Supplementary material

Electronic supplementary material to: Counting rules and out-of-pocket payment under Japan’s 2026 high-cost medical expense reform: a simulation of breast cancer treatment courses

Journal: Breast Cancer

Authors: Akemi Hara, Akihiko Ozaki, Masahiro Wada, Ayu Ajitomi, Mika Nashimoto, Masanori Oshi, Yudai Kaneda, Toyoaki Sawano, Tomohiro Kurokawa, Kenji Gonda, Kazunoshin Tachibana, Masaru Noda, Tohru Ohtake, Yoshinori Nakata, Hiroaki Shimmura

Corresponding author: Akihiko Ozaki, Breast and Thyroid Center, Jyoban Hospital of Tokiwa Foundation, 57 Kaminodai, Jobankamiyunagaya-machi, Iwaki, Fukushima 972-8322, Japan. E-mail: ozakiakihiko@gmail.com

Contents

- Glossary
- A.1 Policy setting
  - A.1.1 Population and legal basis
  - A.1.2 Monthly out-of-pocket cap
  - A.1.3 Repeat-month cap
  - A.1.4 Annual cap
  - A.1.5 Timing of the income category determination
  - A.1.6 The second stage, effective 1 August 2027
- A.2 Treatment profiles
  - A.2.1 Derivation
  - A.2.2 Assumptions common to all profiles
  - A.2.3 Differences between the reviewed specification and the costed content
- A.3 Cost estimation
  - A.3.1 The selection rule
  - A.3.2 Drug costs
  - A.3.3 Fee-schedule items

    - A.3.4 Inpatient costs
    - A.3.5 Attribution of cost within the month
    - A.3.6 Implementation notes
    - A.3.7 Assumptions and boundary
- A.4 Construction of the monthly cost series
    - A.4.1 Unit of construction
    - A.4.2 The start day within the month
    - A.4.3 The start month as a design dimension
    - A.4.4 Observation period and computational horizon
    - A.4.5 Composition of a series
- A.5 Calculation of out-of-pocket payment
    - A.5.1 The order within a month
    - A.5.2 Treatment of the equality case
    - A.5.3 Annual layer
    - A.5.4 The two reform states
- A.6 Outcome definitions
    - A.6.1 The between-caps range
    - A.6.2 Margin
    - A.6.3 Primary outcome
    - A.6.4 Secondary outcome
    - A.6.5 Difference in annual out-of-pocket payment
- A.7 Sensitivity analysis and strata
    - A.7.1 The strata and their levels
    - A.7.2 Retired axis: granulocyte colony-stimulating factor
    - A.7.3 The inpatient-stay mapping
    - A.7.4 The P8 neoadjuvant arm
    - A.7.5 The CDK4/6 dispensing schedule and its switch point
    - A.7.6 The start day
- A.8 Analysis plan and reproducibility

2

- A.9 Undetermined items
- A.10 Clinical review record and bundled-payment equivalence threshold
- Supplementary tables
  - Table A1. Full specification of the 15 treatment profiles
  - Table A2. Fee-schedule items: adopted, rejected, and carried as unjustified
  - Table A3. Series counts by treatment profile
  - Table A4. The aggregation unit of the 21,000-yen floor
  - Table A5. Record of the superseded P8 implementation
  - Table A6. The 15 treatment profiles
  - Table A7. Monthly cap, repeat-month cap, and annual cap for enrollees under 70, before and after the reform effective 1 August 2026
  - Table A8. Monthly cap, repeat-month cap, and annual cap for the thirteen income categories effective 1 August 2027
  - Table A9. Units and denominators of the reported quantities
  - Table A10. The designated combination of the primary outcome, 2026 Category C, base stratum
  - Table A11. Verification of the encoding of the law
  - Table A12. Worked example: one course of the metastatic antibody-drug conjugate profile in 2027 Category 1, started in August and in March
  - Table A13. Inpatient-stay mapping: full sweep
  - Table A14. The body-size band
  - Table A15. Retirement of the G-CSF axis
  - Table A16. CDK4/6 switch point: full comparison and the eight sign flips
  - Table A17. (A0), (A), and (B): what each change moved, and the two corrections
  - Table A18. The primary outcome in all 18 income categories
  - Table A19. Differences between the reviewed clinical specification and the costed content
  - Table A20. The three CDK4/6 profiles over the 24-month course: annual-cap attainment and the difference, post-reform minus pre-reform
  - Table A21. Annual-cap attainment and start-month dependence under the fixed year and a rolling window
- Supplementary figures
  - Fig. A1. The monthly cap, the repeat-month cap, and the annual cap over one course
  - Fig. A2. Two regimes of annual-cap engagement

3

- Fig. A3. The dispensing interval reverses the direction of the reform

Glossary

Canonical term	First-use definition	Japanese source term
total medical expenditure	the full tariff-priced cost of care in a month, before any cost sharing	総医療費（10 割）
coinsurance amount	30% of total medical expenditure, the patient’s liability before any cap	3 割負担額
monthly cap	the statutory ceiling on a patient’s monthly out-of-pocket payment	月額自己負担上限
fixed component	the constant first term of the monthly cap formula	定額
reference amount	the total medical expenditure at which the coinsurance amount equals the fixed component; also the point at which the monthly cap begins to bind	基準額
repeat-month cap	the reduced monthly cap applied from the fourth qualifying month within a rolling 12-month window	多数回該当
annual cap	the ceiling on household out-of-pocket payment over a fixed 1 August to 31 July year	年間上限
annual-cap band	the set of monthly income categories that share one annual cap amount; its boundaries do not coincide with the monthly category boundaries	年間上限の帯
August–July year	one fixed 1 August to 31 July calculation period of the annual cap; a course of treatment spans one to three years depending on its calendar start month and its horizon	年間上限の計算期間（暦窓）
main path	the twelve treatment profiles P1 to P12, whose series are built by the principal generator on a common 12-month horizon. P13 to P15 are built by separate routines on their own horizons and are referred to by name, never as the main path. The rebuilt P8 is generated by its own routine but is evaluated on the same 12-month horizon and is counted in the main path	主経路（P1～P12）
calendar start month	the calendar month in which treatment begins, enumerated over all twelve values as a design dimension	開始暦月
course start	one series begun on one of 24 start dates, the 1st or the 15th of each calendar month; the unit over which Table 1 summarises the change in out-of-pocket payment	セル（1 系列×1 開始条件）
qualifying month	a month in which high-cost medical benefit is paid, i.e. the coinsurance amount exceeds the applicable monthly cap	発動月
benchmark increase	three times the rise in the monthly cap for an income category. It is the additional payment of a patient who pays the monthly cap in the first three qualifying months, and it is not published	目安（月額上限の引上げ幅×3）

Canonical term	First-use definition	Japanese source term
	by the government	
between-caps range	the interval [repeat-month cap ÷ 0.3, fixed component ÷ 0.3) of monthly total medical expenditure	逆転区間
margin	the distance in yen from a month’s total medical expenditure to an endpoint of the between-caps range	余裕
nearest_edge	the unsigned distance to the closer of the two endpoints	nearest_edge
treatment profile	one of the 15 clinically specified courses of breast cancer treatment, P1 to P15	治療プロファイル
treatment phase	a contiguous segment of a treatment profile with a distinct cost regime, e.g. the anti-HER2 maintenance phase	治療フェーズ
stratum	a sensitivity axis whose levels are crossed to generate series	層
design dimension	an axis whose levels are enumerated in full and reported across rather than summarised over, because no single level represents the others	設計次元
sweep axis	an axis varied outside the main enumeration, over a finer grid than any stratum, to characterise the dispersion of one outcome; reported as a median and a range rather than level by level	掃引軸
standard monthly remuneration	the graded wage measure that determines the income category	標準報酬月額
Category A to Category E	the five income categories in force from 1 August 2026, ordered from Category A, the highest standard monthly remuneration, to Category E, exempt from municipal inhabitant tax. Where the main text names them descriptively, the highest category is Category A, upper-middle is Category B, and middle is Category C	区分ア から 区分オ
Category 1 to Category 13	the thirteen income categories in force from 1 August 2027, ordered from Category 1, the highest, to Category 13. Categories 3, 6, and 9 carry the same fixed component, reference amount, repeat-month cap, and annual cap as 2026 Categories A, B, and C respectively	第 1 区分 から 第 13 区分
P1 to P15	the code identifying a treatment profile. Table A6 gives the clinical setting of each and Table A1 the full specification; the main text names them descriptively in Table 1, with the code in parentheses	P1〜P15

5

A.1. Policy setting

A.1.1 Population and legal basis

The population is employees' health insurance (健康保険) enrollees under 70 years of age and their dependents, the group to which the income-category schedule of Article 42, paragraph 1 of the Enforcement Order applies. All income categories are used: the five in force from 1 August 2026 and the thirteen in force from 1 August 2027, 18 in total.

The institutional parameters derive from the Cabinet Order Partially Amending the Enforcement Order of the Health Insurance Act and Related Orders (Cabinet Order No. 240 of 2026), promulgated in the Official Gazette, Extra No. 169 of 29 July 2026. Article 1 of that Order governs the provisions effective 1 August 2026 and Article 2 those effective 1 August 2027. We also used Ministry of Health, Labour and Welfare Ordinance No. 117, Ordinance Partially Amending the Enforcement Regulation of the Health Insurance Act, promulgated in the same Extra No. 169 at pages 150 to 179. All amounts are in yen; medical expenditure is stated at full tariff, and out-of-pocket payment is obtained by applying the caps to 30% of that amount.

The consolidated text on e-Gov, the government legal database, had not incorporated Cabinet Order No. 240 as of 4 August 2026, the most recent order reflected there being No. 219; only pre-reform amounts can be verified against it. We therefore treated the Official Gazette PDF as the primary source. Text extraction from that PDF scrambles the reading order of the multi-column layout, at one point interleaving provisions of the Mariners' Insurance Act Enforcement Order into the middle of Article 2, so we used only a version whose reading order had been reconstructed from glyph coordinates.

A.1.2 Monthly out-of-pocket cap

Article 42, paragraph 1 of the Enforcement Order sets the monthly cap by income category (Table A7). For the three categories at or above 280,000 yen of standard monthly remuneration the cap takes an incremental form,

$$\text{monthly cap} = \text{fixed component} + (\text{total medical expenditure} - \text{reference amount}) \times 1\%,$$

in which the reference amount is exactly the fixed component divided by 0.3. In the post-reform values, 270,300 ÷ 0.3 = 901,000, 179,100 ÷ 0.3 = 597,000, and 85,800 ÷ 0.3 = 286,000. The reference amount is therefore the total medical expenditure at which the coinsurance amount equals the fixed component, and it coincides with the point at which the cap begins to bind. In any month whose total medical expenditure falls below the reference amount the cap does not bind and the patient pays the full 30%. The fixed component is a floor on the cap, not a floor on what the patient pays. This exact coincidence fixes the upper endpoint of the between-caps range defined in section A.6.

Fractions of a yen in the incremental term are rounded down below 50 sen and up at 50 sen or above (Enforcement Order, Article 42, paragraph 1).

A.1.3 Repeat-month cap

The repeat-month cap applies in a month in which high-cost medical benefit is paid for the fourth or later time within the rolling 12-month window ending with the current month (Enforcement Order, Article 42, paragraph 2 and related provisions). It is not a cumulative count over the course of

treatment. The 2026 reform left this amount unchanged in all five categories (Table A7). Benefit paid under the annual cap does not count toward the fourth occurrence, because the proviso to Article 42, paragraph 1, item 1 counts only benefit paid under Article 41, paragraphs 1 to 4.

A.1.4 Annual cap

The annual cap was newly created with effect from 1 August 2026. The amounts are set by the items of Article 42, paragraph 11 of the Enforcement Order; Article 41-3 sets the conditions for payment and does not itself state any amount. Article 41-3 establishes payment to the enrollee holding that status on the reference date, the calculation period (as defined in Article 41-2, paragraph 1), the annual household aggregate of patient charges, the aggregation requirement for those under 70, and the deduction of monthly high-cost medical benefit already paid.

The conditions of application are as follows. The calculation period is a fixed year running from 1 August to 31 July of the following year, unlike the rolling window of the repeat-month cap. Aggregation is at household level. Payment is retrospective, after the reference date of 31 July, and the cap never binds partway through a year. For those under 70, only patient charges of 21,000 yen or more incurred in the same month at the same hospital, clinic, pharmacy, or other provider are aggregated (Enforcement Regulation, Article 109, paragraph 1, item 2).

The treatment of dispensing costs under an outside prescription requires an explicit statement, because the statutory text and our implementation do not resolve to the same granularity. The Cabinet Order lists “one hospital, clinic, pharmacy, or other provider” in parallel, so on the face of the text a pharmacy is a unit separate from the prescribing facility and the dispensed portion would not be aggregated with the prescriber’s outpatient charges. We were unable to obtain the administrative notice that governs aggregation in practice, so that reading cannot be checked against operational rules. **Of the two possible treatments we computed only one.** The implemented calculation applies the 21,000-yen test to the coinsurance amount of the month as a whole rather than provider by provider, which corresponds to aggregating the dispensed portion with the prescriber’s charges. This is the more permissive of the two readings: a month that would fail the test provider by provider can pass it on the combined total, so the choice admits more months into the annual aggregate and makes the annual cap easier to reach than the separate-provider reading would. The separate-provider case was not computed as a full series, but its size is bounded rather than left open. The two readings can differ only where dispensing under an outside prescription is actually incurred, which is three of the twelve profiles built on the main path; in the other nine the pharmacy stream is empty and the readings coincide by construction. For those three we re-aggregated the retained decomposition month by month, testing the medical stream and the dispensing stream separately against the 21,000-yen floor. Because the Cabinet Order gives no rule for apportioning a capped monthly payment between providers, we did not apportion it proportionally, which would be an interior point rather than a bound; we took instead the largest attribution the absence of a rule permits, namely the lesser of the month’s out-of-pocket payment and the combined coinsurance amount of the streams that fail the floor on their own. In these data the two definitions coincide, because in every diverging month the monthly cap does not bind and the out-of-pocket payment equals the coinsurance amount. The divergence is confined to the bone-metastasis profile, the only one of the three whose monthly coinsurance amount reaches the floor at all, and there it is at most 15,151 yen per year against a smallest distance to the annual cap of 161,428 yen. In no profile and no income category does the choice of reading alter whether the annual cap is reached (Table A4). The scheme draws no distinction between medical and dental care, or between inpatient and outpatient care, in either the Cabinet Order or the Enforcement Regulation.

7

Item numbers do not follow the order of the amounts. Item 1 is 530,000 yen and item 2 is 1,680,000 yen. Item 1 is a residual provision covering “persons other than those listed in items 2 to 4”, and both Category C and Category D fall within it; no separate item is provided for Category D.

A.1.5 Timing of the income category determination

The monthly cap and the annual cap determine the income category at different times. The monthly cap uses the standard monthly remuneration of the month in which care was received, each item of Article 42, paragraph 1 carrying that wording separately. The annual cap uses the standard monthly remuneration of the month containing the reference date (Article 42, paragraph 11, items 2 to 4); the opening text of paragraph 11 contains no timing wording. Only the municipal-tax-exempt item is determined instead by the previous fiscal year’s municipal tax for the fiscal year containing the reference date. The category is fixed by standard monthly remuneration, a 50-grade scale from 58,000 to 1,390,000 yen under Article 40 of the Health Insurance Act, so an implementation that assigns categories from indicative annual income does not reproduce the statutory determination.

