**TRABAJO FIN DE MÁSTER EN BIOESTADÍSTICA**

# Simulación de ensayos clínicos de deprescripción

**SEPTIEMBRE 2025**

Florencia Inés Aiello Battan

Tutor: Teresa Pérez

# SIMULACIÓN DE ENSAYOS CLÍNICOS DE DEPRESCRIPCIÓN

Florencia Aiello Battan

# Declaración Responsable sobre Autoría y Uso Ético de Herramientas de Inteligencia Artificial (IA)

Yo, AIELLO BATTAN FLORENCIA

INÉS Con DNI/NIE/PASAPORTE:

Z2073794G

declaro de manera responsable que el presente Trabajo de Fin de Máster (TFM)

Titulado/a

**Simulación de ensayos clínicos de deprescripción**

es el resultado de mi trabajo intelectual personal y creativo, y ha sido elaborado de acuerdo con los principios éticos y las normas de integridad vigentes en la comunidad académica y, más específicamente, en la Universidad Complutense de Madrid.

Soy, pues, autor del material aquí incluido y, cuando no ha sido así y he tomado el material de otra fuente, lo he citado o bien he declarado su procedencia de forma clara -incluidas, en su caso, herramientas de inteligencia artificial-. Las ideas y aportaciones principales incluidas en este trabajo, y que acreditan la adquisición de competencias, son mías y no proceden de otras fuentes o han sido reescritas usando material de otras fuentes.

Asimismo, aseguro que los datos y recursos utilizados son legítimos, verificables y han sido obtenidos de fuentes confiables y autorizadas. Además, he tomado medidas para garantizar la confidencialidad y privacidad de los datos utilizados, evitando cualquier tipo de sesgo o discriminación injusta en el tratamiento de la información.

En Madrid a 07/09/2025

FIRMA

# Table of contents

# Tables



# Figures

# Resumen

Los ensayos clínicos aleatorizados sobre deprescripción son complejos de realizar por la heterogeneidad de las intervenciones, la inconsistencia en los resultados y la susceptibilidad al sesgo. Proponemos un marco basado en simulaciones para emular ensayos objetivo y evaluar los efectos de la deprescripción. El enfoque integra tres componentes: *(i) el diseño del ensayo objetivo*, que define elegibilidad, intervenciones, resultados y seguimiento según principios de inferencia causal; *(ii) la generación de poblaciones sintéticas* a partir de distribuciones multivariantes informadas por ensayos previos para preservar correlaciones empíricas; y *(iii) la modelización del tiempo hasta el evento* mediante modelos de supervivencia paramétricos en una estructura multi estado semi-Markov. Esto permite transiciones dinámicas entre deprescripción, eventos adversos y muerte.

La evaluación metodológica reveló dos conclusiones clave. Primero, es esencial especificar los riesgos particulares de cada resultado. Segundo, la fiabilidad inferencial depende del tamaño de la muestra: una especificación correcta proporciona estimaciones estables en cohortes pequeñas, mientras que una especificación errónea en cohortes grandes provoca un sesgo amplificado y una pérdida de cobertura.

Este marco refuerza la inferencia causal basada en simulaciones y ofrece un enfoque reproducible y extensible para investigar deprescripción y otros contextos en los que los ensayos aleatorios no son viables.

**Palabras claves**: desprescripción, emulación de ensayos clínicos, inferencia basada en simulación, modelos multiestado, modelos paramétricos de supervivencia, inferencia causal.

# Abstract

Randomized deprescribing trials are challenging due to heterogeneous interventions, inconsistent outcome definitions, and vulnerability to bias. We propose a simulation-based framework for emulating target trials to evaluate the effects of deprescribing. The framework integrates three key elements: *(i) the target trial design*, which specifies eligibility, interventions, outcomes, and follow-up to replicate causal inference principles; *(ii) generation of synthetic population data* from multivariate distributions informed by previous clinical essays to preserve empirical correlations; and *(iii) time-to-event modelling* with covariate-adjusted parametric survival models embedded in a semi-Markov multistate structure. This enables dynamic transitions between discontinuation, adverse events, and death, with event risks adapting to prior history.

The methodological evaluation revealed two key findings. Firstly, outcome-specific hazard specification is essential. Gompertz models were more closely aligned with stroke and bleeding, whereas Weibull models better represented mortality. Secondly, inferential reliability depended on sample size; correct specification yielded stable estimates in small cohorts, whereas misspecification in large cohorts resulted in amplified bias and loss of coverage.

This framework strengthens simulation-based causal inference by combining target trial emulation with flexible hazard modelling. It provides a reproducible, extensible approach for deprescribing research and other contexts where randomized trials are infeasible.

**Keywords**: deprescribing, target trial emulation, simulation-based inference, multistate models, parametric survival models, causal inference.

# 1. Introduction

Polypharmacy, defined as the use of five or more drugs or the prescription of unnecessary drugs, is associated with an increased risk of adverse events. The effects of ageing on pharmacokinetics and pharmacodynamics, together with the high prevalence of polypharmacy, increase the risk of adverse effects in older adults (1,2). In Spain, the prevalence of polypharmacy is between 34.2 and 40.8% (3). In the international literature, the prevalence of polypharmacy is estimated to affect one third of those over 65 and half of those over 85 years of age (1).

While the potential harms of polypharmacy have been documented, the benefits of deprescribing remain unclear due to the difficulty in studying them (4). Deprescribing is a supervised and systematic process of withdrawing, reducing or replacing inappropriate medications in order to minimize adverse effects and enhance benefits. It includes the withdrawal of chronic medications that lack benefit, or that have habitually lost beneficial effect due to ageing of the person (1,3,5). Assessing the effects of stopping chronic treatment is crucial, considering both potential benefits and risks. The aim of the deprescribing process is not only to reduce adverse effects but also to maintain or even improve quality of life (6).

Despite the growing interest in deprescribing, its benefits remain still uncertain. Several types of studies have been proposed to analyse the effects of deprescribing. On the one hand, preclinical and clinical studies that are representative of the multimorbid older adult population are required to assess the benefits of medication in this population (2).

Randomized clinical trials to evaluate the effect of deprescribing show large heterogeneity (7). Study interventions include medication review, provider education (8), patient education, or computerized decision support (9,10). The medication review process may involve discontinuation or dose adjustments, with adherence varying significantly across studies (11) (12). In some studies, the intervention is focused in one specific drug class (13,14). Study populations differ largely including older patients in community settings (11,12), post acute care (15,16) or inpatients (17). Outcomes definitions are mostly centred around drug burden or counts of potentially inappropriate medications (8,15) and seldomly on clinical outcomes. The majority of trials that explore changes in quality of life, admission rate, emergency visits rate, adverse events, falls, or survival include them as secondary end-points (12). Few studies

examine adverse drug effects (ADEs) or adverse drug withdrawal effects (ADWEs) (11). Follow up time is diverse ranging from 6 weeks to 18 months (18).

Deprescribing interventions are often successful in reducing the use of potentially inappropriate medications or medication burden but fail to demonstrate clinical benefits. In the Shed-MEDS Randomized Clinical Trial a patient-centered deprescribing intervention was effective in reducing the overall medication burden but no differences were found in ADEs or ADWEs between groups (15). A randomized clinical trial in older inpatients showed no differences in all-cause death, unscheduled hospital visits, or rehospitalization between usual care and medication optimization protocol after 12 months. Total medication counts and potentially inappropriate medications were reduced in the intervention group with no excess adverse events observed (17). A pragmatic randomized control trial found statistically significant improvements on health-related quality of life and mortality after 4 months of the medication review, but that these effects were no longer statistically significant after 13 months (12).

Four systematic reviews and meta-analyses of deprescribing clinical trials consistently describe the low or very low quality of evidence and a large heterogeneity between studies (7,10,18,19). Few studies explored clinical outcomes as the primary end-point (7). The quality of current studies was found to be rather weak mainly due to unclear definitions of polypharmacy, poorly defined and complex interventions, and limited reproducibility (7,18).

One systematic review found no effect on quality of life and minimal effects on clinical outcomes on a short term period of 3 months (19). Another systematic review found that medication review may reduce mortality but did not reduce falls, hospitalizations, or modified health-related quality of life (10). A beneficial effect on clinical outcomes has not been consistently demonstrated in clinical trials. It is unclear whether deprescribing interventions result in any clinically significant improvement (7). Must be taken into account that most studies are not limited to a specific drug class which leads to only reducing low-risk medications with little or no impact on clinical relevant outcomes (16,17).

Another type of analysis has been population-based cost-effectiveness studies. Effectiveness is analysed in units of quality-adjusted life years (QALYs). For proton pump inhibitors, a Hong Kong study found that deprescribing them resulted in lower costs (savings of USD 325 per medication withdrawn) and higher QALYs (0.0249 per medication withdrawn) (5). For antihypertensives in older adults, a UK study found that reducing medication resulted in

lower costs but also lower QALYs due to their association with cardiovascular events. Standard treatment demonstrated a better cost-effectiveness ratio than medication reduction in this population (20).

Due to the large variability between subjects, medications to be evaluated, comorbidities, and clinical scenarios, predicting the benefits or risks of deprescribing is complex. Single variable analyses are not sufficient to adequately estimate the effects of deprescribing. Given the challenges in conducting high-quality deprescribing trials and the heterogeneity in study designs and patient populations, alternative approaches are necessary.

