

Supplementary information

The long return on discovery science in cancer therapeutics

In the format provided by
the authors and unedited

Supplementary Methods

Generation of Figure 1

Figure 1 compares progression-free survival (PFS) across successive randomised trials of HER2-directed therapy in HER2-positive metastatic breast cancer. The figure reports no new data.

Published Kaplan–Meier curves were digitised directly from the source figures of the referenced publications using a script-based image-digitisation workflow: each source figure’s axes were calibrated from its tick marks, curves were traced by colour or line style with manually verified anchor points, and the monotone step structure of the Kaplan–Meier estimator was enforced. Extracted curves were cross-checked against the published median PFS values. The curves were then re-plotted on a common scale (0–55 months; 0–100% PFS) in R (version 4.3.3) using ggplot2, with survival represented as step functions. Censoring marks are not reproduced.

Curves represent approximate reconstructions of the published data, intended for qualitative comparison rather than re-analysis; at-risk tables were not reconstructed. The extracted step coordinates and the complete digitisation and plotting code are provided as Source Data.

Data sources

1. Paclitaxel: Slamon DJ, et al. Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. *N Engl J Med* 2001;344:783–792 (paclitaxel-alone arm, paclitaxel subgroup).
2. Trastuzumab + taxane (TH) and taxane + carboplatin + trastuzumab (TCH): Valero V, et al. Multicenter phase III randomized trial comparing docetaxel and trastuzumab with docetaxel, carboplatin, and trastuzumab as first-line chemotherapy for patients with HER2-gene-amplified metastatic breast cancer (BCIRG 007 study). *J Clin Oncol* 2011;29:149–156.
3. Trastuzumab emtansine (T-DM1) and T-DM1 + pertuzumab: Perez EA, et al. Trastuzumab emtansine with or without pertuzumab versus trastuzumab plus taxane for human epidermal growth factor receptor 2-positive, advanced breast cancer: primary results from the phase III MARIANNE study. *J Clin Oncol* 2017;35:141–148.
4. Taxane + trastuzumab + pertuzumab (THP) and trastuzumab deruxtecan + pertuzumab: Tolaney SM, et al. Trastuzumab deruxtecan plus pertuzumab for HER2-positive metastatic breast cancer (DESTINY-Breast09). *N Engl J Med* 2026;394:551–562.

```

#
=====

# Figure 1 — Progression-free survival across HER2-directed therapies (2001-2025)

# Kaplan-Meier curves re-drawn with tidyverse / ggplot2 (geom_step).

#

# INPUT : km_pfs_data.csv (regimen, year_label, month, pfs) — step corners
# OUTPUT: figure1_km.png / figure1_km.pdf

suppressPackageStartupMessages({
  if (requireNamespace("tidyverse", quietly = TRUE)) {
    library(tidyverse)
  } else {
    library(ggplot2); library(dplyr); library(readr)
  }
})

km <- readr::read_csv("km_pfs_data.csv", show_col_types = FALSE)

# ---- legend order (top to bottom, as in the original) -----
regimen_levels <- c(
  "Trastuzumab deruxtecan + pertuzumab",
  "Taxane + trastuzumab + pertuzumab",
  "Trastuzumab emtansine + pertuzumab",
  "Trastuzumab emtansine",
  "Trastuzumab + taxane",
  "Taxane + carboplatin + trastuzumab",
  "Paclitaxel"
)

# ---- draw order (bottom layer first): curves hidden under others in the
# original are drawn first so the layering reproduces the original ----
draw_order <- c(

```

```

"Trastuzumab emtansine + pertuzumab", # grey (under blue 5.5-8.3 mo)
"Trastuzumab emtansine",           # gold (under blue 3-8 mo)
"Taxane + carboplatin + trastuzumab", # blue (under red < 2 mo)
"Trastuzumab + taxane",           # red
"Taxane + trastuzumab + pertuzumab", # green
"Trastuzumab deruxtecan + pertuzumab", # purple
"Paclitaxel"                     # black on top
)

```

```

km <- km |> mutate(regimen = factor(regimen, levels = regimen_levels))