A standard monthly remuneration set at the regular annual determination applies to each month from September of that year to August of the following year (Health Insurance Act, Article 41, paragraph 2). Of the 12 months in the annual cap calculation period, therefore, the 11 months from September to the following July fall in the same determination cycle as the reference date, and only the initial August falls in the preceding cycle. **We held standard monthly remuneration constant across the calculation period and did not model occasional revision under Article 43, so a patient whose grade changes mid-period is not represented; this is restated as a limitation.**

A.1.6 The second stage, effective 1 August 2027

Article 2 of Cabinet Order No. 240 subdivides the monthly cap categories from five into thirteen and reduces the repeat-month cap from 44,400 to 34,500 yen in Category 12, the category for enrollees with standard monthly remuneration below 160,000 yen; the municipal-tax-exempt Category 13 retains 24,600 yen, so the reduction is not in the lowest category. The annual cap, by contrast, retains only five amounts, and their boundaries do not coincide with those of the thirteen monthly categories. Assignment of an annual cap to each of the thirteen categories is a derivation rather than a transcription: only five items set annual amounts and no provision sets an annual cap per monthly category, so we assigned them by containment of the standard monthly remuneration range, having first confirmed that each of the thirteen ranges falls within exactly one of the five items. The residual provision, item 1, covers two monthly categories in 2026 and five in 2027.

The referent of an item number also changes between the two stages. Item 4 in 2026 denotes municipal-tax-exempt persons at 290,000 yen, whereas item 4 in 2027 denotes enrollees with standard monthly remuneration below 160,000 yen at 410,000 yen, the former item 4 having been renumbered as item 5. The amending text takes the form “after item 3 of the same paragraph, add the following item” and contains no string “item 4”; matching item numbers by the order of the amounts gives the wrong correspondence. The five 2026 categories and the thirteen 2027 categories are not in one-to-one correspondence, and the only unique mapping is Category E to Category 13. The complete parameter table for the thirteen categories is given in Table A8.

8

A.2. Treatment profiles

A.2.1 Derivation

Fifteen treatment profiles, P1 to P15, span the breast cancer settings in which cost sharing is at issue: surgery alone, ductal carcinoma in situ, adjuvant chemotherapy with and without radiotherapy, endocrine maintenance, HER2-positive adjuvant and post-neoadjuvant treatment, metastatic HER2-positive treatment, triple-negative neoadjuvant treatment with immunotherapy, bone metastasis, ovarian function suppression, adjuvant and metastatic CDK4/6 inhibition, and total mastectomy. The clinical content was reviewed by breast surgeons and medical oncologists and finalised by adjudication. The record of which items were settled at review is given in this Appendix.

P14 is split into two subprofiles, P14a and P14b, distinguished by the endocrine backbone combined with the CDK4/6 inhibitor: an aromatase inhibitor for P14a, for which generic products are listed, and fulvestrant for P14b, for which the Drug Price List records no generic. The split is a subprofile distinction and not a stratum or a sensitivity axis, because the two backbones differ in whether a generic substitution is available at all. P14b receives a loading schedule that places two administrations in the first month.

Table A6 gives a one-line summary of each profile. Full specifications, comprising agents, doses, administration intervals, numbers of cycles, supportive care, tests, and observation periods, are given in Table A1.

A.2.2 Assumptions common to all profiles

Primary prophylaxis with granulocyte colony-stimulating factor is counted in the base case only for regimens carrying a febrile neutropenia risk of approximately 20% or higher, the level at which the American Society of Clinical Oncology recommends primary prophylaxis, with dose-dense regimens treated as high risk.

Anti-HER2 therapy uses the subcutaneous fixed-dose combination of pertuzumab and trastuzumab as the base, with the intravenous arm retained as P6. The clinical specification calls for the course to be followed to completion at 18 three-weekly administrations rather than truncated at 12 months, but the series generator uses a fixed 12-month horizon for every profile on the main path, so only the administrations falling inside that horizon are costed: **13 of the 18 at a day-1 start and 12 at a day-15 start**, identically in P6 and P7, with the remaining five or six falling beyond the horizon. This is entry 2 of Table A19.

Surgery is breast-conserving resection with sentinel lymph node biopsy using combined dye and radioisotope localisation in the base case, and the inpatient stay is 3 days for conserving surgery and 6 days for mastectomy.

Antiemetic support follows the guideline standard for the emetogenic risk of each regimen, with dexamethasone supplied as Decadron injection 6.6 mg, the sodium phosphate ester, and dosed by risk category as confirmed at clinical review: 9.9 mg on day 1 followed by 6.6 mg daily on days 2 to 4 for highly emetogenic regimens (anthracycline- or carboplatin-containing), 9.9 mg on day 1 followed by 6.6 mg daily on days 2 to 3 for moderately emetogenic regimens as the base with a single 6.6 mg dose carried as a sensitivity level, and a single 6.6 mg dose for low-risk regimens. Vials are not

9

shared between days, so 9.9 mg is costed as two vials. Each dexamethasone day is generated as a dated event, so a multi-day course beginning late in a calendar month attributes its later days to the following month. The assignment of regimens the review did not name individually to a risk category is provisional (section A.9). Carboplatin is approximated at a fixed 450 mg as the equivalent of AUC 5.

Bone-modifying therapy is denosumab 120 mg every 4 weeks in the base case, with zoledronic acid presented as the alternative.

Surveillance is minimal, comprising mammography once a year, with breast ultrasound every 6 months during anti-HER2 therapy.

Maintenance visits occur every 3 months in the base case, with a 6-month interval under refill prescription as a sensitivity level, and the frequency of scheduled visits, blood tests, and imaging is set at one of three levels with the middle level as the base.

Both taxane schedules are carried, weekly paclitaxel and 3-weekly docetaxel.

Asymptomatic bone scintigraphy and routine computed tomography are not counted in any profile. This is a substantive choice rather than an omission: the earlier version of the analysis carried a stratum for the presence or absence of bone scintigraphy, and one judgment depended on it, so the stratum was retired and that judgment was re-evaluated in the base stratum.

A.2.3 Differences between the reviewed specification and the costed content

The 15 profiles were fixed by clinical review, but the analysis costs what the series generator builds, and the two are not identical everywhere. Table A19 sets out every difference we have identified, with the direction of the discrepancy and, where one can be computed without re-running the analysis, an upper bound on what it could change. One entry, the observation period, remains undeterminable in magnitude for the truncated profiles and is restated in section A.4.4; entry 1, formerly undeterminable, was resolved by the rebuild of P8 during the analysis. The bounds are compared against the distance from each month’s total medical expenditure to the endpoints of the between-caps range, and against the distance from the annual cumulative to the annual cap; both distances are existing quantities of the analysis and are defined in sections A.5 and A.6.

A.3. Cost estimation

The cost of each treatment profile is built from three streams, drug products priced from the Drug Price List, procedures and tests priced from the fee schedule, and inpatient days priced from the inpatient fee band; every component is attached to a dated event and attributed to the calendar month containing that date. Within each stream the components were chosen by a selection rule fixed in advance, so that no component was chosen by an extremal criterion as the primary basis of selection.

A.3.1 The selection rule

A single generic name typically carries several strengths, presentations, and manufacturers, so the cost of a regimen is determined as much by the rule used to pick among them as by the regimen itself. A rule that selects by an extremal criterion, the cheapest product, the highest fee, or the first N candidates, embeds the direction of the eventual result in the input. The risk is sharpest for a study that contrasts brand and generic products: unless the

two arms are selected by the same rule, the contrast is an artefact of the rule rather than a property of the prices. The Drug Price List makes this concrete. In seven ingredient-and-strength pairs the highest listed generic price exceeds the lowest listed brand price, so a cost-minimising rule would pick a cheaper product on one side and a dearer one on the other and would report a difference that the rule itself created.

The rule has five clauses and a fixed order of decision. First, the candidate set is closed before any price is examined: every strength of the ingredient listed in the Drug Price List is enumerated, and restriction on grounds of clinical availability is permitted only where a reason can be written down. Second, among the combinations whose total dose reaches or exceeds the required dose, the combination with the least waste is chosen. Third, and only where two combinations tie on waste, the one with the lower total drug price is chosen. Fourth, brand and generic products are not selected between at all: both are computed and retained as the two levels of a sensitivity stratum. Fifth, every resulting row records the reason for the selection and its source.

Three features of the mechanism deserve statement because they are what keep the rule from collapsing into an extremal one. The tie-break of clause three is explicit and subordinate: it is evaluated only after the waste criterion has failed to discriminate, and each row records whether it fired, so a reader can separate the primary criterion from the tie-break. Candidate sets are built at the level of the individual product rather than the strength, so two brands sharing a strength are both carried instead of being collapsed to the cheaper. And identification runs down to the salt and the ester. The ingredient field of the price list folds salts and esters together, so for injectables the candidate set is re-narrowed by the full generic name; dexamethasone, for instance, is restricted to the sodium phosphate ester, which excludes dexamethasone palmitate, a preparation with a different indication that a search on the ingredient name alone would admit.

The scope of the combination search is set by the basis on which the dose is expressed, a closed set of three values. Where the dose is per square metre of body surface area or per kilogram of body weight, a quantity is computed and then assembled from the strengths on hand, so the combination search applies. Where the dose is per body, the listed strength is itself the approved unit of administration and nothing is assembled, so no combination search is performed and a single presentation is taken. The three values are enforced in code: a dose basis outside the set raises an exception rather than falling through to the unscaled branch, which prevents a misspelling or a full-width character from passing silently as a fixed dose. Agents that are costed as a flat amount per administration and therefore never enter the search, namely the oral endocrine agents, the LH-RH depot, and denosumab, are designated by brand name; their uniqueness in the price table is checked mechanically and a second matching row raises an exception.

The rule is consumed by every judgment that depends on where a month’s total medical expenditure sits relative to the endpoints of the between-caps range, and by the comparison between brand and generic arms that defines the secondary outcome of section A.6. Because both arms pass through the identical rule and differ only in the clause that filters on brand status, the comparison is not a product of the selection procedure.

A.3.2 Drug costs

Unit prices are those of the National Health Insurance Drug Price List effective 1 August 2026. Most items were read from the list as posted on 1 August; fulvestrant was read from the posting of 13 August, which is a later publication of the same list rather than a later price. We verified that the

11

fulvestrant entry, its brand name, strength, manufacturer, and price, is identical in the two postings, so the effective date is 1 August throughout and only the document version differs. Amounts are stated at full tariff.

Doses are expressed on one of the three bases described above. Body surface area and body weight enter only for the ingredients whose dose is expressed per square metre or per kilogram; carboplatin is an exception, being fixed at 450 mg as the equivalent of AUC 5, and is therefore outside the reach of the body-size stratum even though it is a cytotoxic agent.

The brand-to-generic switch is applied to the twelve injectable ingredients carried in the vial selection table. That table holds 69 rows, of which 39 are brand and 30 generic; in the nine remaining combinations no generic product is listed and the two arms are not distinguished, a fact recorded in the row rather than resolved silently. The oral endocrine agents, the LH-RH depot, and denosumab lie outside this scope: they are costed by brand name and take the same value in both arms even where generic products are listed. The consequence is bounded rather than assumed, and is entry 4 of Table A19.

A.3.3 Fee-schedule items

Procedure and test costs are taken from the medical procedure master of the fiscal 2026 fee schedule, the file dated 31 July 2026, whose provisions apply from 1 June 2026. Points are converted to yen at ten yen per point.

Items are designated by procedure code rather than selected by rank. The rule has two branches. An item is adopted only where the notified billing requirements determine it uniquely; where they do not, the item is not adopted but recorded as unjustified and carried as a two-level stratum, computed once with the item included and once without. Seven items fall in the second branch, among them the choice between tumour markers, the slice-count band of computed tomography, and whether the outpatient management supplement is billable on a day when an injection is given. The full inventory of adopted and rejected codes, with the number of candidates considered and the reason for each decision, runs to 215 rows and is given in Table A2.

The outpatient cancer chemotherapy management fee is handled differently. Its applicable band depends on facility accreditation and cannot be derived from the type of institution, and the notification setting out the accreditation criteria could not be obtained. Rather than choose a representative band, we enumerate all fifteen bands of the fee proper, excluding supplements and the paediatric bands, and carry them as a stratum crossed with the others. The fifteen levels are not reported one by one; they are collapsed into the summaries of section A.6 along with the other strata, so the uncertainty in the band appears as dispersion in the outcome rather than as a separate result.

Which profiles carry the fee at all is settled by the official interpretive record rather than assumed. Under the fiscal 2022 interpretive question-and-answer on the fee schedule (part 14, question 3, administrative communication of 22 June 2022, confirmed for this study with a hospital billing office), fee B001-2-12 is billable only against antineoplastic agents in the pharmacological classification, and not against fulvestrant, an LH-RH agonist, denosumab, or zoledronic acid. The fee is therefore carried only by the profiles that contain an injectable antineoplastic agent. A mechanical audit of all 15 profiles against this rule found no discrepancy with the implementation, and the fee is not an undetermined quantity in the sense of section A.9.

12

A.3.4 Inpatient costs

Inpatient care is costed on a fee-for-service basis using the acute general inpatient fee band 1, procedure code 190117710 at 1,874 points, that is 18,740 yen per day, multiplied by the length of stay. The registered treatment of length of stay is a mapping from the operation performed, not a sensitivity axis: 3 days for the two breast-conserving procedures and 6 days for the two mastectomy procedures. The series built by the principal generator, however, carry 5 days uniformly, a value retained from the earlier implementation; the mapping is applied in the rebuilt P8, which takes 3 days, and in P15, which takes 6. The discrepancy is entry 7 of Table A19, and recomputing every series built by the principal generator under the mapping is reported as a supplementary sensitivity in section A.7 and Table A13; it changes no judgment.

We did not use the diagnosis procedure combination system, for two reasons. The first is that the institution-specific coefficient has no single defensible value: it varies by hospital, and no published figure can stand in for it without an unsourced choice. The second is that switching to bundled payment would also bring inpatient drugs and tests inside the bundle, which would double-count them against the itemised build-up and would require reconstructing the cost model rather than adjusting it. In the ministry’s fiscal 2024 discharge survey cited in the main text, 1,767 hospitals were reimbursed under the bundled system, so for a patient treated at one of them the inpatient figures here are not what the hospital would bill. The magnitude at which the choice would matter is quantified by an equivalence threshold reported in section A.10 rather than asserted here.

A.3.5 Attribution of cost within the month

Monthly cost is retained decomposed into inpatient medical, outpatient medical, outside dispensing, and other. The decomposition is retained rather than discarded because it is what makes the aggregation-unit question of section A.1.4 answerable without regenerating the series: the two readings of the 21,000-yen floor differ only in how the streams are grouped, so the bound reported there is computed by re-aggregating these columns. It is not passed to the cap calculation, which receives a single monthly total; the 21,000 yen aggregation test of the annual cap is therefore applied to the coinsurance amount of the month as a whole, which is the single, more permissive reading set out in section A.1.4. Outside dispensing is non-zero only in the three profiles whose treatment is an oral endocrine agent.

A.3.6 Implementation notes

Cap computation and the rounding of fractions of a yen are implemented once, in a single shared function, and are not reimplemented at any call site; the rounding follows Article 42, paragraph 1 of the Enforcement Order, rounding down below 50 sen and up at 50 sen or above, and is applied to the incremental term only, the fixed component and the repeat-month cap being stated in whole yen by the Order itself. Two invariants are enforced by exception rather than by convention: a dose basis outside the closed three-value set halts the input build, and a brand-name designation that matches more than one row of the price table halts the series build. Analyses were run under Python 3.13.9 with numpy 2.4.6 and pandas 3.0.3 on Windows 11; the random seed is recorded as 20260811 although the computation makes no random draw.

