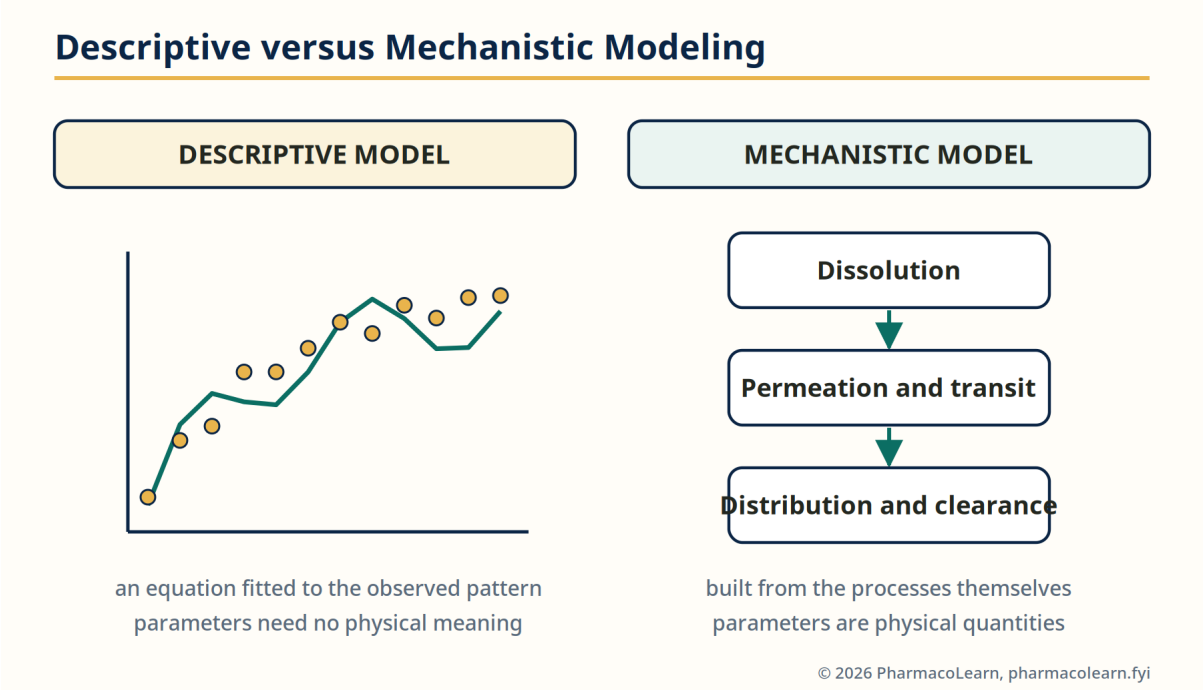

Figure: The two modelling philosophies of Unit 1. The descriptive model follows the data; the mechanistic model follows the process, and can therefore be extrapolated with stated assumptions.

A statistical model is an equation, with parameters, that connects inputs (dose, formulation variables, patient properties) to outputs (concentrations, responses, quality attributes) and leaves a residual for what it cannot explain. The syllabus divides modelling into two philosophies. A **descriptive (empirical) model** describes the pattern in the data without claiming to represent the underlying process: a polynomial fitted to a dissolution profile, or a compartment model used simply because its curve matches the plasma data. Its parameters need have no physical meaning, it is cheap to build, and it usually interpolates well inside the studied range, but it cannot be trusted outside that range, because nothing in it knows why the data bend the way they do. A **mechanistic model** is built from the actual processes believed to generate the data: dissolution, permeation and transit compartments for absorption (Unit 4), enzyme kinetics for metabolism, receptor binding for effect. Its parameters are physical quantities (a permeability, a clearance, a rate constant), it can extrapolate with some justification to new doses, formulations and populations, and it can be falsified process by process when it fails. The price is that it needs more knowledge, more parameters and more data to support them. Practice uses both, often in sequence: a descriptive model early, when little is known, and a mechanistic model when the programme is mature enough to ask "what if" questions.