Simulation techniques and multivariate modeling can overcome the limitations of clinical trials. These models provide a structured way to estimate deprescribing outcomes while incorporating multiple variables, allowing to test different clinical scenarios and long-term effects.

To ensure the reliability of the model, its validity will be assessed using simulation techniques, allowing for the quantification of model performance.

# 2. Objectives

The primary objective of this study is to evaluate the potential risks and benefits of deprescribing through a simulation-based framework. We simulate multivariate, time-to-event data to emulate clinical trial conditions and assess the impact of medication discontinuation. The secondary objectives are to explore how different assumptions such as parametric distributions and sample sizes influence the resulting outcome estimates.

# 3. Methodology

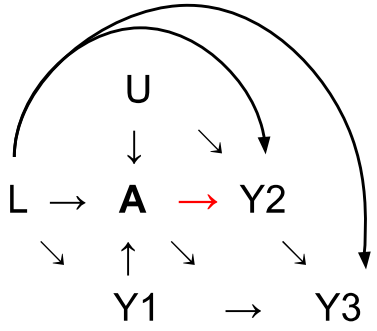
We conducted a review of simulation studies in pharmacoepidemiology that investigated inferential methods within controlled settings. A recurrent limitation was the lack of code availability, which substantially constrained the reproducibility and independent verification of findings. Furthermore, the methodological complexity of most simulations was limited, as relatively few incorporated time-dependent covariates or effects. Although the scope of the review was restricted to a single journal, the number of identified simulation studies was notably small. These observations suggest that the field would benefit from greater emphasis

on methodologically rigorous simulation studies, designed with a level of complexity that more accurately reflects clinical realities, to enhance the evaluation of inferential approaches. These findings are reported in detail in a preprint of the review. (21)

## 3.1 Causal Framework Using a Directed Acyclic Graph (DAG)

Deprescribing is a dynamic process influenced by patient characteristics, clinical decisions, and ADEs. To ensure appropriate adjustment for confounding, we employ a DAG that depicts the relationships between deprescribing (A), ADEs (Y1), ADWEs (Y2), and relevant confounders (Figure 1).

Figure 1. Directed Acyclic Graph




where:

- **A (Exposure):** Drug discontinuation.
- **Y1 (ADEs):** Adverse drug events (e.g., major bleeding event).
- **Y2 (ADWEs):** Ischemic stroke after discontinuation.
- **Y3 (Death):** Death.
- **L (Measured Confounders):** Age, comorbidities.
- **U (Unmeasured Confounders):** Physician preference, medication adherence

In this structure, confounders (L and U) influence both deprescribing (A) and clinical outcomes (Y2), requiring proper adjustment. ADEs (Y1) can also lead to deprescribing, creating potential selection bias. The red arrow represents the estimand of interest.

## 3.2 Target trial emulation

To evaluate the potential effects of deprescribing, we followed the structured target trial emulation framework proposed by Hernán and Robins (22). This methodology treats observational analyses as efforts to replicate a hypothetical randomized trial designed to answer the research question. It specifies key elements such as eligibility criteria, treatment strategies, assignment procedures, follow-up, and outcomes, providing a clear blueprint for study design and analysis. By framing our study as a target trial emulation, we aimed to reduce biases commonly encountered in nonrandomized research and improve causal interpretability. Additionally, we incorporated explanatory modeling and simulation-based causal inference to explore treatment effects under varying assumptions.

As a case study, we will focus on anticoagulation therapy in older adults with subclinical atrial fibrillation. This clinical scenario serves as an ideal example due to its high prevalence in this population, the extensive clinical data available, and the well-defined risks and benefits associated with anticoagulation therapy. Anticoagulation discontinuation carries measurable adverse drug withdrawal effects (e.g., stroke) and adverse drug events (e.g., bleeding), making it a clinically relevant and well-suited candidate for evaluating deprescribing strategies. However, the ultimate goal is to create a generalizable and flexible approach that can be adapted to different populations, treatment classes, and deprescribing scenarios. By ensuring adaptability, this approach can serve as a foundation for future deprescribing research across a broader range of clinical contexts.

This methodology aims to address the limitations of traditional deprescribing studies by leveraging simulation techniques to estimate clinical outcomes while maintaining methodological rigor. By integrating real-world clinical considerations into an explicative modeling framework, this study will provide valuable insights into the potential impact of deprescribing interventions.

The overall methodological steps are outlined below and detailed in the subsequent subsections:

1. Specification of a target trial protocol.
2. Generation of a synthetic population.
3. Simulation of time-to-event outcomes using parametric survival models.

4. Emulation of treatment strategies: continued vs. deprescribing.
5. Validation and performance assessment.
6. Application of multistate models for causal effect estimation.
7. Sensitivity analyses.

The simulation design follows the ADEMP framework (Aims, Data-generating mechanisms, Estimands, Methods, Performance measures) as recommended by Morris et al. (23) to ensure transparency, reproducibility, and pedagogical clarity in simulation-based evaluation of statistical methods (Table 1).

Table 1. ADEMP framework

| | |
|---|---|
| **Aims** | Estimate the causal effects of deprescribing anticoagulation on stroke, bleeding, and mortality. |
| **Data generating mechanism** | Synthetic population data: Simulated using multivariate normal distributions with correlations derived from clinical trials.<br>Time to event data: Time-to-event outcomes were simulated using parametric distributions. |
| **Estimands** | Hazard ratios for stroke, bleeding, and death comparing deprescribing vs. continuation.<br>State transition probabilities in the multistate model |
| **Methods** | Target Trial Emulation<br>Multistate model |
| **Performance measures** | Bias and coverage of HR.<br>Monte Carlo standard errors.<br>Bootstrap validation. |

### 3.2.1 Target trial protocol

We designed a target trial protocol that serves as a blueprint for an ideal randomized controlled trial, ensuring that our simulated study aligns with real-world clinical decision-making.

Our clinical question of interest was to evaluate the efficacy and safety of deprescribing direct-acting oral anticoagulants compared to continued treatment in older adults with subclinical atrial fibrillation.

We consider community-dwelling older adults (≥ 65 years) with subclinical atrial fibrillation and a $CHA_2DS_2$-VASc score of ≥ 2, currently receiving direct-acting oral anticoagulants. Exclusion criteria include other indications for anticoagulation (e.g., deep vein thrombosis), contraindications for direct-acting oral anticoagulants, dual antiplatelet therapy, and a history of clinical atrial fibrillation.

The intervention is defined as abrupt deprescribing direct-acting oral anticoagulants, with follow-up beginning at the start of anticoagulation therapy. The comparator group consists of patients who continue anticoagulation without modification. Patients are followed until death or end of follow-up. Maximum follow-up period will be 10 years.

The primary outcome is time to ischemic stroke. Secondary outcomes include time to all-cause death and time to the first major bleeding event, defined according to the International Society on Thrombosis and Haemostasis criteria. (24)

Baseline characteristics include age, sex, body mass index, $CHA_2DS_2VASc$ score, estimated creatinine clearance, antiplatelet use, history of stroke or transient ischemic attack, diabetes mellitus, arterial hypertension, heart failure, and vascular disease (coronary artery disease, peripheral arterial disease, aortic plaques, or cerebrovascular disease).

The target trial protocol is presented in Table 2.

Table 2. Target trial protocol.

| Protocol element | Description |
|---|---|
| Study population | ● Age ≥ 65 years<br>● At least one of the following cardiovascular conditions leading to $CHA_2DS_2VASc$ score ≥ 2:<br>○ Age ≥ 75 years,<br>○ heart failure (clinically overt or left ventricular ejection fraction < 45%),<br>○ arterial hypertension,<br>○ diabetes mellitus,<br>○ prior stroke or transient ischemic attack,<br>○ vascular disease (peripheral, carotid/cerebral, or aortic plaques<br>○ on transesophageal echocardiogram).<br>● Subclinical atrial fibrillation: Atrial fibrillation detected by an implanted device.<br><br>Patients with history of clinical atrial fibrillation (clinical symptoms or detection by surface ECG), other indications for |

| | |
|---|---|
| | anticoagulation therapy, contraindication for direct-acting oral anticoagulants, on dual-antiplatelet therapy were excluded. |
| Intervention | Abrupt discontinuation of direct-acting oral anticoagulant. Comparator is maintaining treatment with direct-acting oral anticoagulants. |
| Intervention assignment | Randomisation |
| Primary outcome | Time from the start of anticoagulation therapy to ischemic stroke. |
| Secondary outcomes | • Time to first major bleeding event according to International Society on Thrombosis and Haemostasis criteria<br>• Time to all-cause death. |
| Start and end of follow-up | Patients are randomized after one year of anticoagulation therapy and followed until death or the end of the follow-up period. |
| Causal estimand | The main aim is to test the alternative hypothesis that the hazard rate is identical in the two groups.<br>Per protocol effect |
| Statistical analysis | Kaplan-Meier curves for event-free survival will be constructed and compared with the log-rank test. Hazard ratios with 95% confidence intervals will be calculated with the use of Cox proportional hazards model. |

### 3.2.2 Generation of a simulated population

We implemented simulations to generate synthetic data based on realistic assumptions derived from the NOAH-AFNET 6 (25) and ARTESIA (26) trials.

Firstly, we simulated baseline characteristics for older adults (≥ 65 years) with subclinical atrial fibrillation. Marginal distributions were extracted separately for each trial. These were proportions for binary variables. Means with standard deviations for continuous variables were also extracted. A full list of the simulated covariates is provided in Table 3.