A.3.7 Assumptions and boundary

The cost build-up has two boundaries.

13

The first concerns the reach of the brand-to-generic switch, and it follows from the rule system rather than from a judgment about drugs. The switch stops at the injectable products because tablets have no analogue of the primary criterion. For an injectable the waste-minimising clause does the work and the price tie-break is subordinate to it; a tablet generates no waste, so that clause is vacuous and the tie-break would be promoted to primary, which is to say the rule would become cost minimisation. That is the one thing the design forbids, and it would not be harmless here: the nine listed generics of the aromatase inhibitor straddle the brand price above and below, so a cost-minimising rule would move the brand arm and the generic arm in opposite directions and would manufacture the contrast the study sets out to measure. Rather than extend a rule to a class of products it was not built for, we left the oral agents outside the switch.

What that exclusion could change is bounded rather than assumed, and where the bound is not decisive we compute the substitution directly. For each excluded agent we take the widest substitution the price list admits and sum over the agents that can occur in the same calendar month, which gives a per-month bound and a per-period bound for each affected profile. Under the confirmed dispensing schedules the bound alone no longer settles every case: it exceeds the distance to the point at which the monthly cap begins to bind in some combinations of the bone-metastasis profile and of the adjuvant CDK4/6 profile, and, taken at 30%, it exceeds the distance from a year’s cumulative to the annual cap in some combinations of the two CDK4/6 profiles that carry an aromatase inhibitor. A bound that fails to clear the gap shows only that a substitution could matter, so for every affected profile we rebuilt the series at the cheapest and at the most expensive listed generic price and re-evaluated the judgments themselves.

Across 5,184 profile-by-category-by-start combinations, the set of years reaching the annual cap is unchanged everywhere. The set of qualifying months is unchanged everywhere except in 30 combinations of the adjuvant CDK4/6 profile, all of them in the highest income category, 2026 Category A and its 2027 counterpart Category 3, and all of them under the substitution that raises the price, since the most expensive listed generic of the aromatase inhibitor is dearer than the brand. In those 30 combinations the out-of-pocket total over the 24-month course moves by between 2,276 and 2,399 yen. No profile built by the principal generator is affected in either judgment, so no pair entering the comparison of section A.6.4 is affected.

The second concerns the bounds themselves. The upper bounds used throughout this section and in Table A19 are deliberately wider than the selection rule would permit. They admit substitutions the rule excludes, such as two half-strength tablets standing in for one whole-strength tablet, or two low-dose depot injections standing in for one of the strength actually used, and they apply the maximum per-administration difference to every administration that could fall in a single calendar month. The wide band is deliberate: a conclusion that survives a bound looser than the analysis itself does not depend on the analysis’s own conventions. Where the wide and narrow bands differ materially the section states both.

A.4. Construction of the monthly cost series

A.4.1 Unit of construction

Every quantity in sections A.1 and A.6 is defined on a calendar month: the monthly cap binds within a month, the repeat-month cap counts months, and the annual cap sums months inside a fixed year. A cost model that takes the month as its unit of construction therefore looks natural and is wrong for this question. Chemotherapy and antibody regimens are prescribed on cycles of 21 or 28 days, and a cycle length that does not divide the calendar produces months holding one administration and months holding two. That alternation is not noise around a monthly average; it is the mechanism by

14

which a patient’s monthly total medical expenditure moves in and out of the between-caps range defined in section A.6, and a model that assigns one cycle to one month cannot represent it at all.

We therefore generate each administration, each test, each outpatient visit, and each inpatient day as an event carrying a date, and attribute its cost to the calendar month containing that date. Doses follow the cycle length of the regimen from the start date. Oral agents are dispensed at a fixed supply interval and the cost of a dispensing is attributed in full to the month in which it occurs. Follow-up events scheduled n times a year are placed at intervals of 365 divided by n days, a rule that had to be introduced because the earlier construction assigned them to month indices directly and so had no dates to attribute. Admissions are attributed in full to the month containing the first inpatient day; at the lengths of stay used here, 5 days in the profiles built by the principal generator and 3 or 6 days under the mapping of section A.3.4, no admission crosses a month boundary at either start day.

A.4.2 The start day within the month

The day of the month on which treatment begins determines where the two-administration months fall, so it is a parameter of the design and not a property of the patient. It is a nuisance parameter: it has no clinical interpretation, it is not something a patient or a clinician chooses with the cost consequence in view, and it is not a quantity about which the study makes any claim. It is varied because leaving it fixed would let an artefact of the calendar masquerade as a property of the regimen.

It is varied in two different ways, for two different purposes, and the distinction matters because the two feed different outcomes. As a stratum it takes two levels, the first and the fifteenth day of the start month, and these are crossed with the other strata in the enumeration that supports the primary outcome of section A.6. Separately, and outside that enumeration, the start day is a sweep axis taking all 28 values from the first to the twenty-eighth day, used for the secondary outcome, because that outcome proved to depend on the start day strongly enough that a two-level stratum could not characterise it; the sweep is what allows the secondary outcome to be reported as a median with a range rather than as a single value. Which of the two the base case fixed in advance represents, and why the first day is the conservative choice for the secondary outcome, is taken up in section A.6.

A.4.3 The start month as a design dimension

The calendar month in which treatment begins is enumerated over all twelve values. This is not a sensitivity analysis. The annual cap runs on a fixed year from 1 August to 31 July rather than on a rolling window, as set out in section A.1.4, so the calendar start month determines how a course of treatment is cut across August–July years: the same treatment beginning in different months can fall wholly inside one year or be divided across successive years, and the division changes how much of the cumulative total is exposed to the cap. A 12-month course occupies a single year only when it begins in August and two years otherwise; the longer horizons span up to three. The calendar start month is therefore a dimension of the design in the same sense as the income category, and results are reported across it rather than averaged over it.

A.4.4 Observation period and computational horizon

The observation period is a design parameter fixed per profile in advance of computation, not a sensitivity axis, because moving it changes the number of August–July years a course spans and can change the sign of the difference the study reports. The registered periods are 12 months for most profiles,

15

15 for the triple-negative neoadjuvant profile, 16 for the rebuilt P8, 18 for the two HER2-positive adjuvant profiles, 24 for the adjuvant and metastatic CDK4/6 profiles, and 30 for the metastatic HER2-positive profile. The 16 months of P8 are a derivation from its treatment schedule, registered during the analysis before the profile was generated: five months of neoadjuvant TCHP, one surgical month, and ten months holding the 14 three-weekly cycles of trastuzumab emtansine.

The series generator does not implement those periods. It carries a single 12-month horizon for every profile it builds, and the profiles added later are built by separate routines: 24 months for the CDK4/6 profiles, 12 for the mastectomy profile, and 16 for the rebuilt P8, whose full 16-month series is generated as a record while the judgments reported in this paper evaluate it on the common 12-month horizon. The registered periods and the horizon in force therefore diverge for four profiles, by 6 months for each of the two HER2-positive adjuvant profiles, by 18 months for the metastatic HER2-positive profile, by 3 months for the triple-negative neoadjuvant profile, and by 4 months for P8. The divergence is recorded rather than reconciled, as entry 2 of Table A19.

The scope of the discrepancy is settled even though its magnitude is not. Truncation removes months; it does not alter the months that remain. The costs attributed to months 1 to 12 are identical under either horizon because attribution depends only on the dates of events, not on where the series ends. The repeat-month cap looks backward from the current month over the preceding twelve, so the window seen by any month at or before month 12 contains only months that exist in both versions. The annual cap is settled retrospectively at the end of its year and never binds partway through, so it cannot alter a monthly payment inside the horizon. Every difference is therefore confined to the absence of months 13 and later.

From this the direction follows for one outcome and fails to follow for another. Adding months to a year cannot reduce that year’s cumulative total, so a longer horizon can only bring an annual cumulative closer to the annual cap or past it, never further away: attainment of the annual cap is understated by the truncation, one-sidedly. The proportion of months lying inside the between-caps range carries no such guarantee, because lengthening the series changes the denominator as well as the numerator, and the added months are the maintenance months, whose position relative to the interval is exactly what is at issue. For that outcome the direction is undetermined as well as the magnitude. Bounding either would require regenerating the series over the registered horizons, which we did not do.

A.4.5 Composition of a series

A series is one profile at one level of each applicable stratum, generated over the horizon at one start day, and then placed at one of the twelve start months and evaluated under one of the eighteen income categories. Strata are crossed within a profile and differ between profiles, so the total is a sum of per-profile products rather than a single product; the composition is given in section A.7 and the full breakdown in Table A3. Monthly totals are retained together with the four-way decomposition described in section A.3.5, and it is the monthly total that is passed to the cap calculation of section A.5.

A.5. Calculation of out-of-pocket payment

The three instruments of section A.1 do not act independently, and the order in which they are applied determines the result. The calculation is implemented once, in a single shared function, in the order set out here; no step is reimplemented at any call site.

16

A.5.1 The order within a month

For each month in turn, the coinsurance amount is taken as 30% of that month’s total medical expenditure. The ordinary cap is then formed: for the income categories whose cap has an incremental term, it is the fixed component plus 1% of the excess of total medical expenditure over the reference amount, with fractions of a yen resolved as described in section A.3.6; for the categories without an incremental term it is the fixed component alone.

The repeat-month cap is applied next, and it displaces the ordinary cap rather than being applied after it. The year is the 12 months ending with the current month, under the provisions cited in section A.1.3. The current month itself cannot count towards its own qualification, so the test looks at the 11 preceding months: where high-cost medical benefit was paid in three or more of them, the current month is the fourth or later occurrence and the applicable cap for the month becomes the repeat-month amount.

The out-of-pocket payment for the month is then the lesser of the coinsurance amount and the applicable cap. Finally the month is recorded as a qualifying month, or not, and that record is what the following months read when they evaluate their own years.

A.5.2 Treatment of the equality case

A month qualifies when the coinsurance amount **exceeds** the applicable cap. A month in which the two are exactly equal does not qualify: no benefit is paid, the month does not enter the count that the following eleven months read, and the patient pays the coinsurance amount, which in that month happens to equal the cap. This is not a rounding convention but the same convention that makes the between-caps range of section A.6 closed on the left and open on the right, and the two must agree because the upper endpoint of that interval is precisely the total medical expenditure at which coinsurance equals the fixed component. Stating the calculation only as the lesser of the two quantities would leave the equality case unresolved, and this study is about the neighbourhood of exactly that point.

A.5.3 Annual layer

The annual cap is applied after the monthly layer has closed. Before the annual layer is entered, the monthly series is placed on the calendar grid so that each month contributes to the statutory 1 August to 31 July year that contains it; depending on the calendar start month and the horizon, a course spans one to three August–July years, and a year receives only the months that fall inside it. Each month then contributes its out-of-pocket payment to the aggregate for its year, subject to the 21,000-yen floor described in section A.1.4; the running total is compared with the annual cap for the income category, and any excess at the end of the year is refunded. Annual out-of-pocket payment is the sum of the monthly payments less that refund.

Two consequences follow from the ordering rather than from any separate provision. Payment made under the annual cap cannot count towards qualification for the repeat-month cap, which the proviso cited in section A.1.3 also settles as a matter of law, because the record of qualifying months is complete before the annual layer is entered and the annual layer never writes back to it. And the annual cap cannot alter any monthly payment, because it is evaluated only at the close of a year; a patient who will eventually receive an annual refund pays the unreduced monthly amounts throughout.

17

A.5.4 The two reform states

The whole calculation is run twice for every series, once with the pre-reform parameters and once with the post-reform parameters, and the two are differenced. The annual cap exists only after the reform, so the pre-reform run has no annual layer at all; it is not run with an annual cap set to infinity but with the layer absent, which is the same thing numerically and clearer in the record. Differences are reported as post-reform minus pre-reform, so that a positive value is an increase in what the patient pays.

A.6. Outcome definitions

A.6.1 The between-caps range

For an income category with fixed component **F** and repeat-month cap **M**, the **between-caps range** is the interval of monthly total medical expenditure over which the coinsurance amount lies in the half-open band **[M, F)**, that is **[M ÷ 0.3, F ÷ 0.3)**. In a month whose total falls in it, the coinsurance amount is at least the repeat-month cap but below the fixed component, so the monthly cap does not bind and the month does not qualify.

The interval is half-open on the right because it is defined as the pre-image of a half-open band of the coinsurance amount, and not because of the strict inequality of section A.5.2. That inequality points the other way: a month sitting exactly at the upper endpoint also fails to qualify and so shares the interval's defining property, which would be an argument for closing the interval there. The convention is nonetheless immaterial here, because no month in any series has a total exactly equal to an upper endpoint in any of the 18 income categories.

A further statement holds for a patient **all of whose months fall in the interval**. No month qualifies, so the count that governs the repeat-month cap never advances, and the reduction available from the fourth qualifying month is never reached however long treatment continues. Such a patient pays 30% of total medical expenditure throughout. Compared with a patient whose monthly total exceeds the upper endpoint and who has reached the repeat-month cap, the patient inside the interval pays at least as much per month, and strictly more whenever the monthly total lies above the lower endpoint; at the lower endpoint itself the coinsurance amount equals the repeat-month cap and the two pay the same. The difference is bounded above by **F − M**.

A.6.2 Margin

For a month and an income category, the **margin** is the signed distance in yen from the month's total medical expenditure to an endpoint of the between-caps range; both endpoints are retained with their signs so that approach from below and approach from above can be distinguished. Where a single quantity is needed, nearest_edge is the unsigned distance to the closer endpoint. Distances to the upper endpoint are the ones that bear on whether a month qualifies, since that endpoint is where the monthly cap begins to bind; distances to the lower endpoint bear on interval membership but not on qualification, and the two are not interchangeable.

18

A.6.3 Primary outcome

E4 is, for a treatment profile and a treatment phase, the proportion of **treatment-months** whose total medical expenditure falls in the between-caps range, where a treatment-month is one month of one series. The proportion is pooled: numerator and denominator are counted over all series in the combination rather than formed series by series and averaged. Only months in which cost was actually incurred enter the denominator. Excluding a zero-cost month can only raise the proportion, since such a month lies below the lower endpoint and is never inside the interval; in the designated combination the exclusion removes no month from either of the two rows that meet the criterion, and every other row has a numerator of zero, so the convention does not affect any designation.

The combination is the **base stratum**, which fixes follow-up frequency at the middle of its three levels and body size at the middle of its three levels and leaves every other stratum crossed, including the fifteen bands of the outpatient cancer chemotherapy management fee. The start day is fixed at the first of the month and the start month at August.

Neither of those two fixed choices changes the designation, though one of them changes the proportion. Across all twelve start months the proportion is identical, with numerator 4,320 and denominator 7,200 unchanged. We report that as a measurement rather than deriving it: the multiset of monthly totals in the maintenance phase is not itself invariant, differing at four of the twelve start months, so the constancy of the count inside the interval does not follow from the construction and is an empirical property of these series. Moving the start day from the first to the fifteenth does change the proportion, from 0.600 to 0.667, and the change is entirely in the denominator: the numerator stays at 4,320 while the count of months attributed to the maintenance phase falls from 7,200 to 6,480, with the difference moving to the adjuvant chemotherapy phase, whose count rises from 8,640 to 9,360. The count of months lying inside the interval is invariant to the start day and to the start month alike.

