

How good are the Ministry of Justice prison population projections?

A walk-forward scorecard of eighteen editions, 2008 to 2025

Hamish Paton · Walkforward Research

Version 1.0, 4 September 2026 · Tag: Scored · ORCID 0009-0007-4160-9268 · Data and code: github.com/walkforwardresearch/prisoncap

Summary

- Since 2008 the Ministry of Justice has published eighteen editions of its prison population projections. This note scores every central projection against what happened, at every horizon from one to six years, using the MoJ's own weekly population bulletins as outturn.
- The central projection has been **too high at every horizon**. Mean error is +2.4% at one year, +3.8% at two and +5.2% at five. Thirteen of seventeen scoreable editions had every target too high. Fourteen of fourteen targets from editions published since 2020 were too high, by 7.6% on average. Removing the COVID years and the release-scheme years from the targets roughly halves the overshoot, to +2.3%, but does not remove it: the bias is structural, and the shocks doubled it.
- A forecaster who simply assumed the population would not change would have **beaten the MoJ at every horizon**, scored on identical targets. At two years the no-change error is 42% smaller, and the no-change forecast was closer on 54 of the 75 targets. Nobody outside government appears to run this comparison.
- The projections got the direction of change right 63% of the time overall and 53% at one and two years, barely better than a coin toss at the horizons that matter for capacity planning. They projected a rise in 73% of targets; the population rose in 52%.
- At one and two years, outturn fell below the MoJ's own low scenario 38% of the time. The published range covered the outcome in half of those targets.
- 95% of the error sits in the determinate sentenced population, the component set by release rules. Remand error runs the other way. Because the two partly cancel, gross component error is 1.9 times the net headline error, so the headline bias understates the modelling failure by about half.
- The MoJ can fairly say its projections are conditional on unchanged policy and that policy changed. That is true, and it is also the problem: a projection that is systematically falsified by the publisher's own subsequent actions is not a basis for sizing a £7 billion building programme.

What was scored, and how

The eighteen editions run from *Prison Population Projections 2008–2015 to 2025–2030*. Each publishes a central path at annual reference points; earlier editions publish it as one of three scenario columns, later editions as a single path broken down by custody type. The 2008 to 2019 editions sit as attachments on a single rolling gov.uk page; the six recent editions have their own pages. All were parsed into one table with scenario labels normalised to central, low and high.

Outturn is the weekly prison population bulletin, 787 bulletins from January 2011 to August 2026, parsed into one series and interpolated to each edition's published reference date. The reference date matters: it moves between end June, 1 June, end September, 1 July and end November across editions, and horizons are not comparable unless the published dates are used. Targets before January 2011 cannot be scored, so the 2008 and 2009 editions are scored from their second year.

Whether the projections are measured on the same basis as the bulletins was tested rather than assumed. Ten editions publish their own unrounded base-year actual. Those match the bulletin total, which includes HMPPS-operated immigration removal centres before 2022, more closely than the prisons-only figure: mean absolute gap 258 against 394. Since 2018 the agreement is within 40 people. The bulletin total is therefore used.

The benchmark is a forecaster with no model who, at publication, carries the current population forward unchanged. It is available at the same moment with the same information, and its errors are computed on the same targets. Where an edition publishes no base actual, the bulletin value at 30 June of the edition year is used. For the 2008, 2009 and 2010 editions, whose base years predate the weekly series, the 30 June prison population is taken from the Population in Custody bulletins for June 2009 and June 2010 (83,194, 83,391 and 85,002), so every one of the 75 targets has a like-for-like benchmark.

Horizon is the calendar year of the reference date less the edition year, so a 2020 edition's September 2021 target is one year ahead. Error is projected minus actual; a positive error means the projection was too high.

Results

By horizon

HORIZON	MEAN ERROR	MEAN ABS. % ERROR	SHARE TOO HIGH	MOJ MAE	NO-CHANGE MAE	MOJ / NO- CHANGE	NO-CHANGE CLOSER
1 year	+2.4%	2.9%	87%	2,429	1,666	1.46	73%
2 years	+3.8%	4.7%	87%	3,923	2,275	1.72	67%
3 years	+3.5%	4.3%	86%	3,616	3,222	1.12	64%
4 years	+4.6%	6.1%	79%	5,115	3,595	1.42	71%
5 years	+5.2%	5.7%	78%	4,778	2,630	1.82	100%
6 years	+4.3%	4.6%	86%	3,780	2,439	1.55	57%

Table 1. Central projection against outturn, all editions, 75 targets, both forecasts scored on the same targets. MAE in prisoners. The ratio column is MoJ error over no-change error; above 1 means the no-change forecast was more accurate. The final column is the share of targets on which it was closer.