Table 3. Covariates.

| **Baseline Covariates** | **Type** |
|---|---|
| Age | discrete |
| Sex | binary (coded as 1= female, 0 = male) |
| Estimated creatinine clearance | continuous |
| Antiplatelet use | binary (coded as1 = yes; 0 = no) |
| History of stroke or transient ischemic attack | binary (coded as1 = yes; 0 = no) |
| History of coronary arterial disease | binary (coded as1 = yes; 0 = no) |
| Diabetes mellitus | binary (coded as1 = yes; 0 = no) |
| Arterial hypertension | binary (coded as1 = yes; 0 = no) |
| Heart failure | binary (coded as1 = yes; 0 = no) |

Secondly, to generate the data, we simulated multivariate normal variables that reflect the specified correlations among covariates. Binary variables were derived by applying quantile thresholds corresponding to the observed marginal proportions. Continuous variables (age and estimated creatinine clearance) were generated from standardized normal values and scaled to match their observed distributions. A manually specified correlation matrix incorporated weak-to-moderate associations among variables, including strong correlations between antiplatelet use and both stroke history and coronary artery disease, as well as moderate correlation with diabetes. As the initial correlation matrix was not positive definite, it was adjusted by finding the closest positive definite matrix to ensure valid multivariate sampling could be performed. This process was implemented in R as illustrated below. The same process was repeated for the NOAH trial data and then merged into a single dataset using the same correlation structure. $CHA_2DS_2$-VASc scores were computed for each individual.

```
#Baseline characteristics--------
#Simulated data using Artesia trial data for marginal
distributions
set.seed(1234)
n <- 30000
```

```r
binary_probs <- c(
  sex = 0.361, #Female = 1
  htn = 0.815,
  cad = 0.37,
  dbt = 0.291,
  hf = 0.283,
  hstroke = 0.09,
  antiplatelet = 0.574
)

normal_means <- c(age = 76.8, creat = 71.4)
normal_sds   <- c(age = 7.6,  creat = 28.7)

var_names <- c(names(binary_probs), names(normal_means))
p <- length(var_names)

#Correlation matrix
correlation_matrix <- matrix(c(
  1.00, 0.30, 0.25, 0.20, 0.30, 0.15, 0.10, 0.25, 0.10,  # sex
  0.30, 1.00, 0.35, 0.25, 0.40, 0.20, 0.20, 0.30, 0.15,  # htn
  0.25, 0.35, 1.00, 0.30, 0.35, 0.25, 0.80, 0.30, 0.10,  # cad
  0.20, 0.25, 0.30, 1.00, 0.30, 0.20, 0.50, 0.25, 0.25,  # dbt
  0.30, 0.40, 0.35, 0.30, 1.00, 0.25, 0.35, 0.35, 0.30,  # hf
  0.15, 0.20, 0.25, 0.20, 0.25, 1.00, 0.90, 0.20, 0.10,  # hstroke
  0.10, 0.20, 0.80, 0.50, 0.35, 0.90, 1.00, 0.30, 0.15,
# antiplatelet  Strong correlation with hstroke and cad
  0.25, 0.30, 0.30, 0.25, 0.35, 0.20, 0.30, 1.00, 0.35,  # age
  0.10, 0.15, 0.10, 0.25, 0.30, 0.10, 0.15, 0.35, 1.00   # creat
), nrow = p, byrow = TRUE)

colnames(correlation_matrix) <- rownames(correlation_matrix) <-
var_names

# Fix correlation matrix if not positive definite
correlation_matrix <- as.matrix(nearPD(correlation_matrix, corr =
TRUE)$mat)

#Generate multivariate normal data
Z <- mvrnorm(n = n, mu = rep(0, p), Sigma = correlation_matrix)

#Convert binary variables
```

```
binary_names <- names(binary_probs)
thresholds <- qnorm(1 - binary_probs)

binary_data <- sapply(seq_along(binary_probs), function(i) {
  as.numeric(Z[, i] > thresholds[i])
})
colnames(binary_data) <- binary_names

#Scale normal variables to desired mean and sd
normal_data <- Z[, (length(binary_probs) + 1):p]
normal_data <- scale(normal_data)  # standardize first
normal_data <- sweep(normal_data, 2, normal_sds, `*`)  # multiply
by sd
normal_data <- sweep(normal_data, 2, normal_means, `+`)  # add
mean
colnames(normal_data) <- names(normal_means)

#Combine into a single dataset
sim_data <- data.frame(binary_data, normal_data)

#Compare simulated summary to real artesia summary
sim_artesia_summary <- sim_data |>
  summarise(
    age = mean(age),
    sex = mean(sex),
    htn = mean(htn),
    cad = mean(cad),
    dbt = mean(dbt),
    hf = mean(hf),
    hstroke = mean(hstroke),
    antiplatelet = mean(antiplatelet),
    creat = mean(creat)
  ) |>
  pivot_longer(cols = everything(), names_to = "variable",
values_to = "sim_mean") |>
  left_join(real_artesia, by = "variable")

#Merge data
dataset <- bind_rows(sim_data,sim_data_noah, .id = "trial" ) |>
mutate(id = row_number()) #Trial = 1 artesia, 2 = noah

#Add CHADSVASC score------
```

```r
#CHADSVASC score function
chadsvasc <- function(age,sex,hf,htn,hstroke,cad,dbt){
  score <- 0
  score <- score + ifelse(age >=75,2,0)
  score <- score + ifelse(age > 65 & age < 75,1,0)
  score <- score + ifelse(sex == 1, 1, 0)
  score <- score + ifelse(htn == 1, 1, 0)
  score <- score + ifelse(hf == 1, 1, 0)
  score <- score + ifelse(dbt == 1, 1, 0)
  score <- score + ifelse(hstroke == 1, 2, 0)
  score <- score + ifelse(cad == 1, 1, 0)
  return(score)
}

dataset <- dataset |> mutate(chads =
chadsvasc(age,sex,hf,htn,hstroke,cad,dbt))
```

### 3.2.3 Simulation of Time-to-Event Outcomes

We simulated time-to-event data for each individual for the following events: ischemic stroke, major bleeding, and all-cause death. Time-to-event outcomes were simulated using Gompertz distributions, which best fit the original trial data based on graphical assessments and the Anderson-Darling goodness of fit test. Kaplan-Meier curves for stroke, major bleeding, and all-cause death were digitized from the trials and used to estimate Gompertz shape and rate parameters. The estimated parameters used in the simulation are summarized in Table 4. Each simulation included an event time, a censoring indicator, and the associated baseline covariates. Survival analysis and parametric models followed recommendations from Collett. (27)

Table 4. Gomperz parameters.

Table 4. Gompertz Parameters for Simulated Time-to-Event Outcomes

| Event | Shape | Rate |
|---|---|---|
| Ischemic stroke | 0.452 | 0.102 |
| Major bleeding and All-cause death | 0.566 | 0.108 |

The following code snippet demonstrates how this step was carried out for ischemic stroke. The same procedure was repeated for the other events. Gompertz distribution was the best fit for each event.

```
#Fitting distribution for time-to event data
#Outcome: ischemic stroke
#Time in years, cumulative incidence of stroke (%)
km1 <- survfit(Surv(years,status)~ 1, data = stroke)

#Plot log(time), log(-log(S))
S <- km1$surv
Y <- -log(-log(S))
logT <- -log(km1$time)
plot(Y ~ logT, xlab = "logTime", ylab = "log(-log(S))")

#Plot log(time), -log(S/1-S)
Z<- -log(S/(1- S))
plot(Z ~ logT, xlab = "logTime", ylab = "-log(S/1-S)")

#Estimate hazard function
result <- muhaz (stroke$years, stroke$status, bw.method= "global")
plot (result)   #Hazard rate increases over time

# Fit distribution
fit_weibull <- flexsurvreg(Surv(years, status) ~ 1, dist =
"weibull", data = stroke)
fit_gompertz <- flexsurvreg(Surv(years, status) ~ 1, dist =
"gompertz", data = stroke)

# Compare AIC values
AIC(fit_weibull, fit_gompertz)

#Compare Plots
plot(km1, col = "red", conf.int = FALSE, main="Fitted Gompertz vs
KM", xlab = "Time(years", ylab = "Survival probability")
lines(fit_gompertz, type = "survival", col = "blue")
legend("bottomleft", legend = c("KM", "Gompertz"), col = c("red",
"blue"), lty = 1)

plot(km1, col = "red", conf.int = FALSE, main="Fitted Weibull vs
KM", xlab = "Time(years", ylab = "Survival probability")
```

```
lines(fit_weibull, type = "survival", col = "blue")
legend("bottomleft", legend = c("KM", "Weibull"), col = c("red",
"blue"), lty = 1)

#Best fit: Gompertz distribution

x <- stroke$years

# Define the log likelihood function to be used t be maximized.
logLik=function(param) {
  a=param[1]
  t=param[2]
  ll = log(t) + a*x + (t/a)*(1-exp(a*x))
  ll
}

# Perform the maximum likelihood estimation
res=maxLik(logLik,start=c(0.00121307,0.00173329))
# Estimates of the parameters a and t
aVal=res$estimate[1]
tVal=res$estimate[2]
aVal;tVal

#Anderson-Darling test
gpzFn=function(z) {
  F =  1 - exp(-(tVal/aVal)*(exp(aVal*z)-1))
  F
}

goftest::ad.test(x, null=gpzFn)
```

For the first simulated cohort, Gompertz rate and shape parameters were estimated on the active treatment arms (apixaban or edoxaban) data. Hazard rates were further adjusted at the individual level using log-hazard ratios for each covariate, derived from Friberg et al. (28) This enabled covariate-informed scaling of event risks.

