Cureus
Part of SPRINGER NATURE

Open Access Review Article

Treatment-Free Remission in Chronic Myeloid Leukemia: A Systematic Review of Predictors of Successful Tyrosine Kinase Inhibitor Discontinuation

Review began 09/22/2026
Review ended 10/08/2026
Published 10/17/2026

© **Copyright** 2026
Mohammed Elhaj et al. This is an open access article distributed under the terms of the Creative Commons Attribution License CC-BY 4.0., which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

DOI: 10.7759/cureus.118055

Ebtihal Abdalglil Mohammed Elhaj [1], Riham Elgaali Mohamed Abdelfattah [2], Sara Mohamed Abuelgasim Hassan [3], Islam Mustafa Hamid Mohammed [4], Samih Abdelmutalab Mohamed Abdalla [5], Alaa Khidir Attaelmanan Mahgoub [6], Hanady ME M Osman [7], Nahla Mohammed Wahby [8]

1. Internal Medicine, Riyadh Hospital, Riyadh, SAU 2. General Medicine, Taha Medical Centre, Abu Dhabi, ARE 3. Hematology and Blood Bank, Muhayil General Hospital, Muhayil, SAU 4. Endocrinology, Worcestershire Acute Hospitals NHS Trust, Solihull, GBR 5. Internal Medicine, Al Hammadi Hospital, Riyadh, SAU 6. Obstetrics and Gynecology, Prince Mohammed Bin Abdulaziz Hospital, Al Madinah, SAU 7. Quality Improvement and Patient Safety, Najran Armed Forces Hospital, Ministry of Defense Health Services, Najran, SAU 8. Adult Hematology, King Abdulaziz Medical City, Riyadh, SAU

Corresponding author: Nahla Mohammed Wahby, nahlawahby@gmail.com

Abstract

Treatment-free remission (TFR) has become an explicit therapeutic goal in chronic-phase chronic myeloid leukemia (CP-CML), yet only a minority of patients who start a tyrosine kinase inhibitor (TKI) ultimately achieve durable remission off therapy. Although several clinically relevant predictors have been identified, individual prediction of who will sustain remission remains imperfect. This systematic review was conducted to identify and synthesize original peer-reviewed studies published between 2021 and 2026 that evaluated predictors of successful TKI discontinuation in adults with CP-CML.

PubMed, Embase, Web of Science, and the Cochrane Library were searched for studies published from January 1, 2021, to July 31, 2026. Primary research articles reporting molecular outcomes after TKI cessation together with an analysis of at least one candidate predictor were eligible; reviews, meta-analyses, conference abstracts, editorials, letters, protocols, and preprints were excluded. Risk of bias was appraised with the Risk Of Bias In Non-randomized Studies of Interventions (ROBINS-I) tool. Findings were synthesized narratively without meta-analysis because of heterogeneity in predictor definitions and outcome timing.

Fourteen studies comprising more than 3,000 reported discontinuation attempts across Europe, North America, Asia, and Australia were included. At heterogeneous study-specific follow-up timepoints, reported TFR estimates were generally approximately 55-72%, although direct comparison was limited by differences in follow-up duration and outcome definitions; most molecular recurrences occurred within the first six months after cessation. Duration of deep molecular response (DMR) before cessation was the factor most consistently associated with TFR, supported by duration of prior TKI therapy. Residual disease below conventional detection limits, assessed by digital polymerase chain reaction (PCR) in a small number of studies, and the trajectory of BCR::ABL1 transcripts in the first three months off therapy were associated with outcome and represent promising adjunctive markers, whereas early on-treatment kinetics, transcript type, transcriptomic signatures, and immune parameters were promising but inconsistently replicated. Almost all patients regained major molecular response (MMR) after re-treatment under structured molecular monitoring, and progression was not observed.

Depth and duration of molecular remission, rather than baseline prognostic scores, were most consistently associated with TFR. Prospective validation of composite molecular-immunological models is the principal research priority.

Categories: Hematology
Keywords: bcr::abl1, chronic myeloid leukemia, deep molecular response, digital pcr, predictive factors, systematic review, treatment-free remission, tyrosine kinase inhibitor discontinuation

Introduction And Background

Chronic myeloid leukemia (CML) is a myeloproliferative neoplasm driven by the BCR::ABL1 fusion oncogene arising from the reciprocal translocation t(9;22)(q34;q11). The introduction of BCR::ABL1-targeted tyrosine kinase inhibitors (TKIs) transformed a uniformly fatal disease into a chronic condition, and population-based data now indicate that the life expectancy of patients with chronic-phase CML (CP-CML) who respond to therapy approaches that of the age- and sex-matched general population [1]. This success has shifted the

clinical agenda. Where survival was once the sole endpoint, contemporary management is increasingly concerned with the cumulative burden of indefinite therapy: chronic low-grade toxicity, cardiovascular and metabolic sequelae of second- and third-generation agents, teratogenicity in women of reproductive age, impaired quality of life, and a substantial and recurring economic cost to health systems. Against this background, treatment-free remission (TFR), the maintenance of a major molecular response (MMR) (BCR::ABL1 ⩽0.1% on the International Scale (IS)) after planned cessation of TKI therapy, has been formally incorporated into international recommendations as a legitimate and desirable therapeutic goal for carefully selected patients [2]. Eligibility for discontinuation is generally based on a sustained deep molecular response (DMR), defined as MR4 (BCR::ABL1 ⩽0.01% IS) or deeper, including MR4.5 (⩽0.0032% IS).

The feasibility of discontinuation was first demonstrated in the prospective Stop Imatinib (STIM) trial, in which approximately 40% of patients with sustained undetectable BCR::ABL1 transcripts remained in remission after imatinib cessation, with relapses clustering within the first six months and no progression to the advanced phase [3]. Subsequent multicenter experience, most notably the European Stop Kinase Inhibitors (EURO-SKI) study, confirmed both the safety of the strategy and the relevance of prior treatment duration and duration of DMR to outcome [4]. The redefinition of the re-treatment trigger from loss of undetectable disease to loss of MMR, validated in the A-STIM cohort, further improved the apparent success rate without compromising safety, since fluctuating low-level transcript positivity proved compatible with durable off-therapy control [5]. Parallel immunological work suggested that residual leukemic cells may be held in check by host immune surveillance: natural killer (NK) cell counts and maturation status, and the balance between immune effectors and suppressor populations at the time of cessation, have each been associated with molecular relapse-free survival [6-8]. Taken together, these observations imply that TFR is not simply a function of tumor burden but of the equilibrium between residual disease and immune control.

