

Why Mechanistic Models Complement Statistical Models

The statistical and machine learning models of Chapters 8-11 describe associations in observed data; mechanistic (compartmental) models instead encode explicit assumptions about the biological and behavioural process of disease transmission, and can therefore answer counterfactual questions that a purely statistical model cannot — for example, what would happen to malaria incidence in a state if ITN coverage increased by 20 percentage points, a scenario for which no observed data yet exist. This chapter introduces the compartmental modelling framework that underlies mathematical epidemiology, connecting it to the probability foundations of Chapter 3 and the differential-equation systems that describe how a population moves between disease states over time.

The SIR/SEIR Framework

The basic Susceptible-Infected-Recovered (SIR) model divides a population into compartments and describes the flow between them with a system of ordinary differential equations governed by a transmission rate and a recovery rate. The Susceptible-Exposed-Infectious-Recovered (SEIR) extension adds an exposed

(infected but not yet infectious) compartment, appropriate whenever a disease has a meaningful incubation period, and is the standard starting point for modelling diseases such as measles, COVID-19, and Lassa fever in the Nigerian context. The basic reproduction number, R_0 — the expected number of secondary infections produced by a single infectious individual in an otherwise fully susceptible population — emerges naturally from the model parameters and is the single most-quoted summary statistic in infectious disease modelling, governing whether an outbreak will grow ($R_0 > 1$) or die out ($R_0 < 1$).

Extensions for Vector-Borne and Vaccine-Preventable Disease

SVEIHR and Beyond

Malaria transmission requires an explicit mosquito vector compartment, since transmission depends on the interaction between human and vector populations (the biting rate, vector survival, and the extrinsic incubation period within the mosquito), producing coupled human-vector compartmental systems rather than the simpler single-population SEIR model. For vaccine-preventable diseases, adding a Vaccinated compartment (yielding an SVEIHR-type structure, with H representing hospitalised or severe cases) allows explicit modelling of the impact of a vaccination campaign's coverage and timing on subsequent case trajectories, a structure that has been used to model COVID-19 and Lassa fever dynamics in the Nigerian setting, incorporating hospitalisation as a distinct compartment to connect the model directly to health-system capacity planning (bed occupancy, oxygen demand) rather than case counts alone.

Parameter Estimation and Model Fitting

A compartmental model's parameters (transmission rate, incubation period, recovery rate) are rarely all known with

precision from first principles and must typically be estimated by fitting the model's predicted case trajectory to observed surveillance data — often via least squares or maximum likelihood, optimised numerically since the differential equation system generally has no closed-form solution. This connects directly back to Chapter 10's forecasting methods: a well-fitted SEIR model can generate short- to medium-term projections under specified assumptions, complementing (and cross-validating against) the purely statistical time series forecasts of Chapter 10, though the two approaches answer subtly different questions — a statistical time series model extrapolates observed patterns, while a mechanistic model extrapolates the consequences of an assumed transmission process, including under scenarios never yet observed.

R example

```
# R: a minimal SEIR model solved numerically with deSolve
library(deSolve)

seir_model <- function(time, state, parameters) {
  with(as.list(c(state, parameters)), {
    N <- S + E + I + R
    dS <- -beta * S * I / N
    dE <- beta * S * I / N - sigma * E
    dI <- sigma * E - gamma * I
    dR <- gamma * I
    list(c(dS, dE, dI, dR))
  })
}

parameters <- c(beta = 0.30, sigma = 1/5, gamma = 1/10) # per-day rates
state0 <- c(S = 99990, E = 0, I = 10, R = 0)
times <- seq(0, 180, by = 1)

output <- as.data.frame(ode(y = state0, times = times, func = seir_model, parms = parameters))
R0 <- parameters["beta"] / parameters["gamma"]
cat(sprintf("Basic reproduction number R0 = %.2f\n", R0))
```

Growth Curve Models for Endemic and Chronic Disease Burden

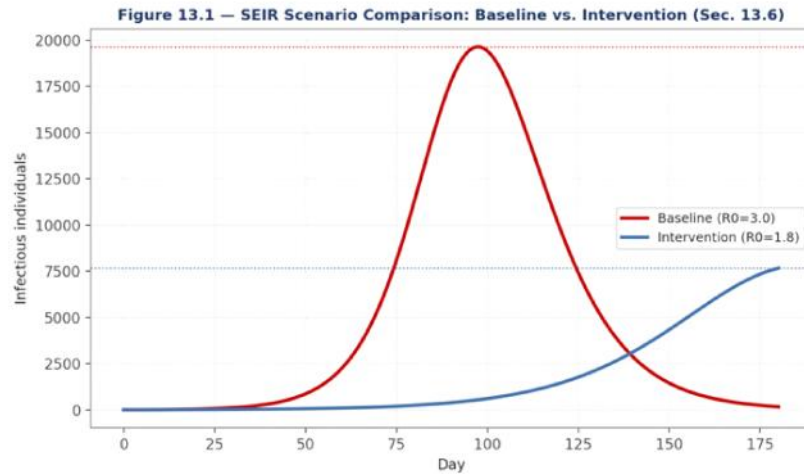
Not every health trajectory is well described by a compartmental epidemic model; many biological growth processes — a child's