Age and creatinine clearance were mean-centered to improve numerical stability in the linear predictor calculations. For each outcome, a separate linear predictor was computed by multiplying the individual's covariates by corresponding log-hazard ratios stored in predefined vectors. These linear predictors were then exponentiated and multiplied by the

baseline hazard rate for each outcome, effectively scaling the event rate for each individual. Finally, survival times were sampled using the rgompertz function with the individual-specific hazard rates and shape parameters, resulting in simulated times to stroke, bleeding, and death.

```
#Generate time to event data---
datasim <- dataset |>
  mutate(
    age_c = age - 77, #Center age
    creat_c = creat - 69, #Center creatinine
    # Linear predictors
    lp_stroke = age_c * log_hr_stroke["age"] + sex *
log_hr_stroke["sex"] +
      htn * log_hr_stroke["htn"] + cad * log_hr_stroke["cad"] +
      dbt * log_hr_stroke["dbt"] + hf * log_hr_stroke["hf"] +
      hstroke * log_hr_stroke["hstroke"] + antiplatelet *
log_hr_stroke["antiplatelet"] +
      creat_c * log_hr_stroke["creat"],

    lp_death = age_c * log_hr_death["age"] + sex *
log_hr_death["sex"] +
      htn * log_hr_death["htn"] + cad * log_hr_death["cad"] +
      dbt * log_hr_death["dbt"] + hf * log_hr_death["hf"] +
      hstroke * log_hr_death["hstroke"] + antiplatelet *
log_hr_death["antiplatelet"] +
      creat_c * log_hr_death["creat"],

    lp_bleeding = age_c * log_hr_bleeding["age"] + sex *
log_hr_bleeding["sex"] +
      htn * log_hr_bleeding["htn"] + cad * log_hr_bleeding["cad"]
+
      dbt * log_hr_bleeding["dbt"] + hf * log_hr_bleeding["hf"] +
      hstroke * log_hr_bleeding["hstroke"] + antiplatelet *
log_hr_bleeding["antiplatelet"] +
      creat_c * log_hr_bleeding["creat"],

    #Survival times
    time_stroke = rgompertz(n(), shape = shape_stroke, rate =
rate_stroke * exp(lp_stroke)),
    time_bleeding = rgompertz(n(), shape = shape_b, rate = rate_b*
exp (lp_bleeding)),
```

```
    time_death =  rgompertz(n(), shape = shape_b, rate = rate_b*
exp (lp_death))
  )
```

### 3.2.4 Emulating of Treatment Strategies

Patients experiencing events during the first year were administratively censored at event time, as they failed to meet the protocol-specified eligibility for randomization. Only those who survived event-free through the 1-year eligibility window (no stroke, bleeding, or death) were included in the analysis cohort.

Eligible patients were randomly assigned (1:1) to either continue anticoagulation therapy or undergo deprescribing, defined as abrupt discontinuation of direct oral anticoagulants. While the continuation arm maintained baseline hazard rates throughout follow-up, the deprescribing arm featured time-varying hazards: during the first year, patients retained the active treatment hazard profile, which then transitioned to the control arm profile (derived from pooled hazard ratios of the NOAH-AFNET 6 and ARTESIA trials) following discontinuation. All participants were subject to administrative censoring at 10 years unless death occurred earlier, with death serving as the sole absorbing state. To account for the prognostic impact of non-fatal events, we implemented post-event survival time adjustments - specifically modeling increased mortality risk following stroke (HR = 1.5) and major bleeding events (HR = 2.0) through piecewise hazard modifications in the Gompertz survival framework.

```
#-TIME-DEPENDENT COVARIATE: DISCONTINUE---------------
datasim_1 <- datasim_1 |> mutate(
  id = 1:nrow(datasim_1), #Id
  d = rbinom(nrow(datasim_1), 1, 0.5),  # 50% discontinue
  discontinue_time = ifelse(d  == 1, 1, NA), #1 year on treatment
before discontinuation
  followup = 10, #10 years follow-up
  time_end = ifelse(time_death < followup, time_death, followup))
#Death as an absorbing state

#--Adjust survival times for discontinuation group
datasim_2 <- datasim_1 |> mutate(
  # Survival times
```

```
  time_stroke = ifelse(d == 1,
                       discontinue_time + rgompertz(n(), shape =
shape_stroke, rate = rate_stroke * exp(lp_stroke) * HR_stroke^d),
                       time_stroke),
  time_bleeding = ifelse(d == 1,
                         discontinue_time + rgompertz(n(), shape =
shape_b, rate = rate_b * exp(lp_bleeding) * HR_bleeding^d),
                         time_bleeding),
  time_death = ifelse(d == 1,
                      discontinue_time + rgompertz(n(), shape =
shape_b, rate = rate_b * exp(lp_death) * HR_death^d),
                      time_death),
#-- Adjust survival time after stroke. HR = 1.5
  time_death_poststroke = case_when(
    time_stroke < time_death ~ time_stroke + rgompertz(n(), shape
= shape_b, rate = rate_b * exp(lp_death + log(1.5)) * HR_death^d),
    TRUE ~ NA_real_
  ),
# --Adjust survival time after bleeding. HR = 2
  time_death_postbleeding = case_when(
    time_bleeding < time_death ~ time_bleeding + rgompertz(n(),
shape = shape_b, rate = rate_b * exp(lp_death + log(2)) *
HR_death^d),
    TRUE ~ NA_real_
  ),
#Ajust death time
time_death = case_when(
  !is.na(time_death_poststroke) & !is.na(time_death_postbleeding)
~ pmin(time_death_poststroke, time_death_postbleeding),
  !is.na(time_death_poststroke) ~ time_death_poststroke,
  !is.na(time_death_postbleeding) ~ time_death_postbleeding,
  TRUE ~ time_death
),
# Adjust time_end
  time_end = pmin(time_death, followup)
)
```

### 3.2.5 Validation and Performance Assessment

To validate the simulation, we compared cumulative incidence curves and survival functions from the simulated dataset to those reported in the original trials. The close alignment

confirmed that our approach reliably reproduced the underlying event distributions and hazard dynamics observed in the NOAH-AFNET 6 and ARTESIA trials.

To evaluate the statistical properties of the estimated hazard ratio for discontinue, we conducted a bootstrap simulation study. A total of 1000 bootstrap samples of the original dataset were generated by sampling individuals with replacement. For each bootstrap replicate, Cox models for stroke, bleeding, and death were re-estimated using the same covariate structure.

For each endpoint, we extracted the estimated hazard ratio (HR) and corresponding 95% confidence interval (CI) for discontinuation. Using the simulated estimates, we calculated the empirical bias (mean difference between the estimated and assumed true HR), coverage probability (proportion of 95% CIs containing the true HR), and Monte Carlo standard errors for bias and coverage. The true HR values were based on the point estimates from the original models. (23)

### 3.2.6 Multi-State Model Approach

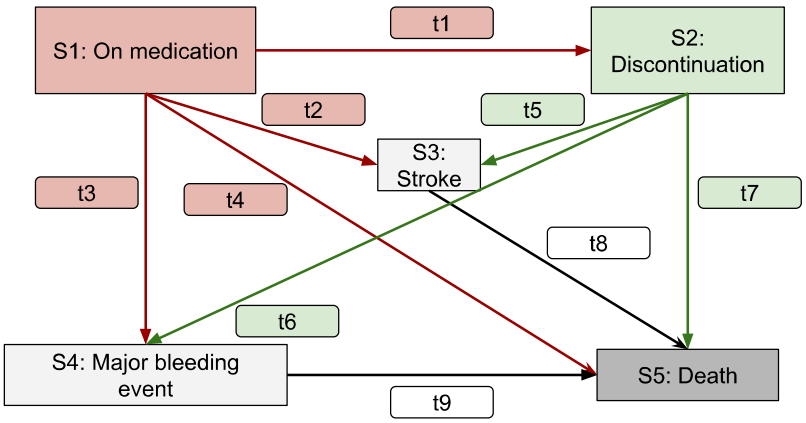
To evaluate the impact of deprescribing anticoagulation therapy in older adults with subclinical atrial fibrillation, we adopt a multi-state modeling approach. This approach allows us to model transitions between clinically meaningful states over time, accounting for competing risks and time-dependent exposures, as described in previous work on multi-state modeling (29–32).

The states of the model will be:

1. **On medication (Baseline State)**: Patients actively receiving direct-acting oral anticoagulants.
2. **Major bleeding Event**: First major bleeding event.
3. **Stroke**: First occurrence of ischemic stroke.
4. **Discontinuation**: Patients who undergo deprescribing.
5. **Death:** Death due to any cause (absorbing state).

Transitions between these states will be modeled using a semi-Markov multi-state model, where transition intensities depend on elapsed time since entering the current state. Figure 2 represents the state-transition diagram.

Figure 2. State-transition diagram.