A month is assigned to exactly one phase and is never divided. Where more than one phase is active within a month, the month is assigned by the fixed priority order **surgery, neoadjuvant chemotherapy, adjuvant chemotherapy, anti-HER2 maintenance, radiotherapy, metastatic or recurrent drug therapy, endocrine maintenance, follow-up only**; a month in which none is active is assigned to follow-up only. Because the assignment is exclusive, the phase-specific proportions are not independent of one another and their sum carries no meaning.

A profile and phase is **designated** when the proportion is 0.50 or greater under the 2026 Category C parameters in the base stratum. Proportions for all 18 income categories are given in Table A18, but the designation is defined on that one combination. The threshold is not consequential: in that combination two rows take the value 0.600 and the remaining 27 are exactly zero, with no intermediate value, so any threshold greater than zero and at most 0.600 designates the same pair.

A.6.4 Secondary outcome

E5 is the proportion of brand-and-generic series pairs whose **annual out-of-pocket payment is higher in the generic arm**. The denominator is all 51,873 pairs, including those whose two arms have identical total medical expenditure. As implemented the numerator conditions on the out-of-pocket difference alone and not on total medical expenditure being lower. The two conditions coincide in these data: the generic arm’s total is lower in 50,220

19

pairs, equal in 1,653, and higher in none. That no pair reverses is a measurement; the reason appears to be that the seven ingredient-and-strength pairs whose listed generic price exceeds the brand price are not the ones a minimum-waste criterion selects.

The switch is applied to the **12 injectable ingredients** of the vial selection table and to those only. The table has 69 rows, and a row is not an ingredient but a combination of ingredient, brand-or-generic side, dose occasion, and body-size level: 39 rows are brand and 30 generic, and in the 9 combinations without a listed generic the two sides are not distinguished. The oral endocrine agents, the LH-RH depot, and denosumab take the same price in both arms, so **E5 is a statement about injectable substitution and not about generic substitution in general**.

What the excluded agents could do is settled by a chain rather than left open. In a month that does not qualify the patient pays 30% of total medical expenditure, which is monotone in that total. In a month that qualifies the patient pays the cap, whose incremental term is also monotone in that total. And provided the set of years in which the annual cap is reached is likewise unchanged, annual out-of-pocket payment is monotone in total medical expenditure. section A.3.7 establishes, by direct computation of the widest substitution the price list admits, that extending the switch to the excluded agents changes the set of years reaching the annual cap in no profile, and changes the set of qualifying months in no profile that enters this comparison; the sole movement, thirty combinations of the adjuvant CDK4/6 profile, arises only under a price-raising substitution and lies outside the pairs of this outcome. No substitution among those agents can therefore raise annual out-of-pocket payment while lowering total medical expenditure.

The phenomenon is **designated as present** in an income category when the proportion is 0.1% or greater. Because the proportion depends strongly on the start day, it is reported as the median over the 28-point start-day sweep together with the minimum and the maximum; the base case of the first of the month is not the median but sits systematically at the conservative end of the range, and is reported as such rather than as a point estimate. The threshold is not consequential here either: across all 28 start days the eight designated categories have a minimum of 0.0349 and the ten undesignated categories are exactly zero, so any threshold greater than zero and not exceeding that minimum produces the same partition. Removing the 1,653 equal-cost pairs from the denominator multiplies every proportion by 1.033 and, the zeros being exact, leaves the partition unchanged; both denominators are reported. The designation is evaluated under the base neoadjuvant arm of P8, a design assumption fixed in advance; its dependence on that assumption is set out in section A.7 and reported with the outcome.

A.6.5 Difference in annual out-of-pocket payment

The difference in annual out-of-pocket payment is reported as **post-reform minus pre-reform**, so that a positive value is an increase in what the patient pays and a negative value a decrease. The pre-reform state has no annual cap, as set out in section A.5.4. Where annual-cap attainment is reported, three qualifications from earlier sections apply: attainment is understated by the 12-month horizon in the direction described in section A.4.4; the choice of aggregation unit for the 21,000-yen floor is the more permissive of the two readings, bounded as described in section A.1.4; and the profiles that never attain the cap do so because their monthly coinsurance amount rarely reaches the floor at all, not because their total cost is low.

Table 1 reports the change in out-of-pocket payment over the course, post-reform minus pre-reform, for each main-path profile in each of the five 2026 income categories. The change is computed for every course start in the brand arm. A course start is one series begun on one of 24 start dates, the 1st or the 15th of each calendar month. For each profile and category, Table 1 gives the median, the minimum, and the maximum over all course starts.

20

Where the number of values is even and the two middle values differ, the median is their midpoint. The change is read against a benchmark increase, three times the rise in the monthly cap for the category (Table A7). The benchmark is 51,330, 33,930, 16,530, 11,700, and 4,500 yen in Categories A to E. It is the additional payment of a patient who reaches the monthly cap in the first three qualifying months and then pays the unchanged repeat-month cap. It is not a value published by the government. Where the change at a course start differs from the benchmark, it is divided month by month into four components. The first is the difference in the monthly cap in months in which the monthly cap applies before and after the reform and the repeat-month cap applies in neither. The second is the difference in months in which the monthly cap applies before the reform but not after it. In those months the patient now pays the full 30% coinsurance. The third is the difference in months in which the repeat-month cap applies before the reform but not after it. The fourth is the refund under the annual cap, which exists only after the reform and is never positive. No other kind of month contributed a non-zero amount, and the four components sum to the change at every such course start. The results are in results/reform_gap/. T2_scheduled.csv holds the benchmark, T3_cells.csv the change at every course start, and T4_decomposition.csv the components. T7_direction_counts.csv holds the counts by direction of the change, and Table1_candidate.csv the values in Table 1. This comparison is not one of the two measures fixed in advance (sections A.6.3 and A.6.4); it was defined after the first results had been seen.

A.7. Sensitivity analysis and strata

A.7.1 The strata and their levels

Strata are crossed within a profile and differ between profiles, as set out in section A.4.5. The axes and their levels are: surgical procedure, 4 levels, for the profiles built by the principal generator that contain surgery, the rebuilt P8 taking breast-conserving surgery as a fixed choice; multigene assay, 2 levels, where surgery and chemotherapy coincide; the outpatient cancer chemotherapy management fee, 15 levels, where an injectable antineoplastic agent is present (section A.3.3); the aseptic preparation and injection fees, 2 levels, on the same condition; body size, 3 levels; the antiemetic combination, 3 levels; the trastuzumab deruxtecan cycle count, 6 levels, in P9 only; the taxane schedule, 2 levels, where a taxane follows anthracycline chemotherapy; follow-up frequency, 3 levels; the P8 imaging add-on, 2 levels; the maintenance dispensing interval, 2 levels, in the endocrine profiles; and the start day within the month, 2 levels. Brand and generic arms are built as parallel enumerations rather than as a stratum, and the calendar start month and the income category are design dimensions (section A.4.3 and Methods 2.3). One full enumeration comprises 51,873 series per arm at one start day. The per-profile composition is given in Table A3.

The three body-size levels are bands that bracket the plausible range of body surface area and body weight, with the middle level as the base (Table A14). They are bounds, not a distribution: no weighting is applied across them, no population frequency is implied, and results at the outer levels are read as an envelope rather than as percentiles.

A.7.2 Retired axis: granulocyte colony-stimulating factor

Granulocyte colony-stimulating factor was initially carried as a 2-level stratum (guideline-based counting versus counting also for the docetaxel-cyclophosphamide regimen) and is not swept in the analysis reported here. The decision to drop the axis was taken after results had been seen, and is recorded as such rather than presented as an original design. The recorded rationale is that the chemotherapy months in every affected profile already

21

lie at or above the upper endpoint of the between-caps range, so adding the agent cannot move their interval position; the unswept axis is restated as a limitation, and the record of the amendment is Table A15.

A.7.3 The inpatient-stay mapping

Recomputing every series built by the principal generator with the length of stay set by the registered mapping, 3 days for breast-conserving procedures and 6 for mastectomy in place of the uniform 5 days, changes no judgment: annual-cap attainment (any or none), between-caps-range membership (any or none), and the modal sign of the reform difference are unchanged in all 216 profile–category combinations. Counts alone move, in 18 combinations, and the full comparison is reported as a supplementary sensitivity. This sweep is reported alongside the base rather than mixed into it.

A.7.4 The P8 neoadjuvant arm

The clinical review specified two neoadjuvant regimens for P8 without stating a priority. The base is TCHP; the alternative, doxorubicin and cyclophosphamide followed by a taxane with trastuzumab and pertuzumab, is computed as a sensitivity arm carrying both taxane schedules. The choice of base is a design assumption fixed before any result was computed, and its consequences are not uniform across the outcomes. The modal sign of the reform difference is unchanged in all 18 income categories. Whether the annual cap is attained at all differs in 5 categories, and whether any month lies in the between-caps range differs in 4, all of them combinations of P8 itself. The set of income categories in which the secondary outcome is designated as present depends on the arm: under the base arm it is the eight categories reported with the outcome, and under the sensitivity arm it extends to Category A and Categories 1 to 3, carried entirely by P8 pairs. The category partition of the secondary outcome is therefore reported conditional on the base arm, and the regimen priority is under clinical confirmation (section A.9).

A.7.5 The CDK4/6 dispensing schedule and its switch point

The dispensing interval for the oral CDK4/6 inhibitors is 14 days during an induction period and 28 days thereafter, as confirmed at clinical review; the earlier uniform 90-day interval is retained only as an outlier scenario. The day on which induction ends has no external anchor, in the product labelling or elsewhere, and is a design assumption: the base is 2 cycles (56 days), chosen on cycle rather than calendar grounds so as not to confound the start-day stratum, with 1 cycle and 3 cycles as sensitivity levels. Across the three switch points, between-caps-range membership and annual-cap attainment are unchanged in all 1,296 combinations; the sign of the reform difference flips in 8 combinations, every one of them a day-1 start in calendar February or March, and the out-of-pocket total over the 24-month course moves in 568 combinations with a maximum of 557,012 yen. The three summary judgments are therefore robust to the switch point severally, and the sign is not; the 8 combinations are named in Table A16.

A.7.6 The start day

The start day enters twice, as described in section A.4.2: a 2-level stratum in the main enumeration and a 28-point sweep axis for the secondary outcome. The first of the month, the base fixed in advance, sits systematically at the conservative end of the sweep: it is the minimum itself in six of the eight categories in which the secondary outcome is designated, and in the remaining two it ranks fourth of 28, the minimum falling on the ninth. Both the position and the ranks are unchanged from the pre-rebuild series.

22

A.8. Analysis plan and reproducibility

The outcome definitions and their thresholds are fixed as constants in the analysis code, 0.50 for the primary designation and 0.001 for the secondary, and no separate analysis-plan document survives alongside them; the record that a definition, threshold, or design assumption was fixed before the corresponding result was computed is the dated decision log and the provenance columns of the parameter files themselves. Design assumptions are registered in a machine-checked allowlist of 93 entries; the loader halts the computation when a parameter row carrying an unregistered assumption, or a superseded row, would enter it.

The computation is deterministic: it makes no random draw, and the recorded seed is vestigial. Rerunning the registered pipeline from a populated state reproduces byte-identical content in all 95 files it regenerates; the remaining 197 files in the results tree are not written by any registered step, and their identity carries no evidence, so the reproducibility claim is confined to the 95.

What is verified is idempotency, not reproduction from scratch, for three reasons. First, the pipeline reads three artifacts generated by earlier-phase scripts that are not registered steps, so an empty results tree cannot be repopulated by the pipeline alone. Second, one registered step is an audit of existing outputs and cannot run against an empty tree by construction. Third, the input parameter files under the parameter directory are outside the pipeline’s scope: the scripts that would regenerate them from their sources overwrite manually sourced rows and are deliberately excluded from the pipeline, so reconstruction of the inputs is not part of the claim. Within its stated scope the pipeline is idempotent; outside it, provenance is carried by the source columns of the parameter files rather than by re-execution.

Two corrections made during the analysis are documented with before-and-after values in Table A17 rather than silently absorbed: the alignment of the August–July years to the statutory 1 August to 31 July calendar, which superseded an earlier computation that cut years from the treatment start month, and a transcription error in an interim log detected when an anchor value failed to reproduce. The manuscript reports only the corrected values, and every number in it is read from a result file rather than from a log.

Two further records bear on the order of decisions. The retirement of the granulocyte colony-stimulating factor axis (section A.7.2 and Table A15) is recorded in the pre-analysis document, file prereg_phaseG.md, item DE-2, revision dated 12 August 2026, which states that the revision was made after results had been seen. The rolling-window counterfactual reported in Results section 3.5 and the Discussion (Table A21) was specified on 15 September 2026, after every other result in this paper had been computed. Its definition document, counterfactual_rolling.md, and its allowlist entry CF_ROLLING_1, of kind post_hoc_counterfactual, are kept with the counterfactual code outside the registered pipeline, and with the year origin set to 1 August the same code reproduced the reported attainment and sign results exactly.

A.9. Undetermined items

Three items remain undetermined at the time of writing. Each is registered as provisional in the parameter files, so the halting loader of section A.8 will surface any unregistered successor.

23

First, the assignment of four regimens to an emetogenic risk category. The clinical review named the anthracycline- and carboplatin-containing regimens as highly emetogenic and the docetaxel-cyclophosphamide regimen as moderately emetogenic; the assignments of the single-agent taxanes, trastuzumab deruxtecan, trastuzumab emtansine, and pembrolizumab follow the general guideline classification and await confirmation. Reassigning any one of them between adjacent categories changes at most 724 yen per administration; the schedule change that introduced these amounts moved no between-caps-range designation, no annual-cap attainment at the profile-by-category level outside the rebuilt P8, and no modal sign, so a reassignment of the same magnitude cannot move those judgments either.

Second, the surveillance imaging frequency of the rebuilt P8. Annual mammography is the base and a semiannual breast ultrasound add-on is the sensitivity level; both levels are enumerated in the stratum product, so the open question is which is standard practice, not what either costs. The difference is two ultrasound examinations per year and was noted at clinical review as small.

Third, the priority between the two neoadjuvant regimens of P8. Unlike the first two items, this choice moves identified judgments, set out in section A.7.4, and the affected outcome is reported conditional on the base arm until the confirmation, which has been requested without disclosure of the computed values, is returned.

One further negative statement belongs here. For the four profiles built by the separate routines, P13, P14a, P14b, and P15, the aggregation-unit test of section A.1.4 cannot be applied, because that path does not retain the four-way cost decomposition; all four dispense at least one oral agent, so the two readings do not coincide by construction for any of them. The rebuilt P8 is not affected: its outside-dispensing stream is empty by construction, so the two readings coincide.

A.10. Clinical review record and bundled-payment equivalence threshold

Reviewer consolidation record. What clinical review settled is recorded where it bears on the analysis rather than gathered here: section A.2.3 and Table A19 set out every difference between the reviewed specification and the costed content, sections A.7.4 and A.7.5 give the amendments that arose from review, and section A.9 lists the items still open, to which the limitations in the Discussion refer.