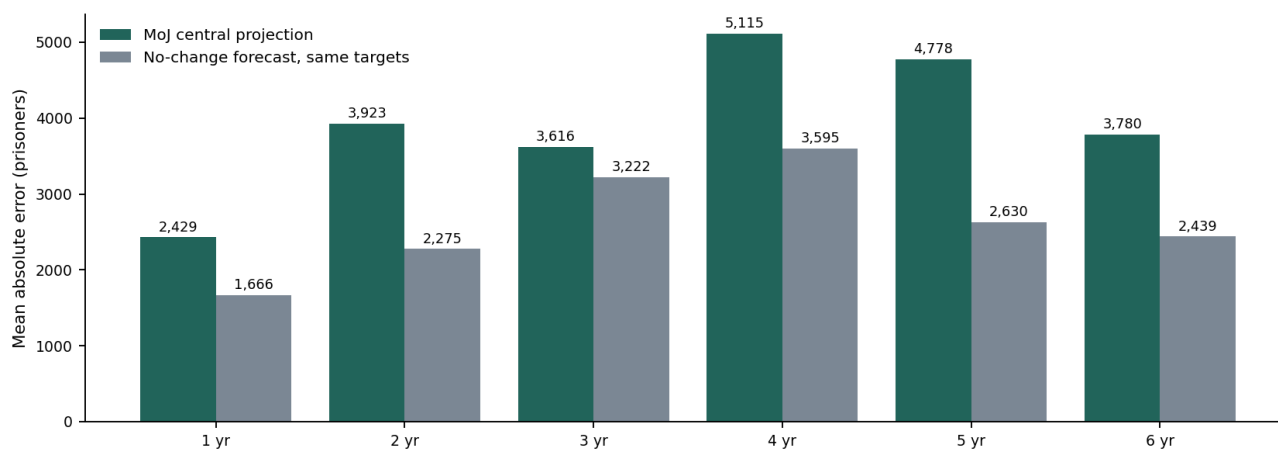


Figure 1. Mean absolute error by horizon. The no-change forecast wins at every horizon.



Figure 2. Every central path since 2008 against the weekly outturn. Editions from 2020 onward, in accent, all project a rise the population has not made.

By era

EDITIONS	MEAN ERROR	MEAN ABS. % ERROR	SHARE TOO HIGH	MOJ MAE	NO-CHANGE MAE	NO-CHANGE CLOSER
2008 to 2014	+2.3%	3.4%	75%	2,895	1,623	29 of 40
2015 to 2019	+4.3%	4.9%	91%	3,984	3,212	15 of 21
2020 to 2025	+7.6%	7.6%	100%	6,506	4,703	10 of 14

Table 2. The bias has worsened in each era, and the no-change forecast wins in every era, including the 2008 to 2014 editions that predate any emergency release scheme. No target from a 2020-onward edition has been too low.

Direction and range

Two further tests do not depend on magnitude. First, direction: did the central path get the sign of the change from the base year right? Across 75 targets it did so 63% of the time, and 53% at both one and two years. The projections called a rise in 73% of targets; the population rose in 52%. At the horizons that bear on capacity decisions, the central path is about as informative on direction as a coin toss, and it is biased toward predicting growth. Second, coverage: forty-five targets have a published low and high scenario. Outturn fell inside that range 67% of the time overall, but only 50% at one and two years, where it fell below the low scenario in 38% of targets. Coverage improves to 86% at five years because the range widens: the mean half-width at three years, 3,343, is about the same as the mean absolute error at three years, 3,616.