To estimate the transition intensities between states, we employ a Cox proportional hazards model for each possible transition. This allows the hazard of transitioning from state *j* to state *k* to vary flexibly over time and to depend on both baseline and time-varying covariates, including treatment discontinuation status.

The hazard function for individual *i*, transitioning from state *j* to *k* at time *t*, is defined as:

$$h_{ijk}(t) = y_{ijk}(t)\, h_{0jk}(t) \exp\left( \beta_{jk,1} discontinue_i(t) + \sum_{m=1}^{M} \beta_{jk,m}^{T} Z_{ijm}(t) \right)$$

$$j,k = 1,2,..,e;\ j \neq k;\ i = 1,2,..,n$$

Where:

- $h_{ijk}(t)$ is the hazard of transitioning from state *j* to state *k* for individual *i* at time *t*
- $y_{ijk}(t)$ is the risk indicator for the $j \rightarrow k$ transition at time *t*
- $h_{0jk}(t)$ is the baseline hazard for transition $j \rightarrow k$
- $discontinue_i(t)$ is a time-varying indicator for treatment discontinuation
- $Z_{ijm}(t)$ is a vector of time-varying covariates measured at *m* different time points
- $\beta_{jk,m}$ are the corresponding regression coefficients for the $j \rightarrow k$ transition

The model is stratified by transition interval and accounts for within-subject correlation using robust (clustered) standard errors.

### 3.2.7 Sensitivity analysis

To assess the robustness of our findings, we conducted sensitivity analyses using alternative model specifications and sample sizes.

First, we evaluated the impact of different assumptions about the underlying time-to-event distribution. In addition to the Gompertz distribution used in the primary analysis, we fit Weibull distributions to the original Kaplan-Meier curves for stroke, bleeding, and death. Shape and scale parameters for the Weibull model were estimated using maximum likelihood techniques based on digitized trial data (Table 5). We then re-simulated survival times using these Weibull parameters and repeated the full simulation pipeline, including treatment assignment, dynamic hazard modeling, and covariate-based hazard adjustment. The results were compared with those from the Gompertz-based simulations to assess sensitivity to distributional assumptions.

Second, we tested the effect of varying sample sizes. This allowed us to evaluate the stability of event rates and treatment effect estimates under different levels of statistical power.

Table 5. Weibull parameters.

| Weibull Parameters for Simulated Time-to-Event Outcomes | | |
|---|---|---|
| Event | Shape | Scale |
| Ischemic stroke | 1.602 | 3.399 |
| Major bleeding and All-cause death | 1.675 | 2.944 |

## 3.3 Software

All statistical analyses were conducted using R (version 4.4.2). The survival, flexsurv, and cmprsk packages were used for time-to-event modeling. The mstate package supported multi-state modeling, while muhaz was employed for hazard function estimation. The rms package aided in regression modeling strategies, and powerSurvEpi was used for power calculations in survival studies. Model diagnostics and goodness-of-fit were assessed with goftest. Tables were formatted for presentation using the gt package. Kaplan-Meier curves from the clinical trials were digitized using WebPlotDigitizer (33) to extract survival data.

# 4. Results

## 4.1 Dataset simulations

We simulated a cohort of older adults (≥65 years) with subclinical atrial fibrillation using parameters derived from the NOAH-AFNET 6 and ARTESIA trials. To ensure an adequate number of stroke events, we estimated the required sample size to achieve 145 events with 80% power at a 5% significance level. Assuming a stroke rate of 0.008–0.012 per patient-year and an estimated HR of 1.395, the required sample size was 29,000 patients (14,500 per group). However, because patients could experience events in the first year before randomization, we initially simulated 60,000 individuals. Table 4 shows the distribution of baseline covariates. The mean age was 77.2 years (SD = 7.2), and 36.8% were female. Baseline characteristics were comparable to those reported in the original clinical trials. After excluding those who experienced an event during the first year, the final cohort included 30,238 individuals.

Table 6. Simulated vs trial baseline characteristics.

| Baseline Characteristics: Simulated vs. Trial Populations | | | |
|---|---|---|---|
| Variable | Simulated Data | ARTESIA Trial | NOAH-AFNET 6 Trial |
| Age (mean (SD)) | 77.2 (7.2) | 76.8 (7.6) | 77.5 (6.7) |
| Sex (%) | 36.8 | 36.1 | 37.4 |
| Hypertension (%) | 84.2 | 81.5 | 86.9 |
| Coronary artery disease (%) | 31.5 | 37.0 | 26.4 |
| Diabetes (%) | 28.2 | 29.1 | 26.9 |
| Heart failure (%) | 28.1 | 28.3 | 27.4 |
| History of stroke (%) | 9.7 | 9.0 | 10.0 |
| Antiplatelet therapy (%) | 55.7 | 57.4 | 53.9 |
| Creatinine (mean (SD)) | 68.7 (26.3) | 71.4 (28.7) | 66.0 (23.4) |

## 4.2 Multistate Transitions and Event Incidence

We modeled 9 transitions using a semi-Markov multistate framework. The five primary clinical states were: 1) on anticoagulation, 2) discontinuation, 3) stroke, 4) major bleeding, and 5) death. Transitions were allowed between baseline and each clinical event, as well as from discontinuation to each adverse event, and from intermediate events to death (Figure 3).

Figure 3. Transition Matrix Table.

| Transition Matrix of the Multistate Model | | | | | |
|---|---|---|---|---|---|
| From State | S1 | Discontinue | Stroke | Bleeding | Death |
| S1 | | 1 | 2 | 3 | 4 |
| Discontinue | | | 5 | 6 | 7 |
| Stroke | | | | | 8 |
| Bleeding | | | | | 9 |
| Death | | | | | |

Among the 30,238 simulated patients, 15,199 individuals (50.26%) discontinued anticoagulation during follow-up. From the baseline state (on treatment), 4,411 patients (14.58%) experienced an ischemic stroke, 5,142 (17.01%) experienced a major bleeding event, and 5,486 (18.14%) died without experiencing an intermediate event.

Among patients who discontinued anticoagulation: 6,435 (42.33%) had a stroke, 3,264 (21.47%) had a major bleeding event, and 5,500 (36.18%) died.

Of those who experienced an intermediate event (stroke or bleeding event), most eventually transitioned to death: 99.54% of patients who had a stroke subsequently died (10,797 out of 10,846) and 99.48% of those who experienced major bleeding also died (8,363 out of 8,406).

## 4.3 Effect of Deprescribing

### 4.3.1 Stroke Risk

Anticoagulation discontinuation was associated with a 36% reduction in stroke risk (HR 0.64, 95% CI 0.63–0.66, $p<0.001$). Age demonstrated a significant consistent per-year risk increase

of 8% for initial strokes (transition 2: HR 1.08, 95% CI 1.08–1.09) and 9% for post-discontinuation strokes (transition 5: HR 1.09, 95% CI 1.08–1.1). Females had significantly higher risk, with hazard ratios of 1.19 (95% CI 1.07–1.32, p=0.001) for transition 2 and 1.28 (95% CI 1.18–1.4, p<0.001) for transition 5. The strongest predictor was prior stroke history, nearly tripling risk for both initial strokes (transition 2: HR 2.99, 95% CI 2.45–3.63) and post-discontinuation events (transition 5: HR 3.63, 95% CI 3.07–4.29). Significant associations were observed for antiplatelet use: 31% increased risk for transition 2 (HR 1.31, 95% CI 1.22–1.41), 27% for transition 5 (HR 1.27, 95% CI 1.2–1.35). Modest and significant associations were also observed for hypertension and diabetes. Creatinine showed minimal but statistically significant effects (both transitions: HR 1.00, p≤0.02), while heart failure and CAD showed no significant associations. Table 7.

### 4.3.2 Major Bleeding

Anticoagulation discontinuation reduced major bleeding risk by 10% (HR 0.9, 95% CI 0.88–0.92, p<0.001). Age was again positively associated with bleeding risk, with hazard ratios of 1.07 (95% CI 1.06–1.08) for transition 3 and 1.05 (95% CI 1.04–1.06) for transition 6. Female sex was associated with a lower bleeding risk (HR 0.8, 95% CI 0.72–0.89 for transition 3; HR 0.75, 95% CI 0.66–0.85 for transition 6). Comorbidities such as hypertension, diabetes, and heart failure also showed positive associations in specific transitions. Antiplatelet use was a particularly strong risk factor for bleeding (HR 1.3, 95% CI 1.22–1.39 for transition 3; HR 1.31, 95% CI 1.21–1.43 for transition 6).Table 8.

### 4.3.3 Mortality

Discontinuation showed no mortality effect (HR 0.98, 95% CI 0.95–1.01, p=0.207). Age consistently showed 7–8% annual risk increase across all transitions (all p<0.001). Table 9 shows the effect of each covariate on mortality for each transition.