growth trajectory, the accumulation of a disease burden indicator, or the trajectory of an intervention's population-level impact over time — are better captured by nonlinear growth curve models such as the logistic, Gompertz, and hyperbolic-type growth models, which share the property of an initial near-exponential phase that decelerates toward an asymptote. These models, closely related in mathematical structure to the logistic-growth term embedded within the SIR framework itself, form a foundational and recurring thread in nonlinear growth modelling research, and they remain valuable wherever a health indicator's trajectory over time needs to be summarised by a small number of interpretable parameters (an initial growth rate, an inflection point, and a carrying capacity or asymptote) rather than by an unconstrained statistical smoother.

Worked Example

Comparing Two Intervention Scenarios with SEIR

Using the SEIR model of Section 13.4 with $\beta = 0.30$, $\sigma = 1/5$, $\gamma = 1/10$ ($R_0 = 3.0$), the model projects a peak of roughly 19,600 simultaneously infectious individuals around day 97 in a population of 100,000. Simulating a hypothetical intervention that reduces the transmission rate by 40% ($\beta = 0.18$, representing, for instance, a combination of case isolation and behaviour change) reduces R_0 to 1.8 and the projected peak to roughly 7,800 infectious individuals, delayed to around day 187 — a 60% reduction in peak burden and roughly three additional months of time for health-system preparation. This is precisely the kind of counterfactual, scenario-comparison question (Section 13.1) that a fitted statistical time series model, which only extrapolates the observed trend, cannot answer, because no data yet exist for the intervention scenario.



PYTHON example

```
# Python: comparing baseline vs. intervention SEIR scenarios with scipy
import numpy as np
from scipy.integrate import odeint

def seir(y, t, beta, sigma, gamma, N):
    S, E, I, R = y
    dS = -beta * S * I / N
    dE = beta * S * I / N - sigma * E
    dI = sigma * E - gamma * I
    dR = gamma * I
    return [dS, dE, dI, dR]

N = 100000
y0 = [N - 10, 0, 10, 0]
t = np.linspace(0, 180, 181)

baseline = odeint(seir, y0, t, args=(0.30, 1/5, 1/10, N))
intervention = odeint(seir, y0, t, args=(0.18, 1/5, 1/10, N))

print(f"Baseline peak infectious: {baseline[:,2].max():.0f} on day {baseline[:,2].argmax()}")
print(f"Intervention peak infectious: {intervention[:,2].max():.0f} on day {intervention[:,2].argmax()}")
```

Worked Example

Fitting a Logistic Growth Curve

Suppose cumulative confirmed cases in a localised outbreak, tracked over 12 weeks, are well approximated by a three-parameter logistic growth curve: cumulative cases(t) = $K / (1 +$

$\exp(-r(t - t_0))$), where K is the eventual total case count (the asymptote), r is the growth rate, and t_0 is the inflection point (the week of fastest growth). Fitting this model by nonlinear least squares to the 12 weeks of data yields $K \approx 480$, $r \approx 0.55$, $t_0 \approx 6.2$ — meaning the model projects the outbreak will ultimately affect around 480 people in total, with the fastest rate of new case accumulation occurring in week 6, after which growth decelerates. This three-number summary is far more interpretable, and more useful for briefing a State Ministry of Health technical committee, than reporting all 12 raw weekly figures.