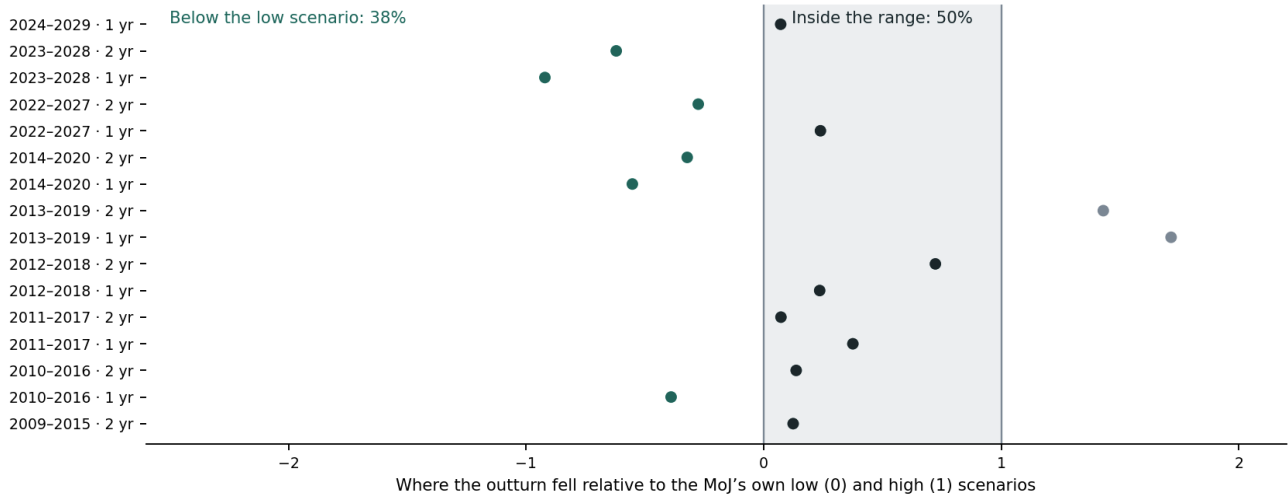


Figure 3. Each dot is a target at one or two years with a published low and high scenario, placed by where the outturn fell relative to that range. Six of sixteen fell below the MoJ's own low scenario.

By edition

EDITION	1 YR	2 YR	3 YR	4 YR	5 YR	TARGETS	NO-CHANGE WINS	ALL TOO HIGH
2008–2015	n/a	n/a	+3.0	+2.7	+5.7	5	4	yes
2009–2015	n/a	+1.8	+1.5	+4.5	+2.9	5	2	yes
2010–2016	+2.0	+1.5	+5.5	+3.7	+2.9	6	5	yes
2011–2017	+0.6	+3.7	+2.3	+2.2	+4.0	6	4	yes
2012–2018	+0.9	–1.4	–2.0	–0.5	–0.7	6	3	no
2013–2019	–2.5	–4.7	–3.7	–5.0	–1.8	6	5	no
2014–2020	+1.9	+4.7	+4.0	+8.1	+8.9	6	6	yes
2015–2021	+1.8	+1.8	+6.3	+7.6	+12.7	5	5	yes
2016–2021	–1.5	+0.9	+1.3	+5.7		4	1	no
2017–2022	+3.8	+4.7	+8.5	+12.0		4	4	yes
2018–2023	+1.2	+5.9	+8.6	+6.3		4	4	yes
2019–2024	+3.5	+3.7	+0.9	–4.8		4	1	no
2020–2026	+5.6	+8.4	+5.9	+10.9	+11.9	5	4	yes
2021–2026	+5.0	+4.2	+7.0	+11.6		4	2	yes
2022–2027	+1.4	+8.3				2	1	yes
2023–2028	+9.8	+14.2				2	2	yes
2024–2029	+2.0					1	1	yes

Table 3. Percentage error of the central path at each horizon, by edition. n/a marks a target dated before January 2011, outside the weekly series and not scoreable; a blank means the edition published no target at that horizon. The 2025–2030 edition has no scoreable target yet: its first, 30 September 2026, is 91,400 against 86,843 on 24 August, and will be scored against the 28 September bulletin when it publishes in early October. The 2012 and 2013 editions, which projected a fall, are the only two that were too low.



Figure 4. The same table as a picture. Each cell is one target; colour is the error as a share of outturn; a dot marks a target where the no-change forecast was closer. Cells marked n/a predate the weekly series.