Despite more than a decade of accumulating evidence, patient selection remains imprecise. Roughly half of eligible patients relapse after a first attempt, and only a minority of all patients who ever start a TKI reach and sustain remission off therapy. Candidate predictors reported across individual studies, prior treatment duration, duration and depth of DMR, Sokal and EUTOS long-term survival (ELTS) risk scores, BCR::ABL1 transcript type, early on-treatment response kinetics, digital polymerase chain reaction (PCR)-quantified residual disease, and immune effector profiles, have frequently been derived in single cohorts, defined inconsistently, and validated irregularly. Two distinct prediction questions should be separated: which patients will achieve the sustained DMR required to become eligible for discontinuation, and which eligible patients will maintain remission once therapy is stopped? The present review addresses the second. The objective of this systematic review was to identify, appraise, and synthesize original peer-reviewed primary research published in the most recent five-year period that examined predictors of successful TKI discontinuation in adults with CP-CML in order to establish which factors are reproducibly associated with sustained TFR, which remain unconfirmed, and where the principal evidence gaps lie.

Review

Materials and methods

Protocol and Reporting

This systematic review was conducted and reported in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 statement [9]. The review was not prospectively registered, which is acknowledged as a limitation. Because the review was restricted to published, anonymized aggregate data, ethical approval and informed consent were not required.

Search Strategy

Four electronic databases were searched: Medical Literature Analysis and Retrieval System Online (MEDLINE) via PubMed, Embase, Web of Science Core Collection, and the Cochrane Library. The search was restricted to studies published from January 1, 2021, onward, and the final search was completed on July 31, 2026. The search combined controlled vocabulary and free-text terms for the population, the exposure, and the outcome. The strategy was adapted to the syntax of each database using Emtree terms for Embase and topic-field searching for Web of Science; the full strategies for all four databases are provided in the Appendices. No language restriction was applied at the search stage. The publication window was defined a priori, before the searches were run, to capture evidence generated under contemporary practice, including standardized IS reporting, loss of MMR as the re-treatment trigger, and current eligibility criteria for discontinuation; foundational trials published before 2021 are therefore discussed narratively for context.

Eligibility Criteria

Eligibility was framed using the Population, Intervention, Comparator, Outcomes, and Study design (PICOS) framework [10], as summarized in Table *1*. Studies were included if they enrolled adults (⩾18 years) with CP-CML who discontinued TKI therapy after achieving a sustained DMR; reported molecular outcomes after cessation (TFR rate, molecular recurrence-free survival, or loss of MMR); and reported an analysis of at least one candidate predictor of outcome, whether clinical, molecular, or immunological. DMR was accepted as

defined by each study (MR4 or deeper, including MR4.5), and the depth used was recorded during data extraction. Both prospective and retrospective observational cohorts, registry analyses, and single-arm interventional discontinuation trials were eligible. Studies were excluded if they addressed pediatric populations, advanced-phase or blast-crisis disease, or Philadelphia-negative disease; if they reported discontinuation outcomes without any predictor analysis; or if they were narrative reviews, systematic reviews, meta-analyses, conference abstracts, editorials, commentaries, letters to the editor, study protocols, preprints, case reports, or case series of fewer than ten patients. Where cohorts overlapped substantially, the most complete or most recent report was retained.

Domain	Inclusion	Exclusion
Population	Adults (≥18 years) with CP-CML in sustained DMR (≥MR4)	Children; accelerated or blast phase; Ph-negative disease
Intervention/Exposure	Planned TKI discontinuation with ≥1 clinical, molecular, or immunological predictor analyzed	Dose reduction only; no predictor analysis
Comparator	Sustained TFR vs molecular recurrence, or above vs below a predictor threshold	—
Outcomes	TFR rate or molecular recurrence-free survival; regain of MMR; progression	No post-discontinuation molecular outcome
Study design	Original peer-reviewed cohort, registry or single-arm discontinuation studies (≥10 patients), published 2021–July 2026	Reviews, meta-analyses, abstracts, editorials, letters, protocols, preprints, case reports

TABLE 1: Eligibility criteria according to the PICOS framework

CP-CML: chronic-phase chronic myeloid leukemia; DMR: deep molecular response; MMR: major molecular response; MR4: BCR::ABL1 ≤0.01% on the International Scale; Ph: Philadelphia chromosome; TFR: treatment-free remission; TKI: tyrosine kinase inhibitor; PICOS: Population, Intervention, Comparator, Outcomes, and Study design

Study Selection

All records were exported to EndNote (Clarivate, London, UK) and de-duplicated. Records were screened by title against the eligibility criteria; because abstracts were not screened as a separate step, title screening was applied inclusively, and any record whose relevance could not be excluded from its title was advanced to full-text assessment. The full text of every potentially eligible record was retrieved and assessed. Reasons for exclusion at full-text stage were documented and are reported in the PRISMA flow diagram; the list of excluded reports with study-specific reasons is available from the corresponding author on request. Screening and selection were performed by a single reviewer, as a second independent reviewer was not available; the full-text assessment was repeated after an interval of two weeks, and any discordant judgments were resolved by rereading the source article against the written protocol. This intra-reviewer check improves consistency but does not substitute for independent duplicate screening and data extraction, and the single-reviewer approach is acknowledged as a major methodological limitation that may have introduced selection and extraction errors. Two additional reports published within the search window were identified during peer review and were assessed in full against the same eligibility criteria.