R example

```
# R: fitting a logistic growth curve to cumulative outbreak case counts
week <- 1:12
cumulative_cases <- c(3, 6, 14, 28, 52, 95, 165, 260, 350, 410, 450, 468)

logistic_fit <- nls(cumulative_cases ~ K / (1 + exp(-r * (week - t0))),
  start = list(K = 500, r = 0.5, t0 = 6))
summary(logistic_fit)

# Project two additional weeks beyond the observed data
predict(logistic_fit, newdata = list(week = c(13, 14)))
```

Stochastic vs. Deterministic Epidemic Models

The SEIR model presented in this chapter is deterministic: given a fixed set of parameters and starting conditions, it always produces exactly the same trajectory. Real outbreaks, especially small ones, are subject to genuine chance — a single early case might, by chance, infect zero or several secondary cases, and this stochasticity can mean a small outbreak with R_0 slightly above 1 fizzles out purely by chance rather than growing, a possibility a deterministic model cannot represent. Stochastic compartmental models (simulated many times with random variation in each transmission event) are preferred whenever case counts are small, such as in the earliest days of a novel outbreak, precisely because they can represent this chance-driven possibility of early extinction that a deterministic model always misses.

Extended Case Study

Retrospectively Validating a Model Against What Actually Happened

The most rigorous test of any mechanistic model is a genuinely prospective one: fitting the model using only data available up to a certain point, generating a forecast, and then — critically, without any subsequent adjustment — comparing that forecast against what was later actually observed. Consider a hypothetical but realistic example: at the start of a localised cholera outbreak, a modelling team fits an SEIR model using the first three weeks of case data and published parameter estimates from comparable prior outbreaks, and issues a projection that the outbreak will peak around week 7 at approximately 340 simultaneously active cases before declining, assuming no major change in current control measures.

When the actual outbreak data for weeks 4 through 12 subsequently become available, the true peak occurred slightly earlier (week 6) and somewhat lower (270 cases) than projected — a discrepancy that a careful team investigates rather than either dismissing as unimportant or treating as evidence the entire modelling exercise failed. Tracing the discrepancy, the team finds that a water-source remediation intervention was implemented in week 5, one week earlier than the response plan had originally scheduled, which is fully consistent with, and indeed a very plausible explanation for, both the earlier and lower observed peak relative to the no-additional-intervention baseline projection. This retrospective validation exercise, properly conducted and honestly reported (rather than quietly revising the original projection after the fact to make it appear more accurate than it was), serves two purposes: it builds genuine, evidence-based confidence in the model's parameter assumptions for use in the NEXT outbreak, and it provides a concrete, quantified estimate of the water-source intervention's apparent impact (a roughly 20% reduction in peak

case count, achieved by acting one week earlier than originally planned) that can directly inform how quickly to prioritise this specific intervention type in future cholera response protocols.

Herd Immunity Threshold and Its Practical Limits

A direct and often-quoted consequence of the SEIR framework's R_0 (Section 13.2) is the herd immunity threshold — the proportion of a population that must be immune (through prior infection or vaccination) for the effective reproduction number to fall below 1, calculated as $1 - 1/R_0$. For a disease with $R_0 = 15$ (measles, at the high end of commonly cited estimates), this implies roughly 93% of the population must be immune to interrupt sustained transmission, a very high bar that explains why even modest declines in measles vaccination coverage can permit outbreaks despite still-high overall coverage. It is important to note two practical limits on this simple formula: it assumes homogeneous mixing (every individual equally likely to contact every other), which real populations violate through household, school, and geographic clustering, meaning local pockets of low immunity can sustain transmission even when the region-wide average coverage exceeds the calculated threshold; and it describes the threshold for INTERRUPTING sustained transmission, not the threshold for preventing any outbreak whatsoever, since even a population above the herd immunity threshold can still experience a limited, self-terminating chain of cases following an imported case.

Agent-Based Models

An Alternative to Compartmental Modelling

The compartmental (SEIR-type) models emphasised throughout this chapter represent a population in aggregate, tracking only the total count of individuals in each disease state; an alternative modelling paradigm, agent-based modelling, instead simulates

each individual in a population separately, with their own specific location, contact network, and behaviour, allowing much richer representation of heterogeneous mixing patterns (Section 13.10's homogeneous-mixing limitation), realistic household and workplace contact structures, and individually targeted interventions (such as ring vaccination around a specific detected case) that a compartmental model's aggregate structure cannot represent at all. Agent-based models are considerably more computationally intensive to build and run, and require correspondingly more detailed input data about population structure and contact patterns to parameterise credibly, which is why compartmental models remain the more practical default choice for most routine forecasting and scenario-comparison tasks (as in Section 13.6), with agent-based modelling reserved for situations where the specific spatial or network structure of transmission is itself central to the policy question being asked — for example, evaluating a targeted household-level intervention strategy where the compartmental model's aggregate, well-mixed population assumption would be too coarse to meaningfully represent the intervention being studied.