Feature	Descriptive (empirical) model	Mechanistic model
Basis	Pattern in the observed data	The physical and biological processes generating the data
Parameters	Curve fitting constants, often without physical meaning	Physical quantities: permeability, clearance, rate constants

Feature	Descriptive (empirical) model	Mechanistic model
Data needed	The dataset being described	Process knowledge plus data to estimate the physical parameters
Interpolation	Usually good inside the studied range	Good, and structured by the process
Extrapolation	Unsafe outside the studied range	Possible with stated assumptions, and testable
Typical examples	Polynomial fit to a dissolution profile; empirical PK curve fit	PBPK model; ACAT gastrointestinal absorption model; receptor occupancy PD model

1.4 Statistical parameters and estimation

Every model carries **parameters**: the fixed numbers that must be learned from data before the model can predict anything, such as the slope and intercept of a calibration line, the elimination rate constant and volume of a compartment model, or the coefficients of a factorial design equation (Unit 3). **Estimation** is the statistical procedure that chooses parameter values making the model agree best with the observations. Two estimation principles dominate pharmaceutical work. Least squares chooses the values that minimise the sum of squared differences (residuals) between observed and predicted values; weighted least squares gives less influence to less precise observations, which matters for concentration data whose error grows with the value. Maximum likelihood chooses the values under which the observed data are most probable, given an assumed error distribution, and is the engine inside population modelling software (section 1.9). Estimated parameters are never exact: each carries uncertainty, expressed as a standard error, and a fit must be judged not by whether the curve touches the points but by whether the residuals are patternless, the parameters are estimated with usable precision, and the model predicts new data.

1.5 Confidence regions

Because parameters are estimated from finite, noisy data, a point estimate is only half an answer. A **confidence interval** gives the range of values for one parameter compatible with the data at a stated confidence level, conventionally 95 percent: loosely, the range within which the true value is believed to lie, given the data and the model. When several parameters are estimated together, the corresponding object is a **confidence region**, a joint region in parameter space (an ellipse for two parameters, a volume for more). Regions matter because estimated parameters are usually correlated: raising one estimate can be compensated by lowering another while the fit barely changes, so two parameters can each look precise alone yet be poorly determined together, a warning sign in any model used for decisions. Narrow regions come from informative data, good designs (next sections) and models no more complex than the question requires.

1.6 Nonlinearity at the optimum

Much pharmaceutical optimization searches for an optimum: the formulation, dose or set of conditions giving the best response. Real response surfaces are curved, and near an optimum they are at their most curved, because a maximum or minimum is by definition where the surface bends over. This is the problem of **nonlinearity at the optimum**. Linear approximations, which are adequate on the flank of a surface, fail at the top: the estimated optimum position becomes uncertain, small parameter changes move the predicted optimum a long way, and the confidence region around the optimum is asymmetric and larger than linear theory suggests. Practical consequences follow. An optimum found by fitting should be confirmed by experiment at and around the predicted point, not trusted from the equation alone. Optimization should report not just the best point but the region of near optimal conditions, which is exactly

the idea the design space of Quality by Design formalises (section 1.10). And where two very different settings give nearly equal responses, the flatter, more robust one is usually the better manufacturing choice.

1.7 Sensitivity analysis

Sensitivity analysis asks how much a model's output changes when its inputs or parameters change. It is the standard way of finding out what a model actually depends on. In the local form, one parameter at a time is varied by a small amount around its estimate while the others are held fixed, and the resulting change in output (a plasma concentration, a predicted response, a quality attribute) is recorded, often drawn as a ranked bar or tornado chart. In the global form, all parameters are varied together across their plausible ranges, typically by simulation, so interactions between parameters are captured as well. A prediction whose output is violently sensitive to a parameter nobody has measured is not a prediction but a guess, and sensitivity analysis is how that is found out before the clinic does it.