Data Extraction

Data were extracted into a piloted standardized spreadsheet. Extracted fields comprised first author, year of publication, country or region, study design and setting, recruitment period, sample size, eligibility thresholds for discontinuation (minimum TKI duration and minimum DMR duration and depth), TKI class at cessation, definition of molecular recurrence and of the re-treatment trigger, molecular monitoring schedule and assay sensitivity, duration of follow-up, the TFR rate with its timepoint, the candidate predictors assessed, the effect estimates reported for those predictors with confidence intervals, p-values and adjustment status (univariable or multivariable), and safety outcomes including regain of response after re-treatment and any progression. Where an article reported both univariable and multivariable analyses, the multivariable estimate was extracted preferentially. Numerical data reported only graphically were not estimated from figures. The extraction template is available from the corresponding author on request.

Data Synthesis

A quantitative meta-analysis was not undertaken. Candidate predictors were defined using different thresholds across studies (for example, DMR duration dichotomized at two, three and five years), outcomes

were assessed at different time points (12, 24, 36, and 60 months), molecular recurrence was variably defined as loss of MMR or loss of MR4; and assay sensitivity differed materially between cohorts. Pooling under these conditions would have produced a spuriously precise summary estimate. Outcome metrics (TFR proportion, molecular relapse-free survival, cumulative incidence of recurrence and proportion restarting therapy) were recorded as reported and were not converted or treated as interchangeable, and DMR depth (MR4 or MR4.5) was not harmonized across studies, so TFR rates were compared descriptively only. Results were therefore synthesized narratively and grouped by predictor domain: treatment-related, molecular, and immunological. Direction and consistency of association were tabulated across studies, and the weight of evidence for each predictor domain was described in terms of the number of independent cohorts in which it was tested and confirmed. A predictor was described as independently associated only when it retained significance in multivariable analysis, and the absence of a statistically significant association was interpreted as no demonstrated association rather than evidence of no effect. For predictors measured after cessation, susceptibility to immortal-time bias was considered, because such measurements are available only in patients still in remission at the time of sampling.

Risk of Bias Assessment

Risk of bias was assessed using the Risk Of Bias In Non-randomized Studies of Interventions (ROBINS-I) tool, which was selected because all included studies were non-randomized, comprising observational cohorts, registry analyses, and single-arm interventional discontinuation trials, and because each study evaluated predictors measured within a defined discontinuation exposure. A prognostic-factor tool such as Quality In Prognosis Studies (QUIPS) would be more conceptually aligned with the review question; however, the ROBINS-I domains of confounding, selection of participants, missing data, and selection of the reported result correspond closely to the QUIPS domains of study confounding, study participation, study attrition, and statistical analysis and reporting. The choice of tool is acknowledged as a limitation. Each study was appraised across the seven ROBINS-I domains: confounding, selection of participants, classification of interventions, deviations from intended interventions, missing data, measurement of outcomes, and selection of the reported result. Domain-level judgments of low, moderate, serious, or critical risk of bias were assigned, and an overall judgement was derived using the standard ROBINS-I algorithm, whereby the overall rating corresponds to the most severe domain-level rating. Domain-level and overall judgments were visualized using the RobVis tool. Appraisal was performed by the same single reviewer.

Results

Study Selection

The database searches returned 521 records: 296 from PubMed/MEDLINE, 102 from Embase, 118 from Web of Science, and five from the Cochrane Library. After 316 duplicate records were removed in EndNote, 205 records were screened by title, of which 153 were excluded as irrelevant to the review question. Fifty-two reports were sought for retrieval, all of which were obtained, and all 52 were assessed in full for eligibility. Forty reports were excluded: 19 were reviews, systematic reviews, or meta-analyses; eight were conference abstracts or letters to the editor, including an otherwise relevant multicenter Italian analysis of early post-cessation transcript levels published in correspondence format; seven reported discontinuation outcomes without any predictor analysis; and six were duplicate or superseded reports of cohorts included elsewhere. Twelve studies met all eligibility criteria from the database searches. Two further reports published within the search window, identified during peer review, were assessed in full and met all eligibility criteria, giving 14 included studies (Figure *1*).

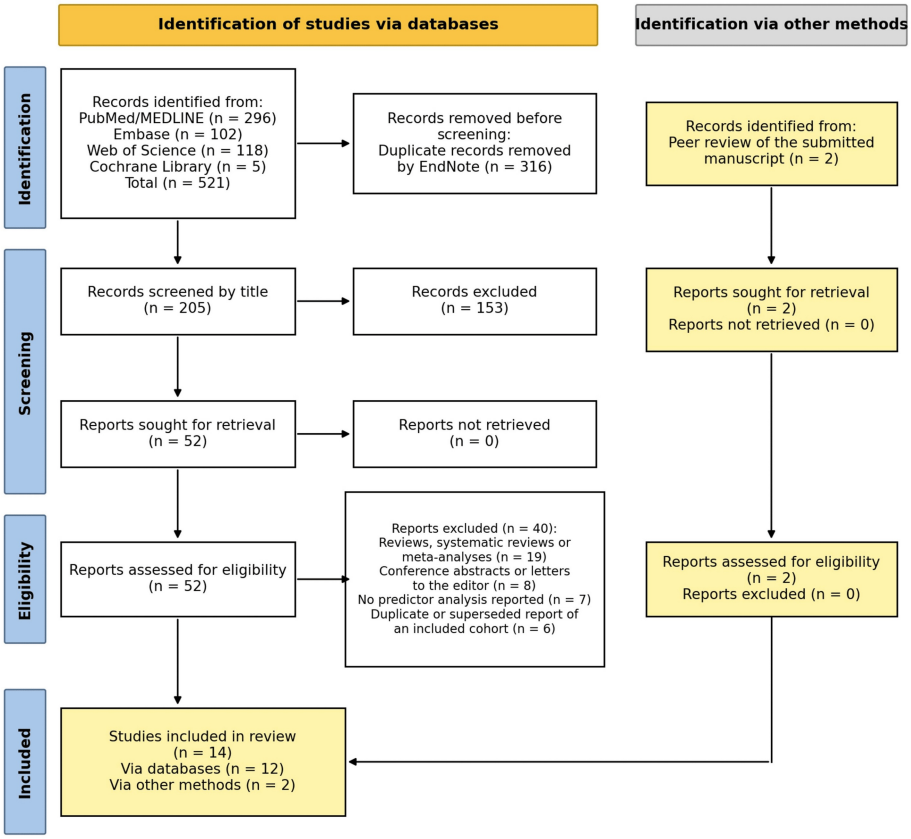

FIGURE 1: PRISMA 2020 flow diagram of study identification, screening and inclusion

PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses; MEDLINE: Medical Literature Analysis and Retrieval System Online

Characteristics of the Included Studies

The characteristics of the 14 included studies [11-24] are summarized in Table 2. The studies were published between 2021 and 2026 and originated from Japan (three), France (three), Australia in collaboration with the Republic of Korea, the United States, the Republic of Korea, Spain with Italian participation, Italy, a pan-European consortium, a transnational European-Canadian consortium, and an international analysis of two multinational nilotinib discontinuation trials. Together they reported on more than 3,000 discontinuation attempts. This figure is the arithmetic sum of the discontinuing patients reported by each study; it includes repeat attempts in one cohort [18], was not de-duplicated, and overlap between trial, registry, and institutional cohorts cannot be excluded (for example, the transcriptomic development cohort [24] was drawn from the STIM2 trial [13]). It should therefore be regarded as an approximate total of reported attempts rather than a count of unique patients. Six studies were retrospective single-center or multicenter cohorts, four were prospective observational or registry cohorts, two were prospective single-arm interventional discontinuation trials, and two were biomarker or prediction-model analyses nested within prospective trial cohorts. Cohort sizes ranged from 45 to 795 patients. Eligibility thresholds for discontinuation were broadly concordant, typically requiring at least three to five years of TKI exposure and one to three years of sustained MR4 or deeper, although the depth defining DMR differed between MR4 and MR4.5, a distinction that was recorded but not harmonized and that materially affected reported success rates. Median follow-up after cessation ranged from 12 months to more than six years.

Study	Country	Design	N	Predictors assessed
Shanmuganathan 2021 [11]	Australia/Korea	Prospective cohort	115	Early BCR::ABL1 halving time
Atallah 2021 [12]	USA	Single-arm trial (LAST)	172	Residual disease by RT-qPCR/ddPCR; PROs
Dulucq 2022 [13]	France	Prospective trial (STIM2)	218	TKI duration; response depth; recurrence kinetics
Park 2023 [14]	Korea	Multicenter cohort	66	Transcript type; ddPCR; DMR duration
Marcé 2024 [15]	Spain/Italy	Multicenter retrospective cohort	263 (107 stopped)	Transcript type; imatinib duration
Mahon 2024 [16]	Europe	Prospective trial (EURO-SKI final)	728	TKI and DMR duration; novel factors
Saugues 2024 [17]	France	Real-world retrospective cohort	408 (70 stopped)	ELTS; halving time; M3/M6 residual disease
Kockerols 2024 [18]	Netherlands/Germany/Canada	Multicenter validation cohort	186	dPCR threshold 0.0023% IS
Ureshino 2025 [19]	Japan	Retrospective cohort (ISAC-TFR)	57	Post-stop BCR::ABL1; TCR/BCR repertoire
Laganà 2025 [20]	Italy	Single-center retrospective cohort	118	Sokal; DMR duration; baseline score
Takahashi 2026 [21]	Japan	Multicenter observational (J-SKI)	795	Clinical factors; DMR duration; post-stop MR4.5
Jo 2026 [22]	Japan	Prospective immunomonitoring	45	CTL phenotype; TCR Vβ; NK, Treg, MDSC
Radich 2026 [23]	International (ENESTfreedom/ENESTop)	Retrospective biomarker analysis of two prospective single-arm trials	270 analyzed	Whole-blood gene expression (RNA-seq) at screening and before cessation; transcriptome panel 12 weeks after cessation
Alcazer 2026 [24]	France	Prediction-model development (STIM2) with external real-world validation	166 (96 + 70)	50-gene peripheral blood transcriptomic signature at cessation

TABLE 2: Characteristics of the included studies

PCR: polymerase chain reaction; CTL: cytotoxic T lymphocyte; ddPCR: droplet digital PCR; DMR: deep molecular response; dPCR: digital PCR; ELTS: EUTOS long-term survival; IS: International Scale; M3/M6: months 3 and 6; MDSC: myeloid-derived suppressor cell; MR4.5: BCR::ABL1 ≤0.0032% IS; NK: natural killer; PROs: patient-reported outcomes; RNA-seq: RNA sequencing; RT-qPCR: quantitative reverse-transcription PCR; TCR: T-cell receptor; TKI: tyrosine kinase inhibitor; Treg: regulatory T cell

Reported outcomes, summarized in Table *3*, comprised TFR proportions, molecular relapse-free survival, cumulative incidence of recurrence, and the proportion restarting therapy, which are related but not equivalent measures. At heterogeneous study-specific time points, reported TFR estimates were generally approximately 55-72%, with lower rates in smaller second-generation TKI subgroups and in cohorts with prolonged follow-up, and the great majority of molecular recurrences occurred within the first six months off therapy. Across all cohorts that reported re-treatment outcomes, virtually every patient who lost MMR regained it after TKI resumption, and progression to accelerated phase or blast crisis was not observed.