Bundled-payment equivalence threshold. Inpatient care is costed fee-for-service (Methods, Study background). We did not switch to the bundled diagnosis procedure combination payment, for two reasons: the institution-specific coefficient has no single defensible value, and bundling would bring inpatient drugs and tests inside the bundle, requiring the itemised build-up to be reconstructed rather than adjusted. Instead of asserting a direction, we solved for the inpatient payment that would matter. For every combination of income category, surgical procedure, and length of stay in which the classification of the surgery month depends on whether the inpatient fee is counted at all, we computed the per-day inpatient payment that would carry the total medical cost of the surgery month across a between-caps-range endpoint. Across the 31 crossings so enumerated, the lower endpoint is reached at 658.83 to 1,317.67 points per day, that is 6,588 to 13,177 yen, and the upper endpoint at 532.17 to 23,751.00 points per day, that is 5,322 to 237,510 yen. At the registered length of stay, three days for the two breast-conserving procedures and six for the two mastectomy procedures, the lower endpoint is reached at 1,317.67 points per day in all four qualifying combinations and the upper endpoint at 1,064.33 to

24

23,751.00. No bundled per-diem rate and no institution-specific coefficient enters this calculation. The threshold admits the comparison this analysis does not make: a per-diem rate and an institution-specific coefficient for a particular hospital can be set against it.

Supplementary tables

Table A1. Full specification of the 15 treatment profiles*

The table lists, for each profile P1 to P15, every agent with its dose, dose basis, cycle length, number of cycles and treatment setting; the supportive-care regimen; the scheduled tests; the surgical procedure and length of stay where applicable; and the registered observation period. It is generated mechanically from the archived analysis files (38 rows), the archived analysis files (14 rows), the archived analysis files (2 rows), the archived analysis files (3 rows), the archived analysis files, and the archived analysis files (15 rows), each row of which carries its own source column.

* Doses are the planned totals under complete delivery; dose reduction, delay, and discontinuation are not modelled. Rows whose superseded_by field is populated are excluded from the listing and are retained in the parameter files as a record. Source: the parameter files named above, as adjudicated at clinical review and subsequently amended.

Table A2. Fee-schedule items: adopted, rejected, and carried as unjustified*

215 rows, one per procedure code considered, giving the code, its name, the points and the yen equivalent at ten yen per point, the number of candidate codes in the same category, the decision (adopted, rejected, or carried as a two-level stratum), and the reason. Seven items fall in the third class, among them the choice between tumour markers, the slice-count band of computed tomography, and whether the outpatient management supplement is billable on a day when an injection is given.

* Points are converted at ten yen per point. Source: medical procedure master of the fiscal 2026 fee schedule (file dated 31 July 2026, provisions applying from 1 June 2026), via the archived analysis files and the archived analysis files.

Table A3. Series counts by treatment profile*

Profile	Series (per arm, one start day)	Strata crossed
P1	12	surgical procedure (4) × follow-up frequency (3)
P2	12	surgical procedure (4) × follow-up frequency (3)
P3	6,480	surgical procedure (4) × multigene assay (2) × E11 (15) × E12 (2) × body size (3) × antiemetic (3) × follow-up frequency (3)
P4	6,480	as P3
P5	3	follow-up frequency (3) only
P6	12,960	as P3, with taxane schedule (2)
P7	12,960	as P6

25

Profile	Series (per arm, one start day)	Strata crossed
P8	1,620	E11 (15) × E12 (2) × body size (3) × antiemetic (3) × follow-up frequency (3) × P8 imaging add-on (2)
P9	4,860	E11 (15) × E12 (2) × body size (3) × antiemetic (3) × T-DXd cycle count (6)
P10	6,480	as P3
P11	3	follow-up frequency (3) only
P12	3	follow-up frequency (3) only
Total	**51,873**	—

[*] Counts are per brand arm at one start day; the enumeration is run for both arms, at two start days, and placed at each of twelve calendar start months and evaluated under each of eighteen income categories. The third column lists every stratum that is crossed for that profile, so the levels shown multiply to the count. P5, P11, and P12 carry follow-up frequency alone: they contain no surgery, so the surgical-procedure and multigene-assay axes do not apply; no injectable antineoplastic agent, so the fifteen bands of the outpatient cancer chemotherapy management fee and the two levels of the aseptic-preparation and injection fees do not apply (Methods, Study background); no cytotoxic dosing by body size or antiemetic regimen, so those axes do not apply either. Their maintenance dispensing interval is a registered two-level axis that is not crossed in this enumeration, which fixes it at the base level; the same holds for P1 and P2, which carry the surgical axis but no drug axes. P8 is the rebuilt profile and is the only one carrying the imaging add-on (Methods, Study background and Variables). The total matches the pair denominator of the secondary outcome (Methods, Outcome). Source: the archived analysis files, the archived analysis files; assembled as the archived analysis files.

Table A4. The aggregation unit of the 21,000-yen floor[*]

Of the profiles that dispense under an outside prescription, only the bone-metastasis profile P11 has any month whose coinsurance amount reaches the 21,000-yen floor, so it is the only profile in which the two readings of the aggregation unit can diverge at all. In each of the 18 income categories, 20 profile-months are counted into the annual aggregate under the implemented reading, of which 2 agree between the readings and 18 diverge.

Quantity	Value	Applies to
Months counted into the annual aggregate	20	P11, each of 18 categories
Months on which the two readings agree	2	as above
Months on which they diverge	18	as above
Largest amount that could fall out of the annual cumulative	15,150.5	as above
Smallest distance from the annual cumulative to the annual cap	161,428.4	as above
Categories in which annual-cap attainment changes	0 of 18	as above

[*] All amounts are in Japanese yen (JPY). The implemented reading applies the 21,000-yen test to the coinsurance amount of the month as a whole, which corresponds to aggregating the dispensed portion with the prescriber’s charges; it is the more permissive of the two readings (Methods, Study

26

background). The upper bound is constructed by attributing to the failing streams the largest share the absence of an apportionment rule permits, not by proportional apportionment. Two profiles that dispense orally, the endocrine-monotherapy profile P5 and the ovarian-suppression profile P12, contribute no counted months at all in any category, because their monthly coinsurance amounts do not reach the floor. Source: the archived analysis files, reproduced as the archived analysis files (18 rows).

Table A5. Record of the superseded P8 implementation*

P8 was rebuilt during the analysis to match the specified setting (Methods, Study background, and Table A19, entry 1). The implementation in force before that date carried the 14-cycle trastuzumab emtansine schedule without the neoadjuvant and surgical components and applied the metastatic follow-up schedule, and so did not match the specification. Its series are retained rather than deleted.

Quantity	Superseded implementation	Rebuilt profile
Series retained	810 series × 12 calendar start months = 9,720 rows	1,620 series per start month
Median annual total medical expenditure (JPY)	6,871,518	—
Annual-cap attainment rate (series instances, 12-month frame)	0.2984	0.3121
Composition	trastuzumab emtansine alone; no surgery, no neoadjuvant therapy; metastatic follow-up	neoadjuvant TCHP, breast-conserving surgery with a 3-day stay, trastuzumab emtansine, adjuvant-type surveillance

* The superseded values show which figures changed against the earlier record and why; they are not an alternative estimate and no judgment in the paper rests on them. The two attainment rates are computed on the same 12-month frame and the same year alignment, so the difference is attributable to the change in composition alone. — the rebuilt profile’s value was not tabulated. Source: the archived analysis files (9,720 rows); the archived analysis files.

Table A6. The 15 treatment profiles*

Profile	Clinical setting	Principal treatment components
P1	Early invasive, surgery only, radiotherapy omitted	Breast-conserving surgery with sentinel node biopsy; no cytotoxic chemotherapy; **no endocrine therapy costed** (see Table A19, entry 3)
P2	Ductal carcinoma in situ	Breast-conserving surgery with sentinel node biopsy; whole-breast irradiation; **no endocrine therapy costed** (see Table A19, entry 3)
P3	Early, surgery and adjuvant chemotherapy	Docetaxel 75 with cyclophosphamide 600 mg/m² every 3 weeks for 4 cycles; no radiotherapy
P4	Early, surgery, chemotherapy, then radiotherapy	As P3, followed by whole-breast irradiation 2 to 4 weeks after chemotherapy
P5	Endocrine maintenance alone	Tamoxifen 20 mg daily; scheduled visits, blood tests, and mammography only

27

Profile	Clinical setting	Principal treatment components
P6	HER2-positive adjuvant, intravenous anti-HER2	Doxorubicin with cyclophosphamide for 4 cycles, then taxane with intravenous trastuzumab and pertuzumab, three-weekly, truncated at the 12-month horizon; the count of months attributed to the maintenance phase depends on the start day (section A.6)
P7	HER2-positive adjuvant, subcutaneous anti-HER2	As P6 with the pertuzumab and trastuzumab subcutaneous fixed-dose combination, truncated at the 12-month horizon; same start-day dependence as P6
P8	HER2-positive, residual disease after neoadjuvant therapy (see footnote)	Neoadjuvant docetaxel, carboplatin, trastuzumab, and pertuzumab (TCHP) for 6 three-weekly cycles; breast-conserving surgery with sentinel node biopsy, 3-day stay; trastuzumab emtansine 3.6 mg/kg every 3 weeks for 14 cycles; adjuvant-type surveillance with annual mammography and no response CT or tumour markers
P9	HER2-positive metastatic or recurrent	Trastuzumab deruxtecan 5.4 mg/kg every 3 weeks; response CT, chest CT for interstitial lung disease surveillance, tumour markers
P10	Triple-negative, neoadjuvant chemotherapy with pembrolizumab	Paclitaxel with carboplatin, then epirubicin with cyclophosphamide; pembrolizumab 200 mg every 3 weeks; surgery in month 7; truncated at the 12-month horizon
P11	Bone metastasis	Anastrozole 1 mg daily with denosumab 120 mg subcutaneously every 4 weeks; response CT, tumour markers
P12	Premenopausal, endocrine therapy with ovarian suppression	Tamoxifen 20 mg daily with leuprorelin 3.75 mg every 28 days, 13 administrations (a monthly, not a 6-month, depot; see Table A19, entry 5)
P13	Adjuvant CDK4/6 inhibition, high-risk hormone receptor positive	Adjuvant endocrine therapy with abemaciclib 150 mg twice daily
P14a	Metastatic or recurrent, CDK4/6 inhibition on an aromatase inhibitor	Aromatase inhibitor with a CDK4/6 inhibitor; response CT, visits, blood tests
P14b	Metastatic or recurrent, CDK4/6 inhibition on fulvestrant	Fulvestrant with a CDK4/6 inhibitor, loading schedule in the first month; response CT, visits, blood tests
P15	Total mastectomy	Total mastectomy with sentinel node biopsy, 6-day stay; radiotherapy not given as a rule; endocrine therapy continued if oestrogen receptor positive

[*] Doses are as specified in the clinical review and are reproduced in full in Table A1 together with administration intervals, cycle numbers, supportive care, scheduled tests, and the observation period for each profile. Numbers such as “18 administrations” and “14 cycles” are the planned totals under complete delivery; dose reduction, delay, and discontinuation are not modelled. P14 comprises two subprofiles distinguished by the endocrine backbone. P8 was rebuilt during the analysis, after the clinical review confirmed its composition, to match the specified setting as shown; the profile

28

computed before that date contained the 14-cycle trastuzumab emtansine schedule without the surgical and neoadjuvant components and with the metastatic follow-up schedule, did not match the specification, and its values are retained in Table A5 as the record of an implementation that failed to match the specification. The clinical review listed two neoadjuvant regimens without stating a priority; taking TCHP as the base is a design assumption, fixed before any result was computed, the alternative (doxorubicin and cyclophosphamide followed by a taxane with trastuzumab and pertuzumab) is carried as a sensitivity arm, and the priority is under clinical confirmation (section A.9). Abbreviations: CDK4/6, cyclin-dependent kinase 4 and 6; CT, computed tomography; HER2, human epidermal growth factor receptor 2. Sources: the finalised specification of the treatment profiles and the reviewer consolidation table, as adjudicated on 10 August 2026.

Table A7. Monthly cap, repeat-month cap, and annual cap for enrollees under 70, before and after the reform effective 1 August 2026*

Category	Standard remuneration	Before: monthly cap	Before: repeat-month cap	Before: annual cap	Before: between-caps-range width	Before: largest monthly difference	From 1 August 2026: monthly cap	From 1 August 2026: repeat-month cap	From 1 August 2026: annual cap	From 1 August 2026: between-caps-range width	From 1 August 2026: largest monthly difference
A	≥ 830,000	252,600 + (M − 842,000) × 1%	140,100	—	375,000	112,500	270,300 + (M − 901,000) × 1%	140,100	1,680,000	434,000	130,200
B	530,000 to 829,999	167,400 + (M − 558,000) × 1%	93,000	—	248,000	74,400	179,100 + (M − 597,000) × 1%	93,000	1,110,000	287,000	86,100
C	280,000 to 529,999	80,100 + (M − 267,000) × 1%	44,400	—	119,000	35,700	85,800 + (M − 286,000) × 1%	44,400	530,000	138,000	41,400
D	< 280,000	57,600	44,400	—	44,000	13,200	61,500	44,400	530,000	57,000	17,100
E	municipal-tax-exempt	35,400	24,600	—	36,000	10,800	36,900	24,600	290,000	41,000	12,300

* All amounts are in Japanese yen (JPY). “Standard remuneration” is standard monthly remuneration, and “before” and “after” are the values in force before and after the reform effective 1 August 2026. M denotes total medical expenditure for the month at full tariff. Categories A to E correspond to the statutory categories ア, イ, ウ, エ, and オ respectively; the Japanese labels are not used elsewhere in this paper. Categories D and E have no incremental term and their cap is the fixed component alone. Out-of-pocket payment is obtained both before and after the reform by applying the cap to 30% of total medical expenditure. Fractions of a yen in the incremental term are rounded down below 50 sen and up at 50 sen or above (Enforcement Order, Article 42, paragraph 1). The repeat-month cap applies from the fourth qualifying month within the rolling 12-month window ending with the current month (same Article, paragraph 2 and related provisions). No annual cap existed before the reform (“—”). The between-caps-range width is (fixed component − repeat-month cap) ÷ 0.3, the length of the interval of monthly total medical expenditure defined in section A.6.1, and the largest

29

monthly difference is fixed component − repeat-month cap, the most by which a patient inside the interval can pay more in a month than a patient on the repeat-month cap; the reform widened the interval in all five categories. Annual cap amounts are set by the items of Article 42, paragraph 11, and the conditions for payment by Article 41-3; Categories C and D both fall within item 1, the residual provision, and share the same amount. Source: Cabinet Order Partially Amending the Enforcement Order of the Health Insurance Act and Related Orders (Cabinet Order No. 240 of 2026), Official Gazette Extra No. 169, 29 July 2026. Parameters for the thirteen categories effective 1 August 2027 (Categories 1 to 13, Article 2 of the same Order) are given in Table A8; they are not placed as a third group of columns here because the thirteen categories do not correspond one to one with Categories A to E. Source of the two interval columns: results/rerun/E3_reform_delta.csv.