1.8 Optimal design

If sensitivity analysis asks what a model depends on, **optimal design** asks how an experiment should be run so that the model is estimated as well as possible. Classical designs (Unit 3) spread experimental runs by habit and symmetry. Optimal design instead starts from the model and a criterion: choose the experimental conditions, the sampling times, doses or factor settings, that maximise the information the data will carry about the parameters. The machinery is the Fisher information matrix, which summarises how precisely the parameters can be estimated from a given design; the most used criterion, D-optimality, chooses the design that maximises the determinant of that matrix, equivalently minimising the volume of the joint confidence region of the parameters. Optimal design is used heavily in pharmacokinetics, where it answers questions such as the few best blood sampling times for estimating clearance and volume in a trial that can afford only sparse samples in each patient, and in formulation work, where it can build a small, efficient design for an awkward factor space that no textbook design fits. Its weakness mirrors its strength: the optimal design is optimal for the assumed model and parameter values, so sound practice checks the design across plausible parameter values and keeps some robustness in reserve.

1.9 Population modeling

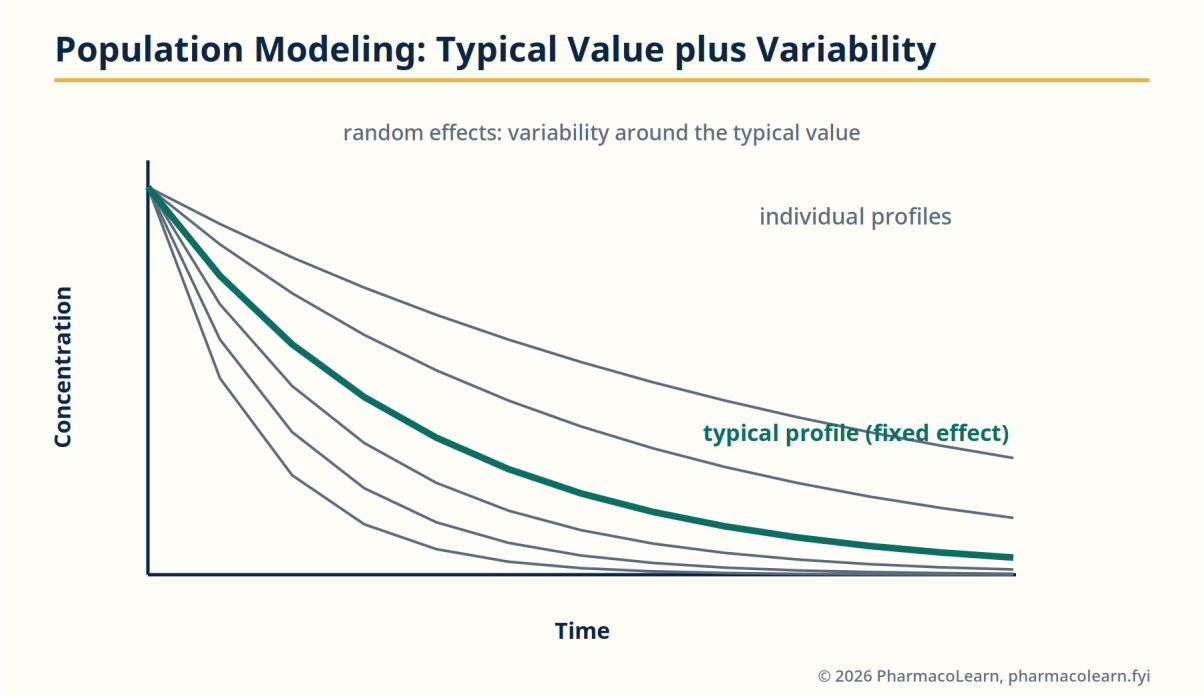

Figure: Population modeling in one picture: every subject has a profile, the model estimates the typical profile and the distribution of variability around it, explained and unexplained.

Classical pharmacokinetic analysis fits one subject at a time. **Population modeling** fits all subjects at once with a model that has two levels: fixed effects, the typical parameter values and the systematic influence of patient properties (covariates) such as weight, age, renal function and genotype; and random effects, the unexplained variability between subjects and the residual variability within a subject over time. The result is not just an average curve but a description of the whole distribution of behaviour in the population, including how much of the variability is explained by measurable covariates. This matters because drug development decisions are population decisions: a dose must work across patients, including the slow metaboliser and the patient with poor kidneys, not just in the average volunteer. Population modelling handles the sparse, unbalanced data that real trials produce (a few samples per patient, at different times), where one subject at a time fitting is impossible, and it is the standard analysis behind dose recommendations in special populations. The approach was pioneered in the NONMEM software tradition, in which the model is written as differential equations for the typical individual plus statistical models for variability, and estimated by likelihood based methods (section 1.4). The same two level thinking, typical value plus structured variability, underlies the virtual trials of Unit 4.