Study	Patients discontinuing	Definition of molecular recurrence	Reported outcome estimate (metric; timepoint)	Timing of recurrence	Regain of response/ progression
					Not the primary

Shanmuganathan 2021 [11]	115	Loss of MMR	TFR proportion: 55% (12 months)	Predominantly early	endpoint; no progression reported
Atallah 2021 [12]	172	Loss of MMR (BCR::ABL1 IS >0.1%)	TFR proportion: 60.8% (minimum 3-year follow-up)	Concentrated in first 6 months	Response regained on re-treatment; no progression
Dulucq 2022 [13]	218	Loss of MMR	Cumulative incidence of recurrence: 58.3% (5 years)	Early <6 months (fast kinetics), 6–24 months, late >24 months (slow kinetics)	Response regained; no progression
Park 2023 [14]	66	Loss of MMR	Molecular relapse-free survival: 57.8% (12 months)	Median follow-up 16.5 months; 43.9% relapsed	All regained MMR at median 2 months
Marcé 2024 [15]	107	Loss of MMR	TFR proportion: 65% at 24 months (imatinib); 33% at 64 months (2G-TKI)	Not separately reported	Not separately reported
Mahon 2024 [16]	728	Loss of MMR	TFR proportion: 61% at 6 months; 46% at 36 months	Predominantly first 6 months	Response regained; no progression
Saugues 2024 [17]	70	Loss of MMR	Proportion restarting therapy for molecular relapse (not a TFR estimate): 40.0%	Including late relapses beyond 24 months	Response regained on re-treatment
Kockerols 2024 [18]	186 (163 first attempts)	Loss of MMR by 36 months	Cumulative incidence of recurrence: 33% (dPCR <0.0023%) vs 70% (≥0.0023%) (36 months)	Within 36 months	Not the primary endpoint
Ureshino 2025 [19]	57	Loss of MMR	TFR proportion: 67.9% after first attempt; 31.2% after second (1 year)	Median time to MMR loss early after cessation	Response regained; no progression
Laganà 2025 [20]	118	Loss of MMR	Recurrence-free proportion: 69.9% overall (median follow-up 72.5 months)	79.7% recurrence-free at 6 months, 74.4% at 24 months	Response regained; no progression
Takahashi 2026 [21]	795	Loss of MMR	TFR rate: 65.2% at 5 years; 60.6% at 10 years	Predominantly early; median 112.5 days to MMR loss in second-attempt subgroup	99% regained MMR within 8 months; no progression
Jo 2026 [22]	45	Loss of MR4	TFR proportion (loss of MR4 as event): 71% (12 months)	Within first year	Not separately reported
Radich 2026 [23]	270 analyzed (164 ENESTfreedom; 106 ENESTop)	Loss of MMR	TFR proportion in parent trials: 52% (ENESTfreedom) and 58% (ENESTop) at 48 weeks	Mostly within 24 weeks	Not reported in this analysis
Alcazer 2026 [24]	166 (96 development; 70 validation)	Loss of TFR by 2 years, as defined in the source cohorts	No pooled TFR rate; model AUROC for 2-year TFR 0.83 (development) and 0.71 (external validation)	Within 2 years (prediction endpoint)	Not reported in this analysis

TABLE 3: Summary of treatment-free remission outcomes across the included studies

Outcome metrics are presented as reported in the source studies and are not directly interchangeable; cumulative incidence of recurrence [13,18] and the proportion restarting therapy [17] are recurrence measures rather than TFR estimates.

PCR: polymerase chain reaction; AUROC: area under the receiver operating characteristic curve; dPCR: digital polymerase chain reaction; IS: International Scale; MMR: major molecular response; MR4: BCR::ABL1 ≤0.01% IS; TFR: treatment-free remission; 2G-TKI: second-generation tyrosine kinase inhibitor

Treatment-Related Predictors

Treatment-related predictors were examined in 11 of the 14 studies and produced the most consistent signal (Table *4*). In the largest observational discontinuation cohort, comprising 795 Japanese patients, multivariable Cox regression identified duration of DMR as the sole independent predictor of successful TFR, with each additional year of DMR conferring a 12.5% relative reduction in the risk of molecular relapse (hazard ratio 0.875, 95% confidence interval 0.829-0.924; p<0.0001); the five-year TFR rate in that cohort was 65.2% [21]. The final analysis of the EURO-SKI trial similarly confirmed duration of DMR before cessation as a dominant prognostic determinant in a large prospective European population [16]. Updated long-term follow-up of the French STIM2 cohort reinforced the contribution of both prior imatinib duration and depth of molecular remission at cessation, while demonstrating that late recurrences beyond two years, though uncommon, do occur and follow markedly slower kinetics than early events [13]. A single-center Italian retrospective analysis of 118 consecutive patients identified sustained DMR duration of less than five years, together with high-risk Sokal score, as independently associated with shorter TFR, and combined these variables into a baseline scoring system for molecular recurrence [20]. A Korean cohort likewise found DMR duration of at least 48 months to be associated with prolonged molecular relapse-free survival [14], and a Spanish multicenter analysis observed that patients treated with imatinib for at least 119 months relapsed significantly less often than those discontinuing earlier [15]. By contrast, several cohorts found no significant independent effect of age, sex, baseline risk score, TKI generation, or line of therapy, and the Japanese registry specifically reported comparable TFR rates for imatinib and second-generation agents [21].

Predictor domain	Specific predictor	Studies reporting a positive association	Studies reporting no significant association	Overall consistency
Treatment-related	Longer duration of deep molecular response before cessation	[13,14,16,20,21]	—	Consistent; the most reproducible association
Treatment-related	Longer duration of prior TKI therapy	[13,15,21,22]	—	Consistent and closely correlated with DMR duration
Treatment-related	TKI generation or line of therapy	—	[21]	No evidence of an independent effect once discontinuation criteria are met
Clinical/baseline	High-risk Sokal score	[20]	[21]	Inconsistent; significant in one single-center cohort only
Clinical/baseline	ELTS score	[17] (2G-TKI subgroup only)	[21]	Weak and inconsistent
Clinical/baseline	Age and sex	[17] (female sex, 2G-TKI subgroup)	[21]	Not reproducible
Molecular (at cessation)	Residual disease below conventional RT-qPCR detection, assessed by ddPCR or dPCR	[12,14,18]	—	Concordant direction in three cohorts using different platforms (RT-qPCR/ddPCR [12], ddPCR [14], dPCR [18]), timings and thresholds; promising adjunctive marker, not yet standardized
Molecular (post-cessation)	Retention of MR4.5 at 1–3 months after stopping	[12,19,21]	—	Consistent in three studies; susceptible to immortal-time bias unless landmark methods are used
Molecular (on-treatment)	Short BCR::ABL1 halving time in first 3 months of therapy	[11]	[17]	Directly contested between two high-quality cohorts
Molecular	BCR::ABL1 transcript type (e14a2 favourable)	[14,15]	[15] (2G-TKI subgroup)	Context-dependent; effect apparent only with prolonged imatinib exposure
Transcriptomic	Peripheral blood gene-expression signature	[24]	[23]	Conflicting; a 50-gene signature showed moderate discrimination on external validation [24], whereas bulk RNA-seq and panel profiling identified no predictive signature [23]
Immunological	Cytotoxic T-lymphocyte activation and TCR Vβ clonality	[22]	[19]	Promising but not replicated
Immunological	NK cells, regulatory T cells, MDSCs	—	[19,22]	Not confirmed within the included studies