Robustness

A hostile reader will ask whether the result is an artefact of two shocks that no projection could have foreseen: the COVID fall of 2020 to 2021, and the emergency releases from October 2023. It is not. The table below re-scores the central path after removing those target years, after removing both, after switching outturn to the prisons-only definition, and after dropping the five largest misses.

CUT	TARGETS	MEAN ERROR	SHARE TOO HIGH	NO-CHANGE CLOSER	WINS AT EVERY HORIZON
All targets	75	+3.8%	84%	72%	yes
Excluding COVID targets, Mar 2020 to Dec 2021	65	+3.2%	82%	72%	yes
Excluding release-scheme targets, Oct 2023 on	66	+3.2%	82%	71%	yes
Excluding both: normal years only	56	+2.3%	79%	71%	yes
Normal years, pre-2020 editions only	52	+2.1%	77%	73%	yes
Prisons-only outturn, IRCs excluded	58	+3.9%	84%	74%	yes
Dropping the five largest absolute errors	70	+3.2%	83%	70%	yes
Horizons one to three only	44	+3.2%	86%	68%	yes

Table 5. The headline survives every cut. The magnitude does not: about 40% of the measured overshoot is attributable to the two shock periods, and the structural overshoot in ordinary years is about +2%.

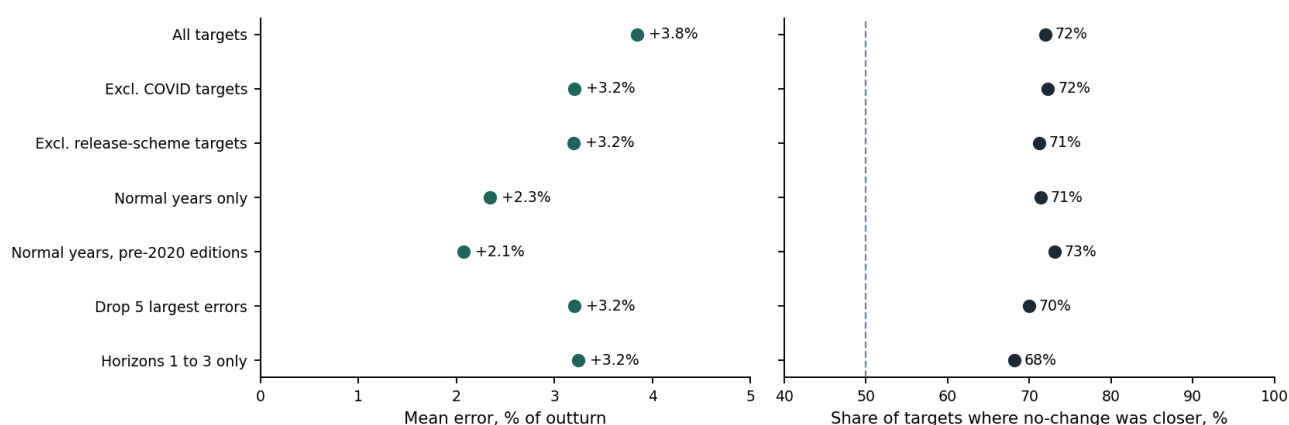


Figure 5. Mean error and the share of targets where no-change was closer, under each cut in Table 5.

Targets within an edition are not independent, so the unit of inference should be the edition. Fifteen of seventeen editions have a positive mean error (one-sided sign test $p = 0.001$). Thirteen of seventeen have a higher mean absolute error than the no-change forecast ($p = 0.02$). The mean of edition-level mean errors is +4.2% with a standard deviation of 3.6, $t = 4.8$ on seventeen observations. Even seventeen overstates independence, because editions share target years and a low outturn year counts against several at once; the honest description is a small, correlated sample that points the same way on every test available.

Two objections cannot be settled from these data. First, the no-change forecast wins because the population was flat; in the trending decades before 2011 it would have lost, and the MoJ's method was built in that era. The scorecard is a statement about the last fifteen years, not about forecasting in general. Second, if capacity binds on the population through channels other than release timing, the pre-2020 overshoot could itself be constrained demand rather than error. The proportion of sentence served held at 60 to 64% throughout 2015 to 2023, which argues against a binding constraint on release, but it does not rule out effects through remand, home detention curfew or sentencing behaviour.