Table 7. Covariate effects on time-to-stroke outcome.

| Covariate effects on time-to-stroke outcome: HR 95%, confidence intervals (CI) and p-values for T2 and T5 | | | | |
|---|---|---|---|---|
| | T2 | | T5 | |
| Covariate | HR 95% (CI) | p-value | HR 95% (CI) | p-value |
| Discontinuation | - | - | 0.64 (0.63 – 0.66) | <0.001 |
| Age (years) | 1.08 (1.08 – 1.09) | <0.001 | 1.09 (1.08 – 1.1) | <0.001 |
| Sex (female) | 1.19 (1.07 – 1.32) | 0.001 | 1.28 (1.18 – 1.4) | <0.001 |
| CHADSVASC score | 1.02 (0.95 – 1.11) | 0.533 | 0.98 (0.91 – 1.05) | 0.560 |
| Hypertension | 1.13 (1.01 – 1.26) | 0.029 | 1.2 (1.09 – 1.32) | <0.001 |
| Coronary arterial disease | 1.09 (0.97 – 1.22) | 0.138 | 1.02 (0.93 – 1.12) | 0.663 |
| Diabetes | 1.16 (1.04 – 1.29) | 0.008 | 1.19 (1.09 – 1.31) | <0.001 |
| Heart failure | 0.96 (0.86 – 1.08) | 0.484 | 0.99 (0.9 – 1.1) | 0.903 |
| History of stroke | 2.99 (2.45 – 3.63) | <0.001 | 3.63 (3.07 – 4.29) | <0.001 |
| Antiplatelets treatment | 1.31 (1.22 – 1.41) | <0.001 | 1.27 (1.2 – 1.35) | <0.001 |
| Creatinine (mg/dl) | 1 (1 – 1) | <0.001 | 1 (1 – 1) | <0.001 |

Table 8. Covariate effects on time-to-bleeding outcome.

| Covariate effects on time-to-bleeding outcome: HR 95%, confidence intervals (CI) and p-values for T3 and T6 | | | | |
|---|---|---|---|---|
| | T3 | | T6 | |
| Covariate | HR 95% (CI) | p-value | HR 95% (CI) | p-value |
| Discontinuation | - | - | 0.9 (0.88 – 0.92) | <0.001 |
| Age (years) | 1.07 (1.06 – 1.08) | <0.001 | 1.05 (1.04 – 1.06) | <0.001 |
| Sex (female) | 0.8 (0.72 – 0.89) | <0.001 | 0.75 (0.66 – 0.85) | <0.001 |
| CHADSVASC score | 0.96 (0.89 – 1.04) | 0.337 | 1.09 (1 – 1.18) | 0.061 |
| Hypertension | 1.25 (1.12 – 1.38) | <0.001 | 1.06 (0.93 – 1.19) | 0.388 |
| Coronary arterial disease | 1.03 (0.93 – 1.15) | 0.557 | 0.96 (0.85 – 1.1) | 0.575 |
| Diabetes | 1.21 (1.09 – 1.35) | <0.001 | 1.09 (0.96 – 1.24) | 0.173 |
| Heart failure | 1.14 (1.02 – 1.27) | 0.018 | 1.02 (0.89 – 1.16) | 0.786 |
| History of stroke | 1.49 (1.2 – 1.86) | <0.001 | 1.03 (0.78 – 1.36) | 0.842 |
| Antiplatelets treatment | 1.3 (1.22 – 1.39) | <0.001 | 1.31 (1.21 – 1.43) | <0.001 |
| Creatinine (mg/dl) | 1.01 (1.01 – 1.01) | <0.001 | 1.01 (1.01 – 1.01) | <0.001 |

Table 9. Covariate effects on time-to-death outcome.

| Covariate effects on time-to-death outcome: HR 95%, confidence intervals (CI) and p-values for T4, T7, T8 and T9 | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | T4 | | T7 | | T8 | | T9 | |
| Covariate | HR 95% (CI) | p-value | HR 95% (CI) | p-value | HR 95% (CI) | p-value | HR 95% (CI) | p-value |
| Discontinuation | - | - | 0.98 (0.95 – 1.01) | 0.207 | - | - | - | - |
| Age (years) | 1.08 (1.08 – 1.09) | <0.001 | 1.08 (1.07 – 1.08) | <0.001 | 1.07 (1.07 – 1.08) | <0.001 | 1.07 (1.07 – 1.08) | <0.001 |
| Sex (female) | 1.29 (1.18 – 1.4) | <0.001 | 1.25 (1.15 – 1.35) | <0.001 | 1.01 (0.95 – 1.07) | 0.752 | 1 (0.93 – 1.07) | 0.917 |
| CHADSVASC score | 0.89 (0.84 – 0.94) | <0.001 | 0.87 (0.82 – 0.92) | <0.001 | 1.07 (1.02 – 1.12) | 0.003 | 1.1 (1.05 – 1.15) | <0.001 |
| Hypertension | 1.24 (1.14 – 1.36) | <0.001 | 1.35 (1.24 – 1.47) | <0.001 | 1.06 (0.99 – 1.13) | 0.089 | 1.04 (0.98 – 1.12) | 0.199 |
| Coronary arterial disease | 1.3 (1.18 – 1.42) | <0.001 | 1.36 (1.25 – 1.49) | <0.001 | 1.13 (1.05 – 1.21) | <0.001 | 1.1 (1.03 – 1.19) | 0.007 |
| Diabetes | 1.42 (1.29 – 1.55) | <0.001 | 1.32 (1.21 – 1.44) | <0.001 | 1.18 (1.1 – 1.26) | <0.001 | 1.18 (1.09 – 1.26) | <0.001 |
| Heart failure | 1.52 (1.39 – 1.67) | <0.001 | 1.58 (1.44 – 1.73) | <0.001 | 1.32 (1.23 – 1.41) | <0.001 | 1.25 (1.16 – 1.35) | <0.001 |
| History of stroke | 1.6 (1.32 – 1.95) | <0.001 | 1.67 (1.37 – 2.05) | <0.001 | 1.29 (1.15 – 1.46) | <0.001 | 1.28 (1.11 – 1.48) | <0.001 |
| Antiplatelets treatment | 1.09 (1.02 – 1.16) | 0.009 | 1.03 (0.97 – 1.09) | 0.366 | 1.1 (1.05 – 1.14) | <0.001 | 1.16 (1.11 – 1.21) | <0.001 |
| Creatinine (mg/dl) | 1.01 (1.01 – 1.01) | <0.001 | 1.01 (1 – 1.01) | <0.001 | 1.01 (1.01 – 1.01) | <0.001 | 1.01 (1.01 – 1.01) | <0.001 |

## 4.4 Bootstrap Performance Assessment

Using 1000 bootstrap replicates, we estimated the empirical bias, coverage probability, and Monte Carlo standard errors for the estimated hazard ratios of discontinue in the stroke, bleeding, and death models. Results are presented separately for Gompertz and Weibull baseline hazard assumptions and evaluated under three sample size scenarios (small: $n$ = 10,000; baseline: $n$ = 30,238; large: $n$ = 100,000).

### 4.4.1 Hazard Ratio Estimates Across Parametric Specifications

Table 10 summarizes the estimated hazard ratios (HRs) and their respective 95% confidence intervals (CIs) for the effect of treatment discontinuation under Gompertz and Weibull baseline hazard models. Estimates are stratified by outcome (stroke, bleeding, death) and sample size (small: $n$ = 10,000; baseline: $n$ = 30,238; large: $n$ = 100,000).

For stroke, the Gompertz model produced HRs ranging from 0.64 (95% CI: 0.63–0.66) to 0.66 (95% CI: 0.62–0.70), indicating a moderate protective effect of discontinuation. The Weibull model produced lower HRs, between 0.47 (95% CI: 0.46–0.48) and 0.48 (95% CI: 0.47–0.50), suggesting a stronger protective effect. Within-model estimates remained consistent across sample sizes, whereas between-model discrepancies were substantial.

For bleeding, the Gompertz-based estimates were closer to the null, with HRs of 0.90 (95% CI: 0.88–0.92) to 0.92 (95% CI: 0.90–0.93). Weibull-based estimates were systematically lower, ranging from 0.80 (95% CI: 0.74–0.85) to 0.82 (95% CI: 0.80–0.84). As with stroke, sample size had little impact on the magnitude of the estimates, though the choice of parametric distribution consistently affected the results.

For death, Gompertz estimates centered around the null, with HRs ranging from 0.97 (95% CI: 0.91–1.05) to 1.00 (95% CI: 0.98–1.03). By contrast, Weibull estimates indicated a slight increase in risk, with HRs ranging from 1.01 (95% CI: 0.93–1.09) to 1.06 (95% CI: 1.03–1.10).

Table 10. Comparison between Gompertz and Weibull distributions on deprescribing effect

**Hazard ratio 95% estimates for the effect of discontinuation**

Comparison under Gompertz and Weibull distributional assumptions across sample sizes

| | Gompertz Model | Weibull Model |
|---|---|---|
| **Stroke** | | |
| Small | 0.66 (0.62–0.70) | 0.47 (0.44–0.51) |
| Baseline | 0.64 (0.63–0.66) | 0.48 (0.47–0.50) |
| Large | 0.66 (0.65–0.67) | 0.47 (0.46–0.48) |
| **Bleeding** | | |
| Small | 0.91 (0.86–0.97) | 0.80 (0.74–0.85) |
| Baseline | 0.90 (0.88–0.92) | 0.82 (0.80–0.84) |
| Large | 0.92 (0.90–0.93) | 0.81 (0.79–0.83) |
| **Death** | | |
| Small | 0.97 (0.91–1.05) | 1.01 (0.93–1.09) |
| Baseline | 0.98 (0.95–1.01) | 1.06 (1.03–1.10) |
| Large | 1.00 (0.98–1.03) | 1.05 (1.03–1.08) |

Baseline size corresponds to the main simulation sample (n = 30,238). Small and large represent sensitivity analyses with n = 10,000 and n = 100,000, respectively.