TABLE 4: Direction and consistency of candidate predictors of treatment-free remission

Numbers in square brackets refer to the reference list. Absence of a statistically significant association does not establish absence of effect.

PCR: polymerase chain reaction; ddPCR: droplet digital PCR; dPCR: digital PCR; DMR: deep molecular response; ELTS: EUTOS long-term survival; MDSCs: myeloid-derived suppressor cells; MR4.5: BCR::ABL1 ≤0.0032% on the International Scale; NK: natural killer; PCR: polymerase chain reaction; RNA-seq: RNA sequencing; RT-qPCR: quantitative reverse-transcription PCR; TCR: T-cell receptor; TKI: tyrosine kinase inhibitor; 2G-TKI: second-generation tyrosine kinase inhibitor

Molecular Predictors at and After Cessation

Molecular predictors measured at or after cessation formed the second major domain. Conventional quantitative reverse-transcription PCR (RT-qPCR) remains the standard method for BCR::ABL1 monitoring before and after discontinuation; three studies additionally quantified residual disease below its detection limit, using different platforms, sampling times, and thresholds [12,14,18]. In an independent transnational validation of a previously derived threshold, 186 patients from Canada, Germany, and the Netherlands were

stratified by a BCR::ABL1 dPCR cut-off of 0.0023% on the IS; among first discontinuation attempts, the probability of molecular recurrence was 33% below the threshold and 70% above it, an association independent of prior treatment duration, MR4 duration, and TKI type (hazard ratio 3.66, 95% confidence interval 2.06-6.51; p<0.001) [18]; this threshold was derived and validated within a single laboratory network. The prospective United States LAST study reached a concordant conclusion using a combination of RT-qPCR and droplet dPCR (ddPCR), showing that detectable BCR::ABL1 at the time of cessation, or at three months thereafter, identified patients at materially higher risk of molecular recurrence; 60.8% of the 172 enrolled patients remained in TFR, and patient-reported outcomes improved after stopping [12]. The trajectory of transcripts in the early off-therapy period proved similarly informative: retention of MR4.5 at one, three, and six months after cessation was strongly associated with superior treatment-free survival in the Japanese registry [21], and in a Japanese single-institution cohort, maintenance of BCR::ABL1 below MR4.5 at three months was associated with a marked reduction in relapse risk (hazard ratio 0.099, 95% confidence interval 0.012-0.829; p=0.033) [19]; because such measurements are available only in patients still in remission at the time of sampling, these estimates may be inflated unless landmark methods are used. Evidence concerning BCR::ABL1 transcript type was more equivocal: a Spanish multicenter analysis found that the e14a2 transcript favored TFR only when combined with prolonged imatinib exposure [15], and a Korean study reported that patients with e13a2 transcripts and detectable disease by ddPCR at cessation had the shortest molecular relapse-free survival [14].

Early On-Treatment Response Kinetics

The predictive value of early on-treatment response kinetics was directly contested between two of the included studies. An Australian-Korean analysis showed that the BCR::ABL1 halving time over the first three months of first-line TKI therapy was the strongest independent predictor of subsequent sustained TFR, with a halving time below 9.35 days conferring an 80% probability of sustained remission compared with 4% for a halving time exceeding 21.85 days, and validated this relationship in an independent dataset [11]. A subsequent real-world French comparative study of 408 patients, however, found that although halving time and residual disease at months three and six were excellent predictors of subsequent DMR, none of these early parameters predicted either the attainment of discontinuation criteria or the maintenance of TFR once treatment had been stopped [17]. The authors of that study interpreted the discrepancy as evidence that the determinants of the initial reduction in tumor burden differ from those governing eradication or immune containment of residual disease during prolonged DMR.

Immunological and Transcriptomic Predictors and Repeat Attempts

Immunological predictors were evaluated in two included studies, with divergent results. A one-year prospective immuno-monitoring study of 45 Japanese patients found that cytotoxic T-lymphocyte activation features, a CD8-predominant profile, memory-over-effector distribution, and highly activated clones on T-cell receptor Vβ repertoire analysis identified patients with the highest likelihood of durable TFR when combined with at least seven years of TKI exposure and at least one year of DMR; 71% of the cohort maintained TFR at 12 months, whereas NK cells, effector regulatory T cells, and myeloid-derived suppressor cells did not discriminate outcome [22]. In contrast, T-cell receptor repertoire sequencing in a retrospective Japanese cohort of 57 patients demonstrated no distinct clonal expansion associated with TFR, identifying instead an age-related decline in T-cell diversity, and concluded that T-cell immunity in patients in sustained DMR closely approximates that of healthy individuals [19].

Two analyses evaluated peripheral blood transcriptomic profiling, also with divergent results. In a development cohort of 96 STIM2 patients, a 50-gene signature measured at imatinib cessation predicted TFR at two years (area under the receiver operating characteristic curve (AUROC) 0.83, 95% confidence interval 0.73-0.93), with lower but retained discrimination in an independent real-world validation cohort of 70 patients (AUROC 0.71, 95% confidence interval 0.58-0.83) [24]. By contrast, retrospective RNA sequencing and transcriptome-panel analyses of samples from the ENESTfreedom and ENESTop nilotinib discontinuation trials, obtained at screening, before cessation, and 12 weeks after cessation, identified no gene-expression signature that discriminated sustained TFR from relapse, with penalised regression models performing no better than random models [23].