```

```

# ---- colours (sampled from the original curve strokes) -----

```

```

pal <- c(
  "Trastuzumab deruxtecan + pertuzumab" = "#5A01E9",
  "Taxane + trastuzumab + pertuzumab" = "#22A044",
  "Trastuzumab emtansine + pertuzumab" = "#8C8C8C",
  "Trastuzumab emtansine" = "#EFB63E",
  "Trastuzumab + taxane" = "#E02B32",
  "Taxane + carboplatin + trastuzumab" = "#2D9FE0",
  "Paclitaxel" = "#000000"
)

```

```

# year-annotation colours (as printed in the original)

```

```

ann_pal <- c(
  "Trastuzumab deruxtecan + pertuzumab" = "#5A01E9",
  "Taxane + trastuzumab + pertuzumab" = "#7FD75D",
  "Trastuzumab emtansine + pertuzumab" = "#9FA4AA",
  "Trastuzumab emtansine" = "#F7C948",
  "Trastuzumab + taxane" = "#E8232B",
  "Taxane + carboplatin + trastuzumab" = "#23B0F2",
  "Paclitaxel" = "#000000"
)

```

)

```
# ---- trial-year annotations (positions as in the original) -----
```

```
label_size <- 8.5
```

```
year_ann <- tibble::tribble(
```

```
  ~regimen,          ~month, ~pfs, ~label, ~lab_size,
```

```
  "Trastuzumab deruxtecan + pertuzumab", 27.8, 73.2, "2025", label_size,
```

```
  "Taxane + trastuzumab + pertuzumab", 33.5, 49.5, "2025", label_size,
```

```
  "Trastuzumab emtansine + pertuzumab", 46.8, 31.0, "2016", label_size,
```

```
  "Trastuzumab emtansine", 48.1, 21.0, "2016", label_size,
```

```
  "Trastuzumab + taxane", 44.0, 14.5, "2011", label_size,
```

```
  "Taxane + carboplatin + trastuzumab", 45.0, 5.9, "2010", label_size,
```

```
  "Paclitaxel", 27.5, 3.2, "2001", label_size
```

```
) |> mutate(regimen = factor(regimen, levels = regimen_levels))
```

```
# ---- build plot -----
```

```
p <- ggplot(km, aes(month, pfs, colour = regimen))
```

```
for (rg in draw_order) {
```

```
  p <- p + geom_step(data = km |> filter(regimen == rg),
```

```
    direction = "hv", linewidth = 1.0)
```

```
}
```

```
p <- p +
```

```
  geom_text(data = year_ann,
```

```
    aes(label = label, colour = regimen, size = lab_size),
```

```
    fontface = "bold", show.legend = FALSE) +
```

```
  scale_size_identity() +
```

```
  scale_colour_manual(values = pal, name = NULL, breaks = regimen_levels) +
```

```
  scale_x_continuous("Months", limits = c(0, 55.3),
```

```
    breaks = seq(0, 55, 5),
```

```

    expand = expansion(mult = c(0, 0.004))) +
scale_y_continuous("Progression-free survival (%)", limits = c(0, 101),
    breaks = seq(0, 100, 10),
    expand = expansion(mult = c(0, 0.005))) +
guides(colour = guide_legend(override.aes = list(linewidth = 2.2),
    label.position = "left")) +
theme_classic(base_size = 16) +
theme(
  axis.line      = element_line(linewidth = 0.7, colour = "black"),
  axis.ticks     = element_line(colour = "black", linewidth = 0.6),
  axis.ticks.length = unit(2.2, "mm"),
  axis.title     = element_text(face = "bold", size = 19),
  axis.text      = element_text(colour = "black", size = 14),
  legend.position = c(0.995, 0.995),
  legend.justification = c(1, 1),
  legend.text     = element_text(size = 12.5, hjust = 1),
  legend.key.width  = unit(1.05, "cm"),
  legend.key.height = unit(0.60, "cm"),
  legend.spacing.y  = unit(0, "cm"),
  legend.background = element_rect(fill = NA, colour = NA),
  plot.margin      = margin(10, 14, 8, 8)
)

ggsave("figure1_km.png", p, width = 11, height = 6.5, dpi = 300, bg = "white")
ggsave("figure1_km.pdf", p, width = 11, height = 6.5, bg = "white")

```