Table A8. Monthly cap, repeat-month cap, and annual cap for the thirteen income categories effective 1 August 2027*

Income category	Fixed component	Reference amount	Repeat-month cap	Annual cap
Category 1	342,000	1,140,000	140,100	1,680,000
Category 2	303,000	1,010,000	140,100	1,680,000
Category 3	270,300	901,000	140,100	1,680,000
Category 4	209,400	698,000	93,000	1,110,000
Category 5	194,400	648,000	93,000	1,110,000
Category 6	179,100	597,000	93,000	1,110,000
Category 7	110,400	368,000	44,400	530,000
Category 8	98,100	327,000	44,400	530,000
Category 9	85,800	286,000	44,400	530,000
Category 10	69,600	—	44,400	530,000
Category 11	65,400	—	44,400	530,000
Category 12	61,500	—	34,500	410,000
Category 13	36,900	—	24,600	290,000

* All amounts are in Japanese yen (JPY). The monthly cap is the fixed component plus 1% of total medical expenditure above the reference amount for Categories 1 to 9, and the fixed component alone for Categories 10 to 13, which have no incremental term (“—”). The reference amount equals the fixed component divided by 0.3 in every category that has one. Annual cap amounts are set by the five items of Article 42, paragraph 11 of the Enforcement Order and are assigned to monthly categories by containment of the standard monthly remuneration range, as described in section A.1.6; five categories share item 1 at 530,000 yen. Source: Cabinet Order No. 240 of 2026, Article 2 (Official Gazette Extra No. 169, 29 July 2026), via the archived analysis files and scheme_params_2027_annual.csv; assembled as the archived analysis files.

Table A9. Units and denominators of the reported quantities*

Quantity	Unit	Composition of the denominator	Denominator	Reported in	Source file
Primary outcome: proportion of	treatment-month	months attributed to the anti-HER2 maintenance	7,200 per profile at a day-1	Results 3.2; Table	phaseH/

30

Quantity	Unit	Composition of the denominator	Denominator	Reported in	Source file
months inside the between-caps range (designated combination)		phase, summed over every series of the base stratum; 2026 Category C; August start	start; 6,480 at a day-15 start	A10	FW4_b_e4_combination.csv
Secondary outcome: proportion of brand-to-generic pairs with a higher annual payment in the generic arm	pair	one pair per main-path series, per income category and start day; August start; 28 start days in the sweep; generic arm total expenditure lower in 50,220, equal in 1,653, higher in 0	51,873 per income category per start day	Results 3.4; Table 2	phaseH/ FW1_e5_pair_totals.csv; phaseH/FJ3B_e5_summary.csv
Annual-cap attainment on the main path: share of years reaching the cap	course-year	51,873 series × 18 income categories × 23 years over the 12 calendar start months (1 year for an August start, 2 for each other start month); day-1 start; brand arm	21,475,422 (reaching: 1,423,481)	Results 3.1 and 3.5; Table 1; Fig. A2; Table A21	phaseH/ FJ3D_b_annual_cap_activity.csv
Profile-by-category combinations with any attainment	combination	12 main-path profiles × 18 income categories	216 (with any attainment: 72)	Results 3.3; Fig. 2	phaseH/ FJ3D_b_annual_cap_activity.csv
Change in out-of-pocket payment over the course, post-reform minus pre-reform, main path: median, minimum, and maximum, and the direction of the change relative to the benchmark increase	course start	series of the profile × 24 start dates (12 calendar start months × the 1st and the 15th), per 2026 income category; brand arm; 12 main-path profiles × 5 income categories	from 72 (P5, P11, P12) to 311,040 (P6, P7) per income category; 60 profile-by-category combinations	Results 3.1; Discussion; Table 1	reform_gap/ Table1_candidate.csv; reform_gap/T3_cells.csv; reform_gap/T4_decomposition.csv; reform_gap/T7_direction_counts.csv
CDK4/6 profiles: annual-cap attainment and sign of the difference over the 24-month course	combination	18 income categories × 12 calendar start months × 2 start days, per profile; base switch point	432 per profile (P13, P14a, P14b); 1,296 in total	Results 3.5; Fig. A2; Table A20	phaseH/ FB2b_switch_point_sensitivity.csv
CDK4/6 medians in one income category	combination of start month and start day	12 calendar start months × 2 start days	24	Results 3.5; Table A20	phaseH/ FW7_p13_median_domain.csv
Rolling-window counterfactual: share of years and of courses reaching the annual cap	course-year; course	years: under the rolling window, the windows a course opens, a course that never qualifies counted as one year; courses: main-path series × income categories × 12 calendar start months; day-1 start; brand arm	11,204,568 years (fixed year: 21,475,422); 11,204,568 courses	Results 3.5; Discussion; Table A21	counterfactual/ CF5_summary.csv

[*] The table gives, for every quantity reported in the main text, the unit that is counted and the denominator over which it is formed. None of the denominators is a count of patients: a series is a constructed monthly cost sequence for one profile at one level of each stratum, a treatment-month is one month of one series, a pair is one brand series and its generic counterpart, a course-year is one August–July year of one series, a course start is one series begun on one of the 24 start dates, a combination is one combination of profile, income category, start month, and start day, and a course is one series at one start month under one income category. Counts are read from the source files named.

31

Table A10. The designated combination of the primary outcome, 2026 Category C, base stratum*

Profile	Phase	Day-1 proportion	Day-15 proportion	Median distance to the nearer interval endpoint, day 1 (JPY)	Same, day 15 (JPY)
P6	anti-HER2 maintenance	0.600	0.667	23,954	21,134
P7	anti-HER2 maintenance	0.600	0.667	23,972	21,152

* Proportions are pooled over treatment-months as defined in section A.6. Every other profile-phase row in this category is exactly zero at both start days.

Table A11. Verification of the encoding of the law*

Check	Implementations compared	Grid	Compared	Disagreeing	Source file
Monthly cap for a given monthly total (fixed component and 1% addition)	shared engine; first-estimate model (no rounding); compared without rounding	18 income categories × 151 monthly totals from 0 to 1,500,000 yen in steps of 10,000	2,718	0	results/phaseG/ G7_2_three_way_crosscheck.csv (list of disagreements); code/phaseG/g7_2.py
Months in which the repeat-month cap applies, consecutive qualifying months	shared engine; first-estimate model; earlier structural schedule	18 income categories × 0 to 24 preceding qualifying months	450	0	results/phaseG/ G7_2_three_way_crosscheck.csv (list of disagreements); code/phaseG/g7_2.py
Months in which the repeat-month cap applies, qualifying months with gaps	shared engine (rolling 12-month window, as in the law); first-estimate model (cumulative count)	108 patterns of qualifying months	108	54**	results/phaseG/ G7_2b_year_discrimination.csv
Between-caps-range endpoints one yen below, at, and one yen above each boundary	floating-point arithmetic; decimal arithmetic	18 income categories × 2 endpoints × 3 points	108	0	results/phaseG/ G7_4_rounding_flip.csv
Statutory parameters: monthly cap, repeat-month cap, and annual cap, 2026 and 2027 (ledger V1-1 to V1-11)	re-derived from the Official Gazette text independently of the parameter files	11 statements	11	0	docs/claims_ledger.md section 1
Smallest number of capped months at which the annual cap is reached (ledger V2-1)	hand calculation; independent implementation	18 income categories	18	0	docs/claims_ledger.md section 2
Worked example of Table A12: every month of one course at two start months	hand calculation in decimal arithmetic; shared engine	26 rows (months and course totals)	26	0	results/counterfactual/ WE2_worked_example_trace.csv

* The checks compare the analysis engine with implementations written separately from it. The first-estimate model does not apply the rounding of the 1% addition, so the monthly cap was compared with the engine’s rounding switched off. The earlier structural schedule returns only the sequence of caps for a patient whose monthly payment is always capped, so it enters the comparison of repeat-month months and not that of the monthly cap. Rerunning the registered pipeline, which shows determinism rather than correctness, is reported in section A.8.

32

[**] The first-estimate model counts qualifying months cumulatively over the course, whereas the law and the engine count them in the rolling 12-month window ending with the current month; the disagreements occur only in patterns whose qualifying months are separated by gaps and are therefore differences of design, not errors of either implementation in its own terms.

Table A12. Worked example: one course of the metastatic antibody-drug conjugate profile in 2027 Category 1, started in August and in March[*]

Start	Treatment month	Calendar month	Annual-cap year	Total medical expenditure	Coinsurance (30%)	Ordinary monthly cap	Applicable cap	Repeat-month cap applied	Out-of-pocket payment	Counted toward the annual cap	Cumulative in the year
August	1	2026-08	1 Aug 2026–31 Jul 2027	1,615,744	484,723	346,757	346,757	no	346,757	346,757	346,757
August	2	2026-09	1 Aug 2026–31 Jul 2027	792,342	237,703	342,000	342,000	no	237,703	237,703	584,460
August	3	2026-10	1 Aug 2026–31 Jul 2027	1,615,744	484,723	346,757	346,757	no	346,757	346,757	931,217
August	4	2026-11	1 Aug 2026–31 Jul 2027	792,342	237,703	342,000	342,000	no	237,703	237,703	1,168,919
August	5	2026-12	1 Aug 2026–31 Jul 2027	1,592,624	477,787	346,526	346,526	no	346,526	346,526	1,515,445
August	6	2027-01	1 Aug 2026–31 Jul 2027	819,432	245,830	342,000	140,100	yes	140,100	140,100	1,655,545
August	7	2027-02	1 Aug 2026–31 Jul 2027	1,584,503	475,351	346,445	140,100	yes	140,100	140,100	1,795,645
August	8	2027-03	1 Aug 2026–31 Jul 2027	796,493	238,948	342,000	140,100	yes	140,100	140,100	1,935,745
August	9	2027-04	1 Aug 2026–31 Jul 2027	796,312	238,894	342,000	140,100	yes	140,100	140,100	2,075,845
August	10	2027-05	1 Aug 2026–31 Jul 2027	1,611,774	483,532	346,718	140,100	yes	140,100	140,100	2,215,945
August	11	2027-06	1 Aug 2026–31 Jul 2027	796,312	238,894	342,000	140,100	yes	140,100	140,100	2,356,045
August	12	2027-07	1 Aug 2026–31 Jul 2027	1,588,654	476,596	346,487	140,100	yes	140,100	140,100	2,496,145
August	Course total	—	—	14,402,272	—	—	—	—	2,496,145	—	—
August	Refund under the annual cap	—	—	—	—	—	—	—	816,145	—	—
August	Net out-of-pocket over	—	—	—	—	—	—	—	1,680,000	—	—

33

Start	Treatment month	Calendar month	Annual-cap year	Total medical expenditure	Coinsurance (30%)	Ordinary monthly cap	Applicable cap	Repeat-month cap applied	Out-of-pocket payment	Counted toward the annual cap	Cumulative in the year
	the course										
March	1	2027-03	1 Aug 2026–31 Jul 2027	1,615,744	484,723	346,757	346,757	no	346,757	346,757	346,757
March	2	2027-04	1 Aug 2026–31 Jul 2027	792,342	237,703	342,000	342,000	no	237,703	237,703	584,460
March	3	2027-05	1 Aug 2026–31 Jul 2027	1,615,744	484,723	346,757	346,757	no	346,757	346,757	931,217
March	4	2027-06	1 Aug 2026–31 Jul 2027	792,342	237,703	342,000	342,000	no	237,703	237,703	1,168,919
March	5	2027-07	1 Aug 2026–31 Jul 2027	1,592,624	477,787	346,526	346,526	no	346,526	346,526	1,515,445
March	6	2027-08	1 Aug 2027–31 Jul 2028	819,432	245,830	342,000	140,100	yes	140,100	140,100	140,100
March	7	2027-09	1 Aug 2027–31 Jul 2028	1,588,473	476,542	346,485	140,100	yes	140,100	140,100	280,200
March	8	2027-10	1 Aug 2027–31 Jul 2028	796,493	238,948	342,000	140,100	yes	140,100	140,100	420,300
March	9	2027-11	1 Aug 2027–31 Jul 2028	819,432	245,830	342,000	140,100	yes	140,100	140,100	560,400
March	10	2027-12	1 Aug 2027–31 Jul 2028	1,588,654	476,596	346,487	140,100	yes	140,100	140,100	700,500
March	11	2028-01	1 Aug 2027–31 Jul 2028	796,312	238,894	342,000	140,100	yes	140,100	140,100	840,600
March	12	2028-02	1 Aug 2027–31 Jul 2028	1,584,684	475,405	346,447	140,100	yes	140,100	140,100	980,700
March	Course total	—	—	14,402,272	—	—	—	—	2,496,145	—	—
March	Refund under the annual cap	—	—	—	—	—	—	—	0	—	—
March	Net out-of-pocket over the course	—	—	—	—	—	—	—	2,496,145	—	—

[*] All amounts are in Japanese yen (JPY), rounded to the nearest yen for display; the calculation carries fractions of a yen. The course is the series selected by a rule fixed before the example was drawn: among all main-path series and income categories at a day-1 start in the brand arm, the one whose out-of-pocket payment over the course differs most between an August and a March start. Its total medical expenditure is the same at both start

34

months. Every row was computed by hand in decimal arithmetic and agreed with the analysis engine (Table A11). Parameters are those of 2027 Category 1 in Table A8, applied to every month as throughout the analysis: the ordinary monthly cap is the fixed component, 342,000, plus 1% of total medical expenditure above the reference amount, 1,140,000, rounded as in section A.1.2; the applicable cap is the repeat-month cap, 140,100, from the fourth qualifying month within the rolling 12-month window; a month counts toward the annual cap only when its coinsurance amount is 21,000 or more; and the annual cap is 1,680,000. A qualifying month is one in which the coinsurance amount exceeds the applicable cap. "—" denotes a combination that does not apply to a total row. Source: results/counterfactual/WE2_worked_example_trace.csv; results/counterfactual/WE1_selection_candidates.csv.

Table A13. Inpatient-stay mapping: full sweep*

Every series built by the principal generator was recomputed with the length of stay set by the registered mapping (3 days for the two breast-conserving procedures, 6 for the two mastectomy procedures) in place of the uniform 5 days carried by the implementation. **No judgment changes in any of the 216 profile–category combinations**: annual-cap attainment (any or none), between-caps-range membership (any or none), and the modal sign of the annual difference are identical. Counts move in 18 combinations, all of them in P6, P7, and P10.

Profile	Income category	Quantity that moves	Uniform 5 days	Registered mapping
P6	2026 Cat. B	attainment count	14,828	14,394
P6	2027 Cat. 1	months inside the interval	739,748	735,968
P6	2027 Cat. 2	months inside the interval	567,540	563,760
P6	2027 Cat. 4	attainment count	25,661	25,271
P6	2027 Cat. 5	attainment count	23,914	23,539
P6	2027 Cat. 6	attainment count	14,828	14,394
P6	2027 Cat. 7	attainment count	90,015	89,973
P6	2027 Cat. 8	attainment count	30,946	30,731
P7	2026 Cat. B	attainment count	16,855	16,515
P7	2027 Cat. 1	months inside the interval	761,780	758,000
P7	2027 Cat. 2	months inside the interval	567,540	563,760
P7	2027 Cat. 4	attainment count	24,951	24,380
P7	2027 Cat. 5	attainment count	24,040	23,436
P7	2027 Cat. 6	attainment count	16,855	16,515
P7	2027 Cat. 7	attainment count	90,713	90,701
P7	2027 Cat. 8	attainment count	31,171	30,978
P10	2027 Cat. 1	attainment count	43,896	43,890
P10	2027 Cat. 2	attainment count	42,574	42,573

35

[*] Counts are series instances pooled over all series and twelve calendar start months at a day-1 start in the brand arm. The largest movement is a decrease of 604 attainment instances (P7, 2027 Category 5) and of 3,780 interval months (P6, 2027 Category 1). Movements are almost all decreases, because the mapping removes two inpatient days from the conserving-surgery admissions that dominate these profiles. The rebuilt P8 already uses the mapping and is unchanged by construction; P15 uses 6 days and is likewise unchanged. Source: the archived analysis files (216 rows) and FK2_admit_changed_combinations.csv (18 rows).