These sensitivity analyses confirm that, although the choice of baseline hazard distribution has a measurable impact on point estimates, particularly for stroke, overall inference regarding the direction and strength of the effects of stopping treatment remains consistent across models.

### 4.4.2 Bias and Coverage Properties of Hazard Ratio Estimates

Table 11 shows the estimated bias, coverage and Monte Carlo standard errors (SE Bias) of the hazard ratio estimates for treatment discontinuation under the two parametric assumptions and for the three different sample sizes and three clinical outcomes.

For stroke, the Gompertz model showed a moderate negative bias (−0.41 to −0.72 on the log-HR scale) and high coverage in the small and baseline scenarios (0.97–0.99). However, coverage dropped markedly in the large-sample scenario (0.46), which is consistent with the accumulation of small model misspecification errors in the asymptotic regime. The Weibull model showed severe positive bias (8.26–13.51) and almost no coverage in the small and large sample scenarios.

For bleeding, Gompertz estimates showed a smaller bias (−0.06 to −0.32) with nominal coverage in small and baseline samples (0.99 - 1.00). However, there was reduced coverage in the large sample (0.64). By contrast, Weibull estimates showed a larger positive bias (1.61–3.43) and reduced coverage, particularly in the large sample (0.34).

For mortality, Gompertz-based estimates showed substantial positive bias (1.62 - 2.78) and extremely poor coverage in the baseline and large samples (<0.01). In contrast, Weibull-based estimates demonstrated smaller bias (0.56- 2.03) and adequate coverage in the small and baseline samples (0.96 - 0.98). However, coverage also collapsed in the large sample (0.05).

These results emphasise the importance of selecting the right model for parametric survival analysis. For stroke and bleeding, the Gompertz model outperformed the Weibull model, producing near-unbiased estimates with high coverage. However, the Weibull model was more appropriate for mortality, as the Gompertz assumption resulted in substantial bias and almost no coverage.

### 4.4.3 Comparative Performance of Parametric Specifications

From a methodological standpoint, the performance of the Gompertz and Weibull models differed systematically depending on the outcome. For non-fatal outcomes (stroke and bleeding), the Gompertz model consistently produced lower bias and more accurate coverage, suggesting that its monotonic hazard function was better aligned with the underlying risk dynamics. In contrast, the Weibull model consistently overstated the protective effect of discontinuation for stroke and underestimated the hazard for bleeding. This led to inflated hazard ratio reductions and attenuated coverage, particularly in larger samples.

The reverse pattern was observed for mortality. Here, the Weibull distribution provided better coverage and was less biased in finite samples. In contrast, Gompertz assumptions imposed structural constraints that distorted estimates and resulted in undercoverage, even in moderate-sized cohorts. This suggests that the Weibull model more accurately captured the risk trajectory of death, whereas the Gompertz form was too restrictive in this context.

The role of sample size was consistent with asymptotic expectations: small and baseline cohorts produced stable estimates with nominal coverage when the models were reasonably well-specified. However, even minor misspecifications were amplified in the large-sample scenario ($n =$ 100,000), resulting in significant loss of coverage. This pattern shows that the increased precision gained by increasing the sample size cannot compensate for structural model misspecification.

Taken together, these findings suggest that there is no single parametric distribution that is universally appropriate for all outcomes. The Gompertz distribution is preferable when hazards are expected to evolve dynamically over time (e.g. stroke or bleeding), whereas the Weibull distribution is better suited to mortality. For simulation studies of deperscribing, outcome-specific modelling should therefore be standard practice, alongside sensitivity analyses under alternative hazard specifications.

Table 11. Comparison between sample sizes and parametric distributions on deprescribing effect.

**Sensitivity Analysis**

Comparison between sample sizes on deprescribing effect

| | Gompertz Model | | | Weibull Model | | |
|---|---|---|---|---|---|---|
| | Small | Baseline[1] | Large | Small | Baseline[1] | Large |
| **Stroke** | | | | | | |
| Bias | −0.464 | −0.406 | −0.717 | 12.105 | 8.263 | 13.512 |
| Coverage | 0.990 | 0.973 | 0.460 | 0.899 | 0.060 | 0.000 |
| SE Bias | 0.024 | 0.009 | 0.005 | 0.504 | 0.124 | 0.145 |
| **Bleeding** | | | | | | |
| Bias | −0.197 | −0.063 | −0.322 | 3.433 | 1.606 | 2.124 |
| Coverage | 0.985 | 1.000 | 0.636 | 0.989 | 0.814 | 0.335 |
| SE Bias | 0.017 | 0.006 | 0.003 | 0.269 | 0.038 | 0.036 |
| **Death** | | | | | | |
| Bias | 1.615 | 2.783 | 2.558 | 1.880 | 0.563 | 2.032 |
| Coverage | 0.961 | 0.004 | 0.000 | 0.984 | 0.957 | 0.047 |
| SE Bias | 0.048 | 0.025 | 0.019 | 0.078 | 0.015 | 0.022 |

[1] Reference size corresponds to the main simulation sample (n = 30,238). Small and large represent sensitivity analyses with n = 10,000 and n = 100,000, respectively.

# 5. Discussion

This study presents a comprehensive methodology for simulating target trials in order to evaluate the impact of deprescribing, as illustrated by the discontinuation of anticoagulation therapy in older adults with subclinical atrial fibrillation. Simulation enables us to address key limitations of traditional studies on this topic, such as heterogeneity in interventions, limited generalisability and inconsistent outcome definitions.

A key strength of the approach is its foundation in the target trial framework, which provides a clear, structured plan for replicating causal inference. By specifying the eligibility criteria, treatment strategies, outcomes and follow-up procedures precisely, we can ensure that the simulated study closely mimics an ideal randomised controlled trial. This structured approach mitigates confounding and selection bias, which is particularly important in studies of drug withdrawal, where reverse causation and unmeasured clinical judgement can distort associations.

Realism was further enhanced by generating a synthetic population based on empirical distributions and correlation structures derived from the ARTESIA and NOAH-AFNET 6 trials. Time-to-event outcomes were modelled using covariate-informed hazard adjustments, enabling event risks to adapt dynamically to individual characteristics instead of being assigned as fixed aggregate rates. Treatment strategies were simulated with a one-year eligibility window to reflect clinical decision-making processes and prevent early events from contaminating the analysis.

A notable innovation of this framework is its use of a semi-Markov multistate model. This model enables us to capture transitions between discontinuation, adverse events and death, and allows prior events to influence subsequent risks. This design is particularly relevant in contexts involving the withdrawal of medication, where treatment cessation, complications, and mortality interact in clinically meaningful ways.

Although decision-analytic models have been used to evaluate the outcomes of deprescribing in certain settings, such as proton pump inhibitor use or antihypertensive therapy in older adults (5,20), these studies often rely on simplified decision trees or Markov models with static assumptions. To our knowledge, this is one of the first studies to use a fully specified target trial emulation framework to simulate the clinical impact of deprescribing. Unlike

previous models, our approach incorporates dynamic treatment decisions and time-varying risks.

Bootstrap validation provided strong evidence that the framework is robust, revealing stable hazard ratio estimates and low bias under correctly specified models. Sensitivity analyses confirmed that parametric assumptions can significantly impact inference, but also demonstrated that outcome-specific choices can mitigate this issue.

This study also shows that parametric survival models are very sensitive to how the baseline hazard is specified. This has direct implications for how accurate the estimates are, how biased they are, and how valid the inferences are. Two key methodological lessons emerge. Firstly, the choice of model should be outcome-specific, as no single distribution can adequately represent the hazard functions across all endpoints. For stroke and bleeding, the Gompertz distribution provided closer alignment with the data-generating process, resulting in lower bias and more reliable coverage. For mortality, however, the Weibull specification was more appropriate, as the Gompertz form imposed structural constraints that resulted in undercoverage. Secondly, sample size moderates inferential reliability. While small and moderate cohorts produced stable estimates when models were correctly specified, even subtle misspecifications resulted in severe coverage loss in large samples, as predicted by asymptotic theory. These findings emphasise the importance of carrying out thorough checks to ensure that models are adequate before scaling simulations, as larger samples amplify the consequences of misspecification.

From a broader methodological perspective, our results suggest the need for greater flexibility in hazard specification within simulation frameworks for causal inference. Semi-parametric methods (e.g., Cox proportional hazards) and flexible parametric approaches (e.g., splines, piecewise exponential models) can reduce the impact of rigid parametric assumptions while maintaining tractability. When using fully parametric models, sensitivity analyses across multiple hazard forms should be standard practice.

Although this methodology has significant advantages, it does have certain limitations. Firstly, simulated trials cannot fully capture real-world complexity, such as adherence patterns, unmeasured confounders, or rare outcomes. Secondly, the validity of the results depends on the accuracy of the input parameters, and biases in the underlying trial data may propagate through the simulation. Nevertheless, the framework is both flexible and

extensible, which makes it a versatile solution for a variety of needs. While it has been applied here to anticoagulation, the framework can be generalised to other therapeutic areas and adapted to populations with multimorbidity or frailty. It can also be extended to prospective trial planning, for example for sample size calculations, subgroup targeting and adaptive designs. Unlike conventional decision-analytic models with static assumptions, this approach allows for dynamic treatment decisions and time-varying risks, providing a more accurate representation of real clinical pathways.