1.10 Quality by Design in pharmaceutical development: introduction and ICH Q8

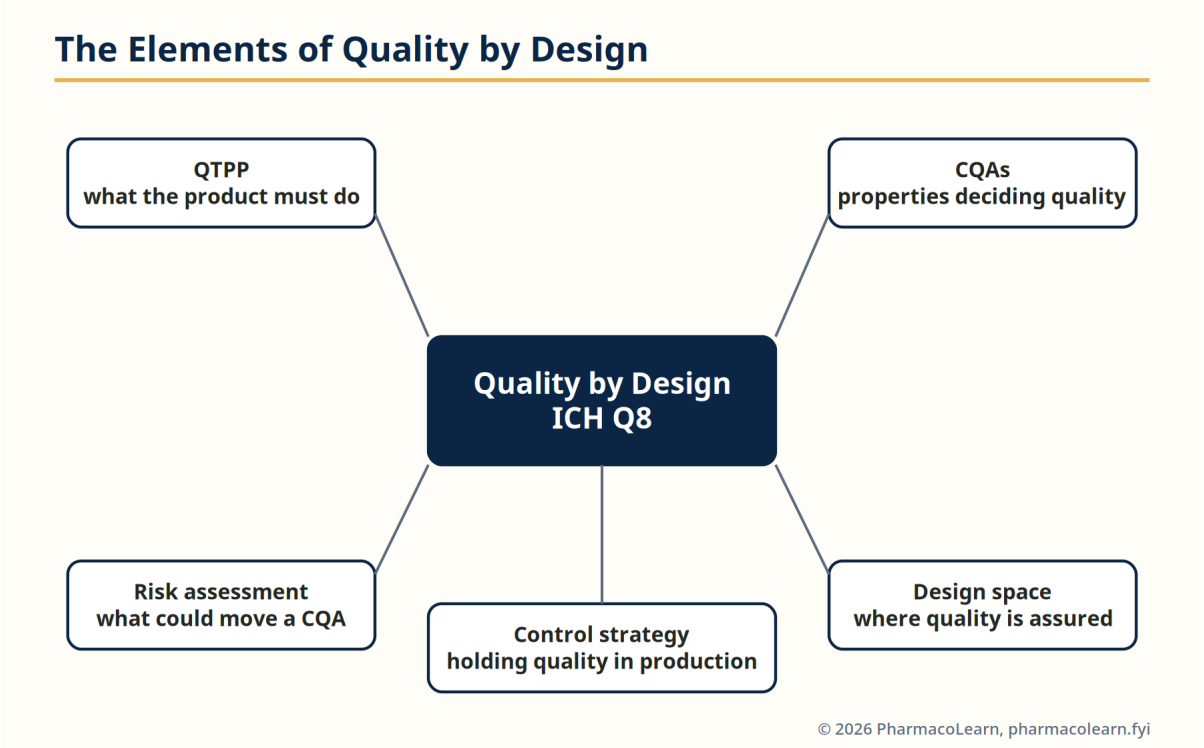

Figure: The elements of Quality by Design under ICH Q8. The QTPP states the goal, CQAs define quality, risk assessment ranks the dangers, the design space marks where quality is assured, and the control strategy holds production inside it.

Quality by Design (QbD) is the systematic approach to development that begins with predefined objectives and emphasises product and process understanding and process control, based on sound science and quality risk management. That definition is from ICH Q8, the International Council for Harmonisation guideline on pharmaceutical development, which is the regulatory core of this topic (the current version, ICH Q8(R2), is read together with ICH Q9 on quality risk management and ICH Q10 on the pharmaceutical quality system). QbD replaced the older quality by testing philosophy, in which quality was verified at the end by testing the finished batch, with the principle that quality should be built in by understanding. Its working elements are:

- **Quality target product profile (QTPP):** a prospective summary of the quality characteristics the product should have to ensure the desired clinical performance: dosage form, route, strength, release profile, stability and the attributes patients and clinicians depend on.
- **Critical quality attributes (CQAs):** the physical, chemical, biological or microbiological properties that must lie within limits to ensure the desired quality, for a tablet typically assay, content uniformity, dissolution and degradation products.
- **Risk assessment:** structured tools (cause and effect diagrams, failure mode and effects analysis) linking material attributes and process parameters to CQAs and ranking what matters, so effort goes where risk is.
- **Critical material attributes and critical process parameters:** the input variables whose variation moves a CQA, identified from the risk assessment and from prior knowledge.
- **Design space:** the multidimensional combination and interaction of input variables and process parameters that has been demonstrated to provide assurance of quality. Working within the design space is not considered a change by regulators; moving outside it is.
- **Control strategy:** the planned set of controls, on materials, process conditions, in process measurements and finished product testing, that together ensure the process performs and the product meets its CQAs, maintained across the product lifecycle with continual improvement.

QbD element	What it answers	Typical output for a tablet
QTPP	What must the product do?	Immediate release tablet, stated strength, dissolution target, 24 month shelf life
CQA	Which properties decide quality?	Assay, content uniformity, dissolution, impurities
Risk assessment	What could move a CQA?	Ranked list: granulation moisture and compression force ranked high
Design space	Where is quality assured?	A region of moisture, force and speed combinations proven acceptable
Control strategy	How is quality held in routine production?	Material specifications, in process controls, PAT checks, end product tests

1.11 Regulatory and industry views on QbD

Regulators and industry arrived at QbD from opposite frustrations. For regulators, the quality by testing system produced products that passed their specifications yet were made by processes nobody fully understood, so any manufacturing change demanded fresh regulatory submissions and approval delays, discouraging improvement. The US FDA's process analytical technology (PAT) initiative and the ICH Q8, Q9 and Q10 guidelines reframed the bargain: a manufacturer who demonstrates genuine process understanding, through risk assessment, designed experiments and a design space, earns regulatory flexibility, including movement within the approved design space without prior approval, and a science and risk based review. The European Medicines Agency holds the same position, and joint FDA and EMA question and answer documents exist precisely because both agencies review QbD applications. Industry's view is the mirror image: QbD development costs more at the start (more experiments, more modelling, more documentation) and pays back across the lifecycle, in fewer failed batches and out of specification results, smoother scale up and technology transfer, faster handling of changes, and a process that stays in control because its behaviour is understood rather than merely observed.

1.12 Scientifically based QbD: examples of application

A scientifically based QbD study is the practical union of this whole unit, so one worked pattern is worth following. Consider an immediate release tablet programme. The QTPP is written first, with the dissolution target the clinical team needs. Prior knowledge and a risk assessment identify the plausible critical variables: granulation moisture, lubricant level and blending time, and compression force, acting on the CQAs of dissolution, hardness related handling quality and content uniformity. A screening design (Unit 3) tests many candidate variables in few runs and keeps the vital few. A factorial or response surface design then maps the survivors, fitting a mechanistic leaning statistical model of dissolution and uniformity against the variables; sensitivity analysis (section 1.7) confirms which effects dominate; confidence regions (section 1.5) show how firmly the boundaries are known; and the design space is declared as the region where the model, with its uncertainty, predicts all CQAs in specification with high probability. Routine manufacture then runs under a control strategy that may include PAT monitoring of granulation, and the knowledge is filed so that post approval movement inside the space is a quality system action, not a new licence application. The same pattern, varying only the product, is reported in the literature for modified release tablets, lyophilised products and biological processes: define the target, assess risk, design experiments, model, define the space, control it, and keep learning across the lifecycle.

```
The QbD workflow of ICH Q8
QTPP: what the product must do
   |
   v
CQAs: the properties that decide quality
   |
   v
Risk assessment: rank materials and parameters
   |
   v
Designed experiments and models (Units 1 and 3)
   |
   v
Design space: where quality is assured
   |
   v
Control strategy and lifecycle learning
```

More free M.Pharm notes and unit wise PDFs at pharmacolearn.fyi
Free video lectures on the PharmacoLearn YouTube channel