Table A14. The body-size band[*]

Level	Body surface area, m²	Body weight, kg	Derivation
low	1.3465	45.1	mean − 1 SD of the 60–69 year age class
mid	1.5161	54.07	simple mean of the age-class means
high	1.6898	64.0	mean + 1 SD of the 40–49 year age class

[*] Body surface area is in square metres and body weight in kilograms; both are computed from the same height-and-weight statistics and are reported to the number of decimal places carried in the parameter file. The three levels are a band that brackets the plausible range, not a distribution: no weighting is applied across them, no population frequency is implied, and results at the outer levels are read as an envelope. Body surface area and body weight are not varied independently, because both are derived from the same height-and-weight statistics; the three levels are one body-size stratum. Carboplatin is fixed at 450 mg as the equivalent of AUC 5 and is outside the reach of this stratum (Methods, Study background). Source: National Health and Nutrition Survey 2019 (statsDataId 0003224177), via the archived analysis files and the archived analysis files.

Table A15. Retirement of the G-CSF axis[*]

Primary prophylaxis with granulocyte colony-stimulating factor was initially carried as a two-level stratum: threshold-based counting, in which the agent is costed only for regimens whose febrile neutropenia risk exceeds 20%, the threshold recommended by the American Society of Clinical Oncology and used in Methods, Variables, and a real-practice level in which it is costed for the docetaxel-and-cyclophosphamide regimen as well. The axis is not swept in the analysis reported here.

Item	Record
Status	Retired; the base case (threshold-based counting at 20%) is the only level computed
When the decision was taken	After results had been seen; recorded as an amendment with that fact stated
Recorded rationale	The chemotherapy months of every affected profile already lie at or above the upper endpoint of the between-caps range, so adding the agent cannot move their interval position
Profiles that would have carried the axis	P3, P4, P6, P7, P8, P9, P10
Consequence for the reported design	The registered series-count expectation was computed with the axis included; the counts in Table A3 are those without it
Where restated	The limitations in the Discussion

36

[*] The entry records that the decision was taken after results were seen, rather than leaving it to be inferred from a design that would otherwise appear smaller from the outset. Source: the archived analysis files, section DE-2, which records the amendment and its date; the archived analysis files for the margins relied on.

Table A16. CDK4/6 switch point: full comparison and the eight sign flips[*]

The dispensing interval switches from 14 days to 28 days at a point with no external anchor. Three levels were computed, 1 cycle (28 days), 2 cycles (56 days, the base), and 3 cycles (84 days), over 1,296 combinations (3 profiles × 18 income categories × 12 calendar start months × 2 start days).

Judgment	Combinations in which the three levels differ
Between-caps-range membership	0 of 1,296
Annual-cap attainment	0 of 1,296
Sign of the reform difference	**8 of 1,296 (0.62%)**
Out-of-pocket total over the 24-month course	568 of 1,296 (43.8%), largest difference 557,012

The eight combinations in which the sign differs, with the out-of-pocket total over the 24-month course at each switch point:

Profile	Calendar start month	Start day	Income category	Sign at 28 / 56 / 84 days	Out-of-pocket total over the course at 28 / 56 / 84 days
P13	February	1	2027 Cat. 4	− / − / +	2,824,822 / 2,824,841 / 2,883,527
P13	March	1	2027 Cat. 4	− / + / +	2,764,076 / 2,825,386 / 2,825,950
P14a	February	1	2027 Cat. 4	− / − / +	2,790,133 / 2,790,152 / 2,841,841
P14a	March	1	2027 Cat. 4	− / + / +	2,734,342 / 2,790,697 / 2,791,261
P14b	February	1	2026 Cat. B	− / − / +	3,044,923 / 3,044,942 / 2,487,930
P14b	February	1	2027 Cat. 4	− / − / +	3,090,123 / 3,090,142 / 2,733,357
P14b	February	1	2027 Cat. 5	− / − / +	3,075,623 / 3,075,642 / 2,704,357
P14b	February	1	2027 Cat. 6	− / − / +	3,044,923 / 3,044,942 / 2,487,930

[*] All amounts are in Japanese yen (JPY). The amount columns are the out-of-pocket total over the profile’s whole 24-month observation period, net of annual-cap refunds in each year the course spans, not a single year’s payment; the three profiles span two to three August–July years depending on the

37

calendar start month. A minus sign is a decrease in what the patient pays after the reform. Every one of the eight is a day-1 start in calendar February or March; no day-15 start and no other calendar month produces a flip. The switch point is a design assumption with no clinical source: the product labelling of neither agent states a monitoring interval of the kind the induction period would correspond to, which was checked verbatim and recorded. The base of 2 cycles was chosen on cycle rather than calendar grounds, so that the induction length would not be confounded with the start-day stratum. Source: the archived analysis files (recomputed for this table as FW6_supp_numbers.csv); the archived analysis files.

Table A17. (A0), (A), and (B): what each change moved, and the two corrections*

Three states of the computation are distinguished so that one comparison carries one cause. **(A0)** is the computation as first performed. **(A)** differs from it in one respect only: the August–July years are aligned to the statutory 1 August to 31 July calendar instead of being cut from the treatment start month. **(B)** differs from (A) by the clinical-review answers of 13 August 2026, namely the dexamethasone schedules, the rebuilt P8, the CDK4/6 dispensing intervals, and the confirmed scope of the outpatient chemotherapy management fee. The paper reports (B) throughout.

S11a. (A0) compared with (A): the year alignment alone

Judgment	(A0), years cut from the start month	(A), years on the statutory calendar
Annual-cap attainment, total instances	4,536,812	1,365,767
Attainment measure reported	41.13% of series instances	6.4606% of years (1,365,767 / 21,140,082)
P6 / P7 / P8 / P9 / P10 attainment instances	1,195,760 / 1,218,736 / 126,360 / 596,276 / 1,399,680	227,147 / 231,664 / 52,208 / 251,422 / 603,326
Modal sign of the annual difference	—	reversed from negative to positive in 31 profile–category combinations (P8 12, P9 4, P10 15)
Annual-cap years spanned by a 24-month course	2	3
Between-caps-range month counts	identical (digest match in all 18 categories)	identical
Secondary outcome (E5)	identical; (A) equals (A0), verified in all 18 categories	identical

S11b. (A) compared with (B): the clinical-review answers alone

Judgment	(A)	(B)
Annual-cap attainment, share of years	6.4606% (1,365,767 / 21,140,082)	6.6284% (1,423,481 / 21,475,422)
Profile-by-category combinations with any attainment	67 of 216	72 of 216
Profiles with no attainment	7	7
Modal sign of the annual difference	—	unchanged in all 216 combinations
Between-caps-range month totals	—	move in 40 of 216 combinations
Secondary outcome: designated categories	8	8, the same set
Series per arm at one start day	51,063	51,873

S11c. The two corrections

38

Correction	What was wrong	Where it was found	Effect
Year alignment	The cap calculation received a cost series that began at the treatment start month, so the 12-month years did not coincide with the statutory 1 August to 31 July period. Attainment was systematically overcounted.	While preparing a boundary argument, when a start-month index was found not to mean the calendar month it had been read as	(A0) → (A), quantified in S11a
Interim transcription	An interim log recorded attainment figures from a run made before the alignment fix, while the result files had been regenerated after it. The stale figures were 1,366,179 years, 6.4625%, and 31 sign reversals stated as 33.	When an anchor value recomputed during the (B) work failed to reproduce the logged figure	Correct values are 1,365,767 years, 6.4606%, and 31 reversals. No result file was affected; the manuscript reports the values read from the result files.

[*] All counts are series instances or years as labelled, never patients. The denominators differ between (A) and (B) because the rebuilt P8 carries more series than the profile it replaced, so attainment shares rather than counts are the comparable quantity. Source: the archived analysis files, FF2_A_vs_B.csv, FF2_summary_three_arms.csv, FJ3D_b_annual_cap_activity.csv, FI5_e4_invariance_digest.csv, FJ3B_e4_invariance_digest.csv.

Table A18. The primary outcome in all 18 income categories[*]

The designation of the primary outcome is defined on one combination, the 2026 Category C base stratum (Methods, Outcome). The proportion itself is reported here for every income category. Only the anti-HER2 maintenance phase of P6 and P7 takes a non-zero proportion in any category; the columns for those two profiles are therefore given in full, and the last column counts the profile-and-phase rows anywhere in the category that contain at least one month inside the interval.

Income category	P6 anti-HER2 maintenance	P7 anti-HER2 maintenance	Median nearest_edge, P6 / P7	Profile-phase rows with any month inside
2026 Category A	0.400	0.400	194,234 / 194,252	11
2026 Category B	0.400	0.400	47,954 / 47,972	12
2026 Category C	**0.600**	**0.600**	**23,954 / 23,972**	2
2026 Category D	0.000	0.000	67,766 / 67,748	0
2026 Category E	0.000	0.000	149,766 / 149,748	2
2027 Category 1	0.400	0.400	194,234 / 194,252	11
2027 Category 2	0.400	0.400	194,234 / 194,252	11
2027 Category 3	0.400	0.400	194,234 / 194,252	11
2027 Category 4	0.400	0.400	47,954 / 47,972	12
2027 Category 5	0.400	0.400	47,954 / 47,972	12
2027 Category 6	0.400	0.400	47,954 / 47,972	12
2027 Category 7	0.600	0.600	105,954 / 105,972	4
2027 Category 8	0.600	0.600	64,954 / 64,972	4

39

Income category	P6 anti-HER2 maintenance	P7 anti-HER2 maintenance	Median nearest_edge, P6 / P7	Profile-phase rows with any month inside
2027 Category 9	0.600	0.600	23,954 / 23,972	2
2027 Category 10	0.000	0.000	40,766 / 40,748	0
2027 Category 11	0.000	0.000	54,766 / 54,748	0
2027 Category 12	0.000	0.000	67,766 / 67,748	0
2027 Category 13	0.000	0.000	149,766 / 149,748	2

[*] Proportions are pooled over treatment-months in the base stratum (follow-up frequency at the middle level, body size at the middle level, calendar start month August, day-1 start), with 7,200 maintenance treatment-months in the denominator for P6 and P7 in every category. nearest_edge is the unsigned distance in yen to the nearer endpoint of the between-caps range, median over those treatment-months. The designation threshold of 0.50 is met in four categories, 2026 Category C and 2027 Categories 7, 8, and 9, all of which share the repeat-month cap of 44,400 yen; the designation reported in the main text is the one defined on 2026 Category C. A category can have zero maintenance months inside the interval and still have other profile-phase rows with months inside, which is why the last column is not zero wherever the P6 and P7 columns are. Source: the archived analysis files, derived from FW4_b_e4_combination.csv.

Table A19. Differences between the reviewed clinical specification and the costed content[*]

Entry 1. P8 scope

Reviewed — HER2-positive disease with residual tumour after neoadjuvant chemotherapy, with breast-conserving surgery and sentinel node biopsy

Costed — Resolved during the analysis: the profile was rebuilt as specified, with neoadjuvant TCHP for 6 cycles, breast-conserving surgery with sentinel node biopsy and a 3-day stay, trastuzumab emtansine for 14 cycles, and adjuvant-type surveillance

Direction — Formerly understated: the superseded implementation lacked the surgical and neoadjuvant components

Upper bound — No longer a discrepancy; the superseded implementation’s values are retained in Supplementary material as a record

Effect on judgments — All judgments reported in this paper use the rebuilt profile

Entry 2. Observation period

Reviewed — Fixed per profile at 12, 15, 16, 18, 24, or 30 months; the 16 months of P8 were derived from its treatment schedule and registered during the analysis

Costed — Fixed 12-month horizon for every profile on the main path, including the judgments on the rebuilt P8, whose full 16-month series is generated as a record; 24 months for P13, P14a, P14b and 12 months for P15 on the separate paths

40

Direction — Truncates P6 and P7 by 6 months, P9 by 18, P10 by 3, and P8 by 4; months 1 to 12 are unaffected in content, so the discrepancy is confined to the absence of months 13 onward

Upper bound — Magnitude requires regenerating the series over the registered horizons

Effect on judgments — Scope is settled and the effect is one-sided for the annual cap: adding months cannot reduce a year's cumulative, so annual-cap attainment is understated. The direction for the proportion of months inside the between-caps range is not determined, because adding months changes the denominator as well as the numerator

Entry 3. P1 and P2 endocrine therapy

Reviewed — Aromatase inhibitor or tamoxifen, continuous

Costed — No endocrine therapy costed

Direction — Understates

Upper bound — 11,555 per 90-day dispensing, being the aromatase inhibitor at 90 days plus the prescription fee and the outpatient base fee

Effect on judgments — 8 of 36 combinations change interval membership, 12 months in each, all in Categories A, 1, 2, and 3; **0 combinations change the set of qualifying months**

Entry 4. Scope of the brand-to-generic switch

Reviewed — Applied uniformly to all target drugs

Costed — Applied to the 12 injectable ingredients carried in the vial selection table; the oral endocrine agents, the LH-RH depot, and denosumab are priced by brand name and do not switch

Direction — Narrows the comparison

Upper bound — 7,803 decrease and 2,142 increase per 90-day dispensing for the aromatase inhibitor; 1,215 and 2,853 for tamoxifen; 6,517 decrease per administration for the LH-RH depot; 0 for denosumab, which has no generic listed

Effect on judgments — Settled by direct recomputation at the extreme listed generic prices rather than by the bound (section A.3.7). Across 5,184 profile-by-category-by-start combinations, **no combination changes the set of years reaching the annual cap**, and the set of qualifying months changes only in **30 combinations of P13**, all in the highest income category and all under the substitution that raises the price. **No profile on the main path is affected**

Entry 5. P12 LH-RH formulation

41

Reviewed — 6-month depot as the base case

Costed — Leuprorelin 3.75 mg, a monthly formulation, every 28 days for 13 administrations

Direction — Overstates: 188,257 per year in drug and administration cost

Upper bound — 39,318 increase in an administration month and 24,263 decrease in a non-administration month if the 6-month depot were substituted

Effect on judgments — 0 of 18 combinations monthly and 0 of 18 annually

Entry 6. Brand-row designation where two brand products share a strength

Reviewed — Not addressed by the review

Costed — Leuprorelin: the higher-priced of the two 3.75 mg brand rows. Denosumab: the lower-priced of the two 120 mg brand rows

Direction — Overstates by 1,889 per administration for leuprorelin; understates by 1,684 for denosumab

Upper bound — 3,778 and 3,368 per calendar month at two administrations; 24,557 and 21,892 per year

Effect on judgments — 0 of 18 combinations monthly and 0 of 18 annually for each

Entry 7. Inpatient stay

Reviewed — 3 days for the two breast-conserving procedures and 6 days for the two mastectomy procedures, as a mapping from the operation performed

Costed — 5 days uniformly for the profiles built by the principal generator; 3 days for the rebuilt P8 and 6 days for P15, which follow the mapping

Direction — Overstates conserving-surgery admissions by 2 days and understates mastectomy admissions by 1 day

Upper bound — 18,740 per inpatient day, that is 37,480 per conserving-surgery admission and 18,740 per mastectomy admission

Effect on judgments — Recomputing every series built by the principal generator under the mapping changes no judgment: annual-cap attainment (any or none), between-caps-range membership (any or none), and the modal sign of the reform difference are unchanged in all 216 profile–category combinations; counts alone move, in 18 combinations (Table A13)

* All amounts are in Japanese yen (JPY). “Reviewed” is the content specified at clinical review and “Costed” the content the series generator builds. Direction is stated relative to the specification: “overstates” means the costed content is higher than the specification would give. Upper bounds are constructed from the Drug Price List and the fee schedule alone and are deliberately wider than the selection rule of section A.3.1 permits; see that

42

section. “Effect on judgments” reports whether the bound is large enough to move a month across the upper endpoint of the between-caps range, which is the point at which the monthly cap begins to bind, or to move an annual cumulative across the annual cap. Combinations are counted over treatment profile by income category across all 18 categories. Entry 2 carries no bound for the truncated profiles and is restated in section A.4.4; entry 1 was resolved by the rebuild of P8; entry 7 rests on full recomputation rather than on a bound. Sources: the finalised specification and reviewer consolidation table; the series generator; EM1_margin_by_month.csv, EM3_flip_judgment.csv, EM4_literal_bounds.csv, and FK2_admit_mapping_sweep.csv.