In summary, these findings emphasise the importance of outcome-specific hazard modelling and systematic robustness checks when designing and interpreting simulation-based evaluations of deprescribing interventions. Rather than defaulting to a single distributional form, researchers should tailor hazard specifications to the outcome under study.

# 6. Conclusions

This work introduces a rigorous, simulation-based framework for emulating target trials and evaluating the effectiveness of interventions. The results highlight that hazard specification depends on the outcome. More broadly, robust inference requires outcome-specific modelling, systematic sensitivity analyses and, where feasible, flexible semi-parametric approaches.

By incorporating causal inference principles into a realistic, multistate modelling framework, this methodology offers a transparent and adaptable means of generating evidence in situations where randomised trials are impractical. This methodology advances the methodological foundation for deprescribing research, supports more informed clinical decision-making and paves the way for future applications in diverse therapeutic contexts.

# 7. References

1. Veronese N, Gallo U, Boccardi V, Demurtas J, Michielon A, Taci X, et al. Efficacy of deprescribing on health outcomes: An umbrella review of systematic reviews with meta-analysis of randomized controlled trials. Ageing Res Rev. 2024 Mar;95:102237.

2. Hilmer SN, Johnell K, Mach J. Pre-clinical Models for Geriatric Pharmacotherapy. Drugs Aging. 2024 Aug;41(8):633–40.

3. Montero C. Deprescripción. Más allá del uso racional del medicamento. Med fam Andal. 2020 Apr;21(1):47–66.

4. Mercer O, Threapleton C, Bennett F, Baker E. Identification of a complex of symptoms frequently caused by commonly prescribed drugs as a potential tool to quantify the benefits of deprescribing. Br J Clin Pharmacol. 2020 June;86(6):1194–5.

5. Xie M, You JHS. Deprescribing of proton pump inhibitors in older patients: A cost-effectiveness analysis. Arab-Zozani M, editor. PLOS ONE. 2024 Oct 7;19(10):e0311658.

6. Thompson W, Pizzola L, Hogel M, Black C, Farrell B. Developing an Evidence-based Deprescribing Guideline: INSTRUCTION MANUAL For Guideline Coordinators. 2018. 76 p.

7. Cole JA, Gonçalves-Bradley DC, Alqahtani M, Barry HE, Cadogan C, Rankin A, et al. Interventions to improve the appropriate use of polypharmacy for older people. Cochrane Effective Practice and Organisation of Care Group, editor. Cochrane Database Syst Rev [Internet]. 2023 Oct 11 [cited 2025 Mar 1];2023(10). Available from: http://doi.wiley.com/10.1002/14651858.CD008165.pub5

8. Martin P, Tamblyn R, Benedetti A, Ahmed S, Tannenbaum C. Effect of a Pharmacist-Led Educational Intervention on Inappropriate Medication Prescriptions in Older Adults: The D-PRESCRIBE Randomized Clinical Trial. JAMA. 2018 Nov 13;320(18):1889–98.

9. Kouladjian O'Donnell L, Gnjidic D, Sawan M, Reeve E, Kelly PJ, Chen TF, et al. Impact of the Goal-directed Medication Review Electronic Decision Support System on Drug Burden Index: A cluster-randomised clinical trial in primary care. Br J Clin Pharmacol. 2021 Mar;87(3):1499–511.

10. Bloomfield HE, Greer N, Linsky AM, Bolduc J, Naidl T, Vardeny O, et al. Deprescribing for Community-Dwelling Older Adults: a Systematic Review and Meta-analysis. J Gen Intern Med. 2020 Nov;35(11):3323–32.

11. Herrinton LJ, Lo K, Alavi M, Alexeeff SE, Butler KM, Chang C, et al. Effectiveness of Bundled Hyperpolypharmacy Deprescribing Compared With Usual Care Among Older Adults: A Randomized Clinical Trial. JAMA Netw Open. 2023 July 3;6(7):e2322505.

12. Kornholt J, Feizi ST, Hansen AS, Laursen JT, Reuther LØ, Petersen TS, et al. Effects of a comprehensive medication review intervention on health-related quality of life and other clinical outcomes in geriatric outpatients with polypharmacy: A pragmatic randomized clinical trial. Br J Clin Pharmacol. 2022 July;88(7):3360–9.

13. Moreira PM, Aguiar EC, Castro PR, Almeida KC, Dourado JA, Paula SM, et al. Optimizing Hypertension Treatment in Older Patients Through Home Blood Pressure Monitoring by Pharmacists in Primary Care: The MINOR Clinical Trial. Clin Ther. 2023

Oct;45(10):941–6.

14. Phelan EA, Williamson BD, Balderson BH, Cook AJ, Piccorelli AV, Fujii MM, et al. Reducing Central Nervous System-Active Medications to Prevent Falls and Injuries Among Older Adults: A Cluster Randomized Clinical Trial. JAMA Netw Open. 2024 July 1;7(7):e2424234.

15. Vasilevskis EE, Shah AS, Hollingsworth EK, Shotwell MS, Kripalani S, Mixon AS, et al. Deprescribing Medications Among Older Adults From End of Hospitalization Through Postacute Care: A Shed-MEDS Randomized Clinical Trial. JAMA Intern Med. 2023 Mar 1;183(3):223–31.

16. McDonald EG, Wu PE, Rashidi B, Wilson MG, Bortolussi-Courval É, Atique A, et al. The MedSafer Study-Electronic Decision Support for Deprescribing in Hospitalized Older Adults: A Cluster Randomized Clinical Trial. JAMA Intern Med. 2022 Mar 1;182(3):265–73.

17. Ie K, Hirose M, Sakai T, Motohashi I, Aihara M, Otsuki T, et al. Medication Optimization Protocol Efficacy for Geriatric Inpatients: A Randomized Clinical Trial. JAMA Netw Open. 2024 July 1;7(7):e2423544.

18. Johansson T, Abuzahra ME, Keller S, Mann E, Faller B, Sommerauer C, et al. Impact of strategies to reduce polypharmacy on clinically relevant endpoints: a systematic review and meta-analysis. Br J Clin Pharmacol. 2016 Aug;82(2):532–48.

19. Huiskes VJB, Burger DM, Van Den Ende CHM, Van Den Bemt BJF. Effectiveness of medication review: a systematic review and meta-analysis of randomized controlled trials. BMC Fam Pract. 2017 Dec;18(1):5.

20. Jowett S, Kodabuckus S, Ford GA, Hobbs FDR, Lown M, Mant J, et al. Cost-Effectiveness of Antihypertensive Deprescribing in Primary Care: a Markov Modelling Study Using Data From the OPTiMISE Trial. Hypertension. 2022 May;79(5):1122–31.

21. Muddiman R, Aiello Battan FI, Tazare J, Schultze A, Boland F, Perez T, et al. Pharmacoepidemiology simulation study practices: A methodological review [Internet]. Epidemiology; 2025 [cited 2025 Sept 7]. Available from: http://medrxiv.org/lookup/doi/10.1101/2025.04.30.25326427

22. Hernán MA, Robins JM. Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available: Table 1. Am J Epidemiol. 2016 Apr 15;183(8):758–64.

23. Morris TP, White IR, Crowther MJ. Using simulation studies to evaluate statistical methods. Stat Med. 2019 May 20;38(11):2074–102.

24. Schulman S, Kearon C. Definition of major bleeding in clinical investigations of antihemostatic medicinal products in non-surgical patients. J Thromb Haemost. 2005 Apr;3(4):692–4.

25. Kirchhof P, Toennis T, Goette A, Camm AJ, Diener HC, Becher N, et al. Anticoagulation with Edoxaban in Patients with Atrial High-Rate Episodes. N Engl J Med. 2023 Sept 28;389(13):1167–79.

26. Healey JS, Lopes RD, Granger CB, Alings M, Rivard L, McIntyre WF, et al. Apixaban for Stroke Prevention in Subclinical Atrial Fibrillation. N Engl J Med. 2024 Jan

11;390(2):107–17.

27. Collett D, Kimber A. Modelling Survival Data in Medical Research, Third Edition. 3rd ed. Boca Raton: Chapman and Hall/CRC; 2015. 1 p. (Chapman and Hall/CRC Texts in Statistical Science Ser).

28. Friberg L, Rosenqvist M, Lip GYH. Evaluation of risk stratification schemes for ischaemic stroke and bleeding in 182 678 patients with atrial fibrillation: the Swedish Atrial Fibrillation cohort study. Eur Heart J. 2012 June;33(12):1500–10.

29. Le-Rademacher JG, Therneau TM, Ou FS. The Utility of Multistate Models: A Flexible Framework for Time-to-Event Data. Curr Epidemiol Rep. 2022 June 29;9(3):183–9.

30. Andersen PK, Keiding N. Multi-state models for event history analysis. Stat Methods Med Res. 2002 Apr;11(2):91–115.

31. Meira-Machado L, De Uña-Álvarez J, Cadarso-Suárez C, Andersen PK. Multi-state models for the analysis of time-to-event data. Stat Methods Med Res. 2009 Apr;18(2):195–222.

32. Nicholls G. A Practioners' User Guide for 'Multi-State Models for Event History Analysis' with Applications in R [Internet]. [cited 2025 Sept 7]. Available from: https://gnicholl.github.io/post/multistatemodelsRtutorial/

33. Rohatgi A. WebPlotDigitizer [Internet]. [cited 2025 Feb 28]. Available from: https://automeris.io/WebPlotDigitizer