Sensitivity Analysis for Mechanistic Model Parameters

Because compartmental model parameters (Section 13.4) are typically estimated with meaningful uncertainty rather than known exactly, a complete mechanistic modelling exercise should include a formal sensitivity analysis: systematically varying each key parameter (transmission rate, incubation period, recovery period) across its plausible range, one at a time or in combination via Latin hypercube sampling, and reporting how much the key outputs of interest (peak size, peak timing, final epidemic size) change in response. A finding that is robust across the full plausible parameter range (for example, 'the intervention reduces peak size by at least 50% under every combination of parameters we consider plausible') is a far stronger basis for a policy

recommendation than a single point estimate computed under one specific, possibly fragile, parameter assumption, and presenting this sensitivity range transparently is considered standard good practice in applied infectious disease modelling.

Network Models of Transmission

Beyond Homogeneous and Compartmental Mixing

The SEIR framework of Section 13.2 assumes homogeneous mixing — every susceptible individual is equally likely to contact every infectious individual — an assumption that is a reasonable first approximation for some diseases but a poor one for others where transmission occurs along specific, structured contact networks (household contacts, sexual networks relevant to HIV transmission, or specific occupational contact patterns). Network-based transmission models instead represent the population as a graph of individuals connected by contact edges, and simulate transmission spreading along this explicit network structure rather than through random homogeneous mixing, capturing phenomena — such as the disproportionate role of highly-connected ‘super-spreader’ individuals, or the protective effect of network structure that limits how far an early cluster of cases can spread — that a compartmental model built on the homogeneous-mixing assumption cannot represent.

Modernising the SEIR Implementation with `solve_ivp`

The SEIR implementations in Sections 13.4 and 13.6 use `deSolve::ode()` and `scipy.integrate.odeint`, both mature and reliable, but SciPy’s newer `solve_ivp` interface offers adaptive step-size control, a choice of integration methods, and built-in dense output that make it the more robust default for a stiff or fast-changing epidemic curve (such as one with a very high transmission rate), where a fixed-step method can silently produce an inaccurate trajectory without any visible error message.

PYTHON example

```
# Python: SEIR with solve_ivp, offering adaptive step-size control over the fixed-step odeint
from scipy.integrate import solve_ivp
import numpy as np

def seir_rhs(t, y, beta, sigma, gamma, N):
    S, E, I, R = y
    return [-beta*S*I/N, beta*S*I/N - sigma*E, sigma*E - gamma*I, gamma*I]

N = 100000
sol = solve_ivp(seir_rhs, t_span=[0, 180], y0=[N-10, 0, 10, 0], args=(0.30, 1/5, 1/10, N),
                method="RK45", dense_output=True, rtol=1e-6, atol=1e-6)

t_smooth = np.linspace(0, 180, 500)
I_smooth = sol.sol(t_smooth)[2]
print(f"Peak infectious (adaptive solver): {I_smooth.max():.0f} on day
{t_smooth[I_smooth.argmax():.0f}]")
```

Chapter Summary

- Compartmental (SIR/SEIR) models encode explicit transmission assumptions and can answer counterfactual ‘what-if’ questions that statistical models cannot.
- The basic reproduction number R_0 governs whether an outbreak grows or declines and emerges from the model’s transmission and recovery parameters.
- Vector-borne (malaria) and vaccine-preventable disease models require extensions — human-vector coupling, or a vaccinated/hospitalised compartment (SVEIHR) — connecting model output directly to health-system capacity planning.
- Model parameters are typically estimated by fitting predicted trajectories to surveillance data; mechanistic and statistical forecasts answer complementary questions.
- Nonlinear growth curve models (logistic, Gompertz, hyperbolic) summarise decelerating growth trajectories with a small number of interpretable parameters.

- Comparing SEIR scenarios under different transmission-rate assumptions is a direct, quantitative way to communicate the projected impact of a proposed intervention.

Mini-Project

Mini-Project 13: Scenario Modelling for an Outbreak Response Brief

Build an SEIR model for a hypothetical or real outbreak and use it to compare at least two intervention scenarios for a decision-maker briefing.

- Step 1.** Choose a disease and set plausible SEIR parameters (transmission rate, incubation period, recovery period) based on published values or reasonable assumptions, stating your sources or reasoning.
- Step 2.** Simulate the baseline (no intervention) trajectory and report the projected peak infectious count and its timing.
- Step 3.** Simulate at least one intervention scenario (e.g., a reduction in transmission rate from case isolation, or a vaccination campaign reducing the susceptible pool) and compare peak size and timing to baseline.
- Step 4.** Separately, fit a logistic growth curve (Section 13.7) to a simulated or real short outbreak time series and report the three fitted parameters.
- Step 5.** Write a one-page briefing note, in plain language suitable for a non-technical decision-maker, summarising your scenario comparison and its practical implications.