Where the error sits

Every edition since 2017 publishes an unrounded base-year actual broken down by custody type: remand, determinate, indeterminate, recall, non-criminal. Where an earlier edition projected that same reference month, the error can be decomposed component by component. Thirteen such pairs exist.

COMPONENT	MEAN ERROR	SHARE OF TOTAL ERROR	LARGEST CONTRIBUTOR IN
Determinate	+3,272	95%	54% of pairs
Other (non-criminal, fine defaulters)	+426	12%	15%
Indeterminate	+174	5%	0%
Recall	-144	-4%	0%
Remand	-292	-8%	31%

Table 4. Component decomposition of projection error, 13 matched pairs. Positive means projected above actual. Non-criminal and fine defaulter prisoners are combined; the non-criminal category fell from 1,422 to 367 over the period and its error is largely definitional churn.

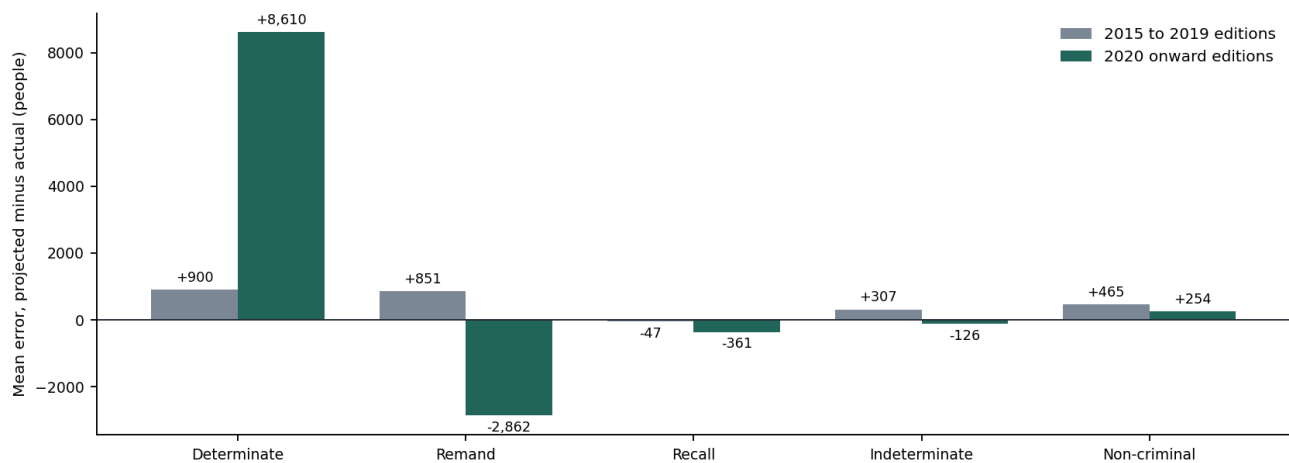


Figure 6. Mean component error by edition era. The 2020-onward editions are +8,610 on determinate and -2,862 on remand; the two partly cancel.

Determinate accounts for 95% of the error and recall is never the largest contributor. Remand has a near-zero mean but a mean absolute error of 1,893: a large miss in both directions that averages out. The two biggest components err in opposite directions, so the gross component error, 6,896 on average, is 1.9 times the net error of 3,627. The worst case is the 2022 edition's November 2023 target: net error 1,165, gross 8,723, a ratio of 7.5. It got the total nearly right by pairing a determinate overshoot of 4,608 with a remand undershoot of 3,086.

The composition shifts across eras. Editions from 2015 to 2019 were mildly high on everything: determinate +900, remand +851. Editions from 2020 onward are +8,610 on determinate and -2,862 on remand. Nineteen successive-edition revisions of a common target date point the same way: determinate accounts for 105% of revision movement and is revised down, by 2,621 on average, edition after edition.

The interpretation is that the determinate overshoot is policy: release points were cut after each edition was published, by ECSL in 2023, SDS40 in 2024 and the progression model in 2026. The remand undershoot is demand: the court backlog. These are separate failures that happen to offset, and reporting a single net bias hides both.