The ISAC-TFR cohort provided the most detailed data on repeat attempts, reporting a one-year TFR rate of 67.9% after a first attempt but only 31.2% after a second [19]; the Japanese registry reported a comparable figure of 41% for second attempts [21].

Risk of Bias in the Included Studies

Risk-of-bias judgments for the 14 included studies are presented in Figure *2*. Seven studies were judged at low overall risk of bias [11-13,16,18,21,22], with low-risk ratings across all seven ROBINS-I domains; these comprised the prospective trials and prospective or registry cohorts with pre-specified eligibility criteria, standardized molecular monitoring and multivariable adjustment. One prediction-model study was judged at moderate overall risk of bias [24]. The remaining six studies were judged at serious overall risk of bias [14,15,17,19,20,23]. Bias due to confounding was the most frequent source of concern, rated serious in three

studies [14,15,19] and moderate in four [17,20,23,24], reflecting incomplete adjustment for the strongly inter-correlated variables of treatment duration, DMR duration and response depth. Bias in the selection of participants was serious in three studies: two real-world cohorts in which the decision to discontinue was made at clinician discretion [17,20] and one extreme-phenotype comparison restricted to consenting patients [23], and moderate in four further studies [14,15,19,24]. Bias due to missing data was rated serious in two studies [17,23], the latter reflecting sample quality-control failures, and moderate in five [14,15,19,20,24], and bias due to deviations from intended interventions was moderate in four retrospective cohorts [15,17,19,20]. Bias in selection of the reported result was serious in one exploratory analysis testing multiple candidate parameters without pre-specification [19] and moderate in five studies [14,15,20,23,24]. By contrast, classification of interventions and measurement of outcomes were judged at low risk of bias in all 14 studies, consistent with the uniform definition of the discontinuation exposure and the use of standardized IS molecular reporting across cohorts.

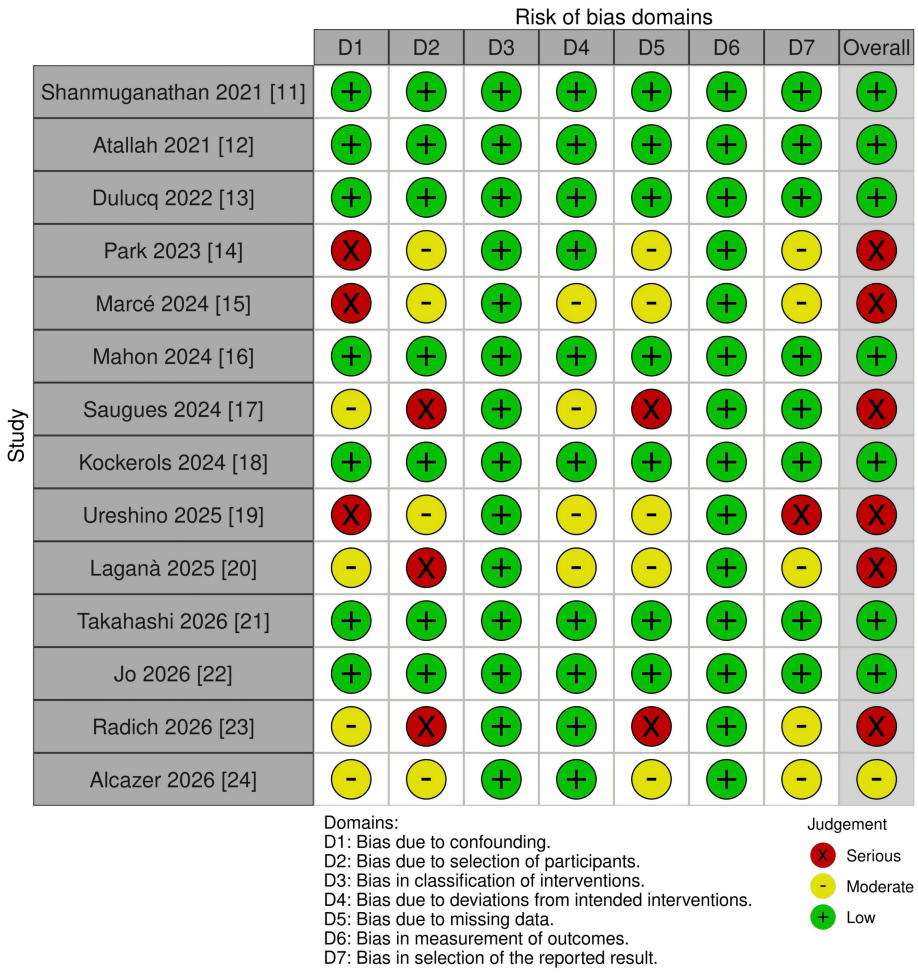

FIGURE 2: Risk of bias assessment of the included studies using the ROBINS-I tool

ROBINS-I: Risk Of Bias In Non-randomized Studies of Interventions

[11-24]

Discussion

Principal Findings

This systematic review of 14 original studies published between 2021 and 2026, encompassing more than 3,000 reported discontinuation attempts across four continents, identifies a consistent pattern among the factors associated with whether a patient with CP-CML will remain in remission after TKI withdrawal. The most consistent finding is that the duration and depth of molecular remission achieved before cessation, rather than any baseline characteristic of the patient or the disease, were associated with outcome. Duration of DMR was associated with TFR in the largest observational cohort assembled to date, where it was the sole

factor retained in multivariable analysis, and was corroborated in the final analysis of the largest prospective European trial, in the updated long-term follow-up of a French national cohort, and in analyses from Italy, Spain, and the Republic of Korea [13-16,20,21]. Duration of prior TKI exposure travels closely with this variable, and the two are statistically entangled: patients cannot accrue years of DMR without years of therapy. Because these variables are strongly intercorrelated and all included studies were observational, the findings represent reproducible associations rather than causal determinants. The evidence for DMR duration derives predominantly from studies at low risk of bias [13,16,21], whereas the evidence for transcript type, baseline scoring systems, and T-cell repertoire derives mainly from studies at serious risk of bias [14,15,19,20], which tempers the weight that can be given to the latter. Their joint effect is consistent with the treatment-duration and response-duration thresholds embedded in contemporary international recommendations [2].