Deliverable: *SEIR simulation code and output for at least two scenarios, a fitted logistic growth curve, and a one-page briefing note.*

End-of-Chapter Exercises

1. Explain, in terms of R_0 , what it means for an outbreak to be ‘controlled’ by an intervention that reduces the effective transmission rate.
2. Why does modelling malaria transmission require a vector compartment, unlike modelling measles transmission?
3. Contrast what a fitted SEIR model and a fitted SARIMA model (Chapter 10) can each tell you about the likely effect of introducing a new vaccine midway through an outbreak.
4. Using the worked example in Section 13.6, calculate the approximate percentage reduction in R_0 achieved by the

intervention scenario, and comment on whether this alone would be enough to push R_0 below 1.

5. In Section 13.7, explain what the fitted parameter $t_0 = 6.2$ tells a health programme manager about the timing of the outbreak's fastest growth phase.
6. Using the Python code in Section 13.6, describe (in words) how you would modify the model to represent a vaccination campaign that moves a fraction of the susceptible population directly into the recovered compartment at day 30.
7. A colleague argues that because the SEIR model's parameters are 'just assumptions,' its output cannot be trusted for policy decisions. Write a short response explaining the value of scenario comparison even when parameters carry uncertainty.
8. Using Section 13.8, explain why a stochastic model might show an outbreak fizzling out even when R_0 is estimated to be slightly above 1.
9. In what stage of an outbreak (early, middle, or late) is the distinction between stochastic and deterministic modelling most practically important, and why?
10. Describe, in outline, how you would modify the SEIR code in Section 13.4 to introduce random variation into the number of new infections at each time step (a basic step toward a stochastic model).
11. In the Section 13.9 case study, explain why the team investigated the discrepancy between the projected and observed peak rather than either ignoring it or immediately doubting the whole model.
12. Why is it important that the team did NOT quietly revise their original projection after the fact to make it look more accurate?

13. What specific piece of real-world information (beyond the case count data itself) explained most of the discrepancy in Section 13.9?
14. How would you use this retrospective validation result to argue for earlier intervention timing in a future, similar outbreak?
15. Explain, in your own words, why a prospective forecast (made before the outcome is known) is a more rigorous test of a model than checking a model's fit to data it was trained on.
16. Calculate the herd immunity threshold for a disease with $R_0 = 5$, using the formula in Section 13.10.
17. Explain why the herd immunity threshold formula assumes homogeneous mixing, and why real populations often violate this assumption.
18. Why can a population above the calculated herd immunity threshold still experience a limited, self-terminating outbreak following an imported case?
19. Explain, using Section 13.10, why a 'local pocket' of low vaccination coverage can sustain measles transmission even when a state's overall average coverage is high.
20. Why is $R_0 = 15$ for measles associated with such a high (93%) herd immunity threshold, compared to a disease with a lower R_0 ?
21. Using Section 13.11, explain one thing an agent-based model can represent that a compartmental (SEIR) model cannot.
22. Why are agent-based models not the default choice for most routine forecasting tasks, according to Section 13.11?

23. Propose a specific Nigerian public health scenario where an agent-based model's ability to represent individual contact networks would be particularly valuable.
24. What additional data would an agent-based model typically require that a compartmental model does not?
25. Explain why 'ring vaccination around a specific detected case' is an example of an intervention better suited to agent-based than compartmental modelling.
26. Summarise, in your own words, why this chapter treats mechanistic and statistical models as complementary rather than competing approaches.
27. Explain the common thread linking the herd immunity threshold discussion (Section 13.10) and the agent-based models discussion (Section 13.11).
28. Why does this chapter emphasise retrospectively validating a model against what actually happened (Section 13.9), rather than treating a model's initial projections as the final word?
29. Propose a scenario from your own field of interest where an agent-based model might capture something a compartmental SEIR model would miss.
30. If you had to explain the practical limits of the herd immunity threshold concept to a policymaker, what would you say?
31. Using Section 13.13, explain why the homogeneous mixing assumption in a standard SEIR model might be a poor fit for HIV transmission.
32. Describe what a 'super-spreader' individual represents in a network transmission model, and why a compartmental model cannot represent this concept directly.

33. Propose a Nigerian disease context where network structure (household, occupational, or social) would likely matter a great deal for transmission dynamics.
34. What additional data would be needed to build a network transmission model that would not be needed for a standard SEIR model?
35. How might the protective effect of network structure described in Section 13.13 relate to the herd immunity threshold discussed earlier (Section 13.10)?