This table is intended to be exhaustive as of the present analysis. It was assembled by comparing the specification against the generator profile by profile rather than by collecting incidental observations, and any further difference found later will be added to it rather than recorded in a footnote elsewhere. No further difference is known and the search was systematic; the absence of any further difference is not claimed. Staleness in the specification documents themselves, where a registered parameter has been superseded by a later decision, is a separate matter and is not listed here.

Table A20. The three CDK4/6 profiles over the 24-month course: annual-cap attainment and the difference, post-reform minus pre-reform*

Profile and dispensing schedule	Reaching the annual cap, n of 432 (%)	Difference negative, n of 432 (%)	Difference positive, n of 432 (%)	Category B, median (JPY)	Category C, reaching the annual cap, n of 24	Category C, median (JPY)	Category C, minimum (JPY)	Category C, maximum (JPY)
Adjuvant (P13), 14/28-day dispensing	432 (100.0%)	323 (74.8%)	109 (25.2%)	−690,914	24	13,730	−122,788	13,730
Metastatic on an aromatase inhibitor (P14a), 14/28-day dispensing	432 (100.0%)	323 (74.8%)	109 (25.2%)	−630,354	24	13,730	−122,377	13,730
Metastatic on fulvestrant (P14b), 14/28-day dispensing	432 (100.0%)	324 (75.0%)	108 (25.0%)	−1,062,289	24	12,921	−126,211	13,730
Adjuvant (P13), 90-day dispensing (superseded)	0 (0.0%)	24 (5.6%)	408 (94.4%)	33,930	0	16,530	16,530	16,530

* All amounts are in Japanese yen (JPY); a negative value is a decrease in what the patient pays. The difference is the out-of-pocket payment over the 24-month course, post-reform minus pre-reform. A combination is one income category at one calendar start month and one start day, so each profile has 432 combinations over the 18 income categories and a 2026 income category has 24. Reaching is attainment of the annual cap in at least one August–July year. No combination had a difference of exactly zero. The 14/28-day rows use the base induction-to-maintenance switch point (section A.7.5); the 90-day row is the superseded uniform 90-day dispensing interval, retained as an outlier scenario. Category B and Category C are the 2026 upper-middle and middle income categories. Medians are rounded to the nearest yen, halves away from zero. Source: results/phaseH/FB2b_switch_point_sensitivity.csv.

43

Table A21. Annual-cap attainment and start-month dependence under the fixed year and a rolling window*

Profile or group	Courses reaching the annual cap, fixed year	Same, rolling window	Years reaching the annual cap, fixed year	Same, rolling window	Combinations in which the start month sets the sign, fixed year	Same, rolling window
P1	0 (0.0%)	0 (0.0%)	—	—	0 of 216	0 of 216
P2	0 (0.0%)	0 (0.0%)	—	—	0 of 216	0 of 216
P3	0 (0.0%)	0 (0.0%)	—	—	0 of 116,640	0 of 116,640
P4	0 (0.0%)	0 (0.0%)	—	—	0 of 116,640	0 of 116,640
P5	0 (0.0%)	0 (0.0%)	—	—	0 of 54	0 of 54
P6	227,504 (8.1%)	986,977 (35.3%)	—	—	23,840 of 233,280	10,610 of 233,280
P7	231,465 (8.3%)	1,006,428 (36.0%)	—	—	19,288 of 233,280	7,943 of 233,280
P8	109,210 (31.2%)	272,316 (77.8%)	—	—	22,680 of 29,160	12 of 29,160
P9	251,010 (23.9%)	595,928 (56.8%)	—	—	39,040 of 87,480	160 of 87,480
P10	604,292 (43.2%)	1,244,280 (88.9%)	—	—	110,160 of 116,640	168 of 116,640
P11	0 (0.0%)	0 (0.0%)	—	—	0 of 54	0 of 54
P12	0 (0.0%)	0 (0.0%)	—	—	0 of 54	0 of 54
P13	216 (100.0%)	168 (77.8%)	—	—	8 of 18	1 of 18
P14a	216 (100.0%)	168 (77.8%)	—	—	8 of 18	1 of 18
P14b	216 (100.0%)	204 (94.4%)	—	—	7 of 18	0 of 18
Main path, all profiles, day-1 start	1,423,481 (12.7%)	4,105,929 (36.6%)	1,423,481 of 21,475,422	4,105,929 of 11,204,568	215,008 of 933,714	18,893 of 933,714
Main path, all profiles, day-15 start	1,324,233 (11.8%)	3,859,296 (34.4%)	1,324,233 of 21,475,422	3,859,296 of 11,204,568	217,916 of 933,714	0 of 933,714
CDK4/6, all three profiles, day-1 start	648 (100.0%)	540 (83.3%)	856 of 1,890	765 of 1,085	23 of 54	2 of 54
CDK4/6, all three profiles, day-15 start	648 (100.0%)	540 (83.3%)	868 of 1,890	918 of 1,053	21 of 54	0 of 54

* The fixed year is the statutory 1 August to 31 July calculation period. The rolling window is a counterfactual in which a year of twelve calendar months opens on the first day of a course's first qualifying month and the next year opens at the next qualifying month after it closes; every other rule, including the 21,000-yen floor, the repeat-month cap, and the timing of the income category, is unchanged. A course is one series at one start month under one income category; profile rows are at a day-1 start in the brand arm, and group rows give both start days. A year is counted where a course has costs: under the fixed year a 12-month course spans one year if it starts in August and two otherwise, and a 24-month course two or three; under the rolling window a course counts the years it opens, and a course that never qualifies counts as one year that does not reach the cap, so year shares are not comparable between the two definitions and courses are the comparable unit. A combination is one series in one income category at one start

44

day, and the start month sets the sign in that combination when the post-reform minus pre-reform difference is positive at some of the 12 start months and negative at others. Under the rolling window, courses of the adjuvant CDK4/6 profile and of the metastatic profile on an aromatase inhibitor in 2026 Category A and 2027 Categories 1 to 3, and of the fulvestrant-based profile in 2027 Category 1, open no year, because no month's coinsurance exceeds the monthly cap, so these courses lose the annual cap; this is why attainment falls for the CDK4/6 profiles. "—" denotes a year count reported only for the groups. The counterfactual is post hoc (section A.8). Source: results/counterfactual/CF1_profile_year_reach.csv, CF3_sign_start_month_dependence.csv, CF5_summary.csv.

45

Supplementary figures

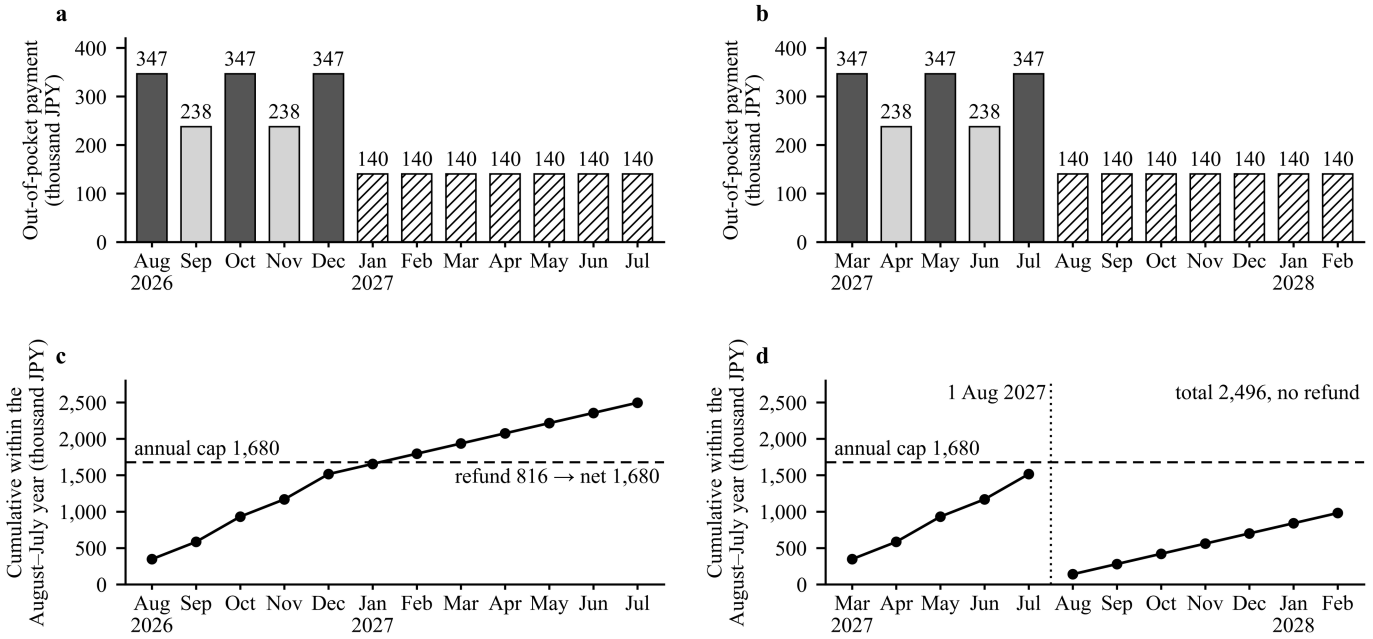
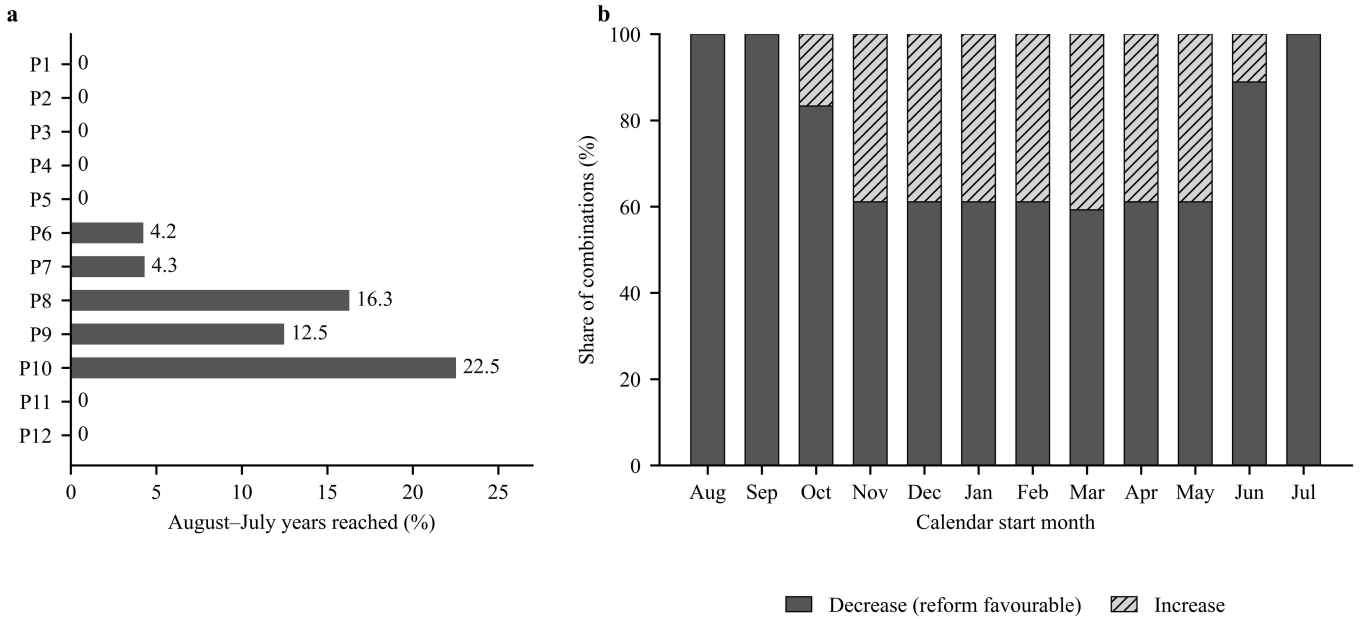


Fig. A1. The monthly cap, the repeat-month cap, and the annual cap over one course. The course is the one in Table A12: one series of the metastatic antibody-drug conjugate profile (P9) in 2027 Category 1, with the same total medical expenditure whether started in August (a, c) or in March (b, d). (a, b) Out-of-pocket payment in each month, in thousands of yen. Dark bars are months in which the coinsurance amount exceeds the ordinary monthly cap, so the patient pays the cap and the month qualifies. Light bars are months below the ordinary cap, in which the patient pays the full 30% coinsurance and the month does not qualify. Hatched bars are months paid at the repeat-month cap, which applies from the fourth qualifying month within the preceding 12 months. (c, d) Cumulative out-of-pocket payment within each August–July year, in thousands of yen. The dashed horizontal line is the annual cap of 1,680,000 yen. In (d) the dotted vertical line marks 1 August 2027, where the annual total restarts; the repeat-month count carries across that date. Started in August, the course passes the annual cap within one year, and the refund of 816,145 yen brings the net payment to 1,680,000 yen. Started in March, the same course is divided between two years, reaches the cap in neither, and the patient pays 2,496,145 yen. Every month passes the 21,000-yen floor. The parameters of 2027 Category 1 (Table A8) are applied to every month, as in Table A12.

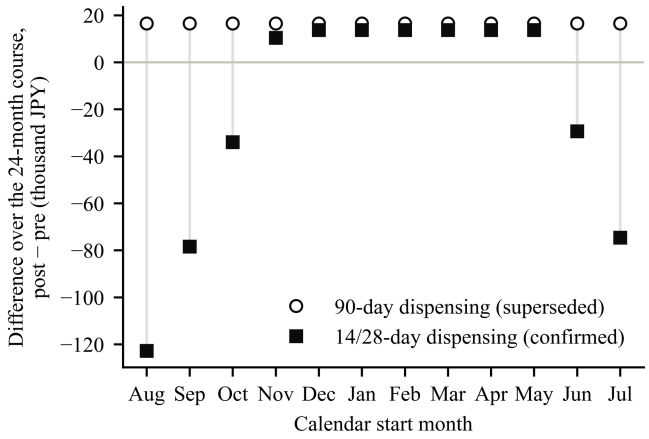
46



904

905 Fig. A2. Two regimes of annual-cap engagement. (a) Share of August–July years reached, by main-path profile; seven profiles never reach the cap. (b)
906 For the three CDK4/6 profiles, in which every combination reaches the cap, the sign of the difference in out-of-pocket payment over the 24-month
907 course by calendar start month; hatched segments are increases. No combination had a difference of exactly zero.

47



908

909 Fig. A3. The dispensing interval reverses the direction of the reform. Difference in out-of-pocket payment over the 24-month course for the adjuvant
910 CDK4/6 profile, 2026 Category C, day-1 start, under the superseded 90-day dispensing assumption and the confirmed 14/28-day schedule, by calendar
911 start month.

48