The conditional-projection defence

The MoJ's strongest reply is that the projections are conditional. They model the population under existing release rules; the rules then change in response; the projection is falsified by design. On this reading the overshoot is a measure of the

policy response, not of forecasting skill, and a naive forecast wins precisely because it silently assumes the government will act.

WHY THE DEFENCE DOES NOT RESCUE THE NUMBER

The 10-Year Prison Capacity Strategy commits up to £7 billion to 14,000 places by 2031 against a demand assumption of about 3,000 additional prisoners a year in the absence of further action. The realised growth rate over the fifteen years of weekly data is 246 a year. No five-year window since 2011 has sustained even 2,000 a year.

That is a fair description of what has happened and this note does not dispute it. But it does not rescue the projections for the purpose they are put to. A counterfactual that has been too high in every year since 2020, and that the government itself falsifies each time it is published, cannot carry that weight.

There is a second reason the defence is incomplete. The no-change forecast wins in the 2008 to 2014 editions too: closer on 29 of 40 targets, with a mean absolute error of 1,623 against the MoJ's 2,895, on identical targets. Those editions predate ECSL, SDS40 and the progression model by a decade. The overshoot predates the policy responses that are said to explain it.

Why it matters now

Fifteen years of weekly data show a prison estate that has run between 96% and 99% full in every calendar year, with headroom above 5,000 in ten weeks in total. The supply side has delivered a net 641 places since January 2011 on this series, or 1,005 between 2010 and 2024 on the Public Accounts Committee's figure, against thousands announced. The demand side is planned against a growth rate that has never been sustained. Both halves of the capacity plan rest on numbers that the record does not support, and the demand half has not previously been scored.

The 2025–2030 edition projects 91,400 for September 2026. The August 2026 bulletin gives 86,843. The nineteenth edition is expected in December 2026 and will be scored on publication.

What would prove this wrong

- **An outturn series on a different basis from the projections.** Tested: the editions' own base actuals match the bulletin total to within 40 people since 2018, and the result holds on the prisons-only definition.
- **The overshoot being entirely the two shocks.** Tested: removing COVID and release-scheme years halves the magnitude to +2.3% and leaves the direction, the share too high and the no-change result unchanged.
- **A scored point going the other way.** The 2025–2030 edition's first target, 30 September 2026, is 91,400; the register will score it in early October, and the nineteenth edition will be scored on the day it is published. If either is closer than no-change, that goes in the next version of this paper.

Data notes and caveats

- Two source defects were found. The 2022–2027 edition's central block stops at November 2024 while its low and high blocks run to 2026; the missing years are recoverable from the monthly appendix. And "Scenario 2" is the central path in the 2012 and 2013 editions but the high path in the 2014 edition, so generic scenario labels must be resolved against the explicit central column rather than read by position.
- The 2011–2017 edition's base-year row is a scenario value, not an actual; it is excluded from the calibration test. The 2016–2021 edition publishes only four forward years.

- Component decomposition rests on thirteen matched pairs, weighted toward the June reference grid of the 2015 to 2019 editions, and the two most extreme rows come from the same edition. The era contrast rests on fewer independent observations than the row count suggests.
- Reference dates vary between the first and last day of the month across editions; decomposition matches on year and month and the largest day gap used is 29 days.
- The no-change benchmark uses the edition's own base actual where published, otherwise the bulletin at 30 June of the edition year. This slightly favours the MoJ where their base is a September or November figure and the population was seasonally higher.
- Everything in this note refreshes from the gov.uk content API. Every input is a versioned attachment or an API-enumerable route; there is no page scraping. The refresh reads every source Notes sheet before rebuilding any series and stops if a definition has changed.

Data and code

Corrections. Five made before publication and recorded in the definitions log: an unscored benchmark for three early editions, a direction-accuracy figure of 43% corrected to 63%, non-like-for-like error tables, a misaligned row, and the magnitude claim qualified by the robustness section.

The weekly series (population, capacity, headroom, operating margin, 2011 to date), the establishment panel (2018 to date), the eighteen projection editions in one table, the scorecard by edition, the decomposition and a 106-entry log of definitional changes and corrections are published with this note. The repository includes the refresh script and a standing forecast register, lodged on 3 September 2026, that will be scored against each weekly bulletin without amendment.

Walk Forward Research · enquiries@walkforwardresearch.com
github.com/walkforwardresearch/prisoncap