Residual Disease and Early Post-cessation Kinetics

The second consistent theme concerns the quantification of residual disease. Conventional quantitative reverse-transcription PCR is limited by its detection floor, and three of the included studies suggest that pushing sensitivity beyond that floor may add prognostic information. dPCR applied immediately before cessation stratified risk with an approximately 3.7-fold higher recurrence hazard that persisted after adjustment for treatment duration, response duration, and drug class [18]. The prospective United States experience arrived at a concordant conclusion using detectability rather than a continuous threshold [12]. However, platforms, sampling times, and thresholds differed between these studies, and dPCR thresholds are not yet standardized across laboratories; these assays should therefore currently be regarded as promising adjunctive prognostic biomarkers rather than established decision-making tools. Read together with the observation that retention of MR4.5 at one, three, and six months after stopping is associated with durable treatment-free survival [19,21], these data are, in our interpretation, compatible with viewing discontinuation as a dynamic rather than a static decision, in which the first months off therapy may provide informative prognostic information. This window has not, however, been prospectively validated as a decision-making tool in any included study.

Conflicting Evidence

A striking feature of the present evidence base is the direct contradiction between two high-quality analyses of early on-treatment kinetics. The demonstration that the BCR::ABL1 halving time during the first three months of first-line therapy predicts eventual sustained TFR, with a twentyfold difference in probability between the fastest and slowest responders and validation in an independent dataset, was an important result [11]. Its apparent refutation in a real-world French cohort of 408 patients, where halving time and residual disease at months three and six predicted subsequent DMR excellently but predicted neither attainment of discontinuation criteria nor maintenance of remission off therapy, cannot be dismissed as a chance finding [17]. The two cohorts differed in the proportion of patients receiving second-generation agents in the first line, in the distribution of ELTS risk categories, in the number of patients who actually discontinued, and in the duration of post-cessation follow-up, the French analysis capturing late relapses that a 12-month endpoint would miss. The reconciliation proposed by the French investigators is that early kinetics measure the sensitivity of the bulk leukemic population to kinase inhibition, whereas TFR depends on the fate of a quiescent, kinase-independent stem cell reservoir and on the host immune response to it; this explanation is biologically plausible but was not directly tested in the included studies and should be regarded as a hypothesis. If it is correct, early response parameters would be useful for identifying patients likely to become eligible for a discontinuation attempt, but poor predictors of what happens once the attempt is made are a distinction that deserves explicit attention in future analyses.

The immunological hypothesis, which has animated much of the translational work in this field, receives only qualified support from the studies included here. Earlier investigations reported associations between NK cell counts and molecular relapse-free survival after imatinib cessation [6], between the proportion of mature NK cells and successful discontinuation [8], and between a composite score of immune effectors and suppressors and TFR [7]. Within the present review, one prospective study found that cytotoxic T-lymphocyte activation phenotype and T-cell receptor Vβ clonality identified the patients most likely to sustain remission, while failing to replicate the NK cell, regulatory T-cell, and myeloid-derived suppressor cell associations in the same cohort [22]. Another found no clonal T-cell signature at all [19]. Transcriptomic profiling showed a similar divergence, with an externally validated signature in one study [24] and no predictive signature in two large trial-derived datasets in another [23]. The disagreement is not necessarily a contradiction. Immune and transcriptomic parameters are measured by heterogeneous platforms, at varying times relative to cessation, often in cohorts of a few dozen patients, and are almost always analyzed without correction for multiple comparisons. What the field lacks is standardized, adequately powered, prospectively specified immune endpoints embedded within large discontinuation cohorts.

Comparison With Earlier Studies

The foundational STIM trial reported that approximately 40% of highly selected patients remained in remission, with relapses concentrated in the first six months [3], and long-term follow-up of that cohort

confirmed the durability of remission in those who passed the early hazard period [25]. The EURO-SKI interim analysis subsequently established treatment duration and DMR duration as prognostic in a far broader population [4]. The studies synthesized here report higher success rates, commonly between 55% and 72%, and the explanation is largely methodological rather than biological. The change in re-treatment trigger from loss of undetectable disease to loss of MMR, validated in the A-STIM cohort, mechanically increases the measured success rate by permitting fluctuating low-level positivity to persist without intervention [5]. Greater assay sensitivity has a subtler but opposite effect: cohorts defining DMR as MR4.5 using highly sensitive standardized assays select a deeper-remission population than those defining it as MR4, and this difference has been invoked to explain why the Japanese registry reported a five-year success rate of 65.2% against the 46% three-year rate of EURO-SKI [21]. Because DMR depth was recorded as defined by each study and was not harmonized in the present synthesis, reported TFR rates are not directly comparable between studies unless the response definition, the assay sensitivity, and the re-treatment trigger are all specified, a point that has considerable practical importance when quoting figures to patients.

Safety and Repeat Attempts

The consistency of the safety signal across every included study deserves emphasis. Essentially all patients who lost MMR regained it after resuming therapy, typically within a few months, and no included cohort observed progression to accelerated phase or blast crisis during discontinuation follow-up. This finding aligns with earlier trial experience [3,4] and supports offering a discontinuation attempt to eligible patients even when the predicted probability of success is moderate. This reassurance is, however, conditional on strict adherence to frequent molecular monitoring and timely re-treatment, as specified in the source protocols, and cannot be assumed where such monitoring is unavailable, or adherence is uncertain. Under these conditions, the relevant considerations become the psychological burden of a failed attempt, the intensity of monitoring required, the well-described TKI withdrawal syndrome, and the value the individual patient places on time off therapy. Evidence that patient-reported outcomes improve after cessation [12] is a relevant counterweight.

Data on repeat attempts remain limited but are clinically pressing, since roughly half of patients who try will fail. The two included studies that addressed this question reported second-attempt success rates of 31.2% and 41%, both substantially below first-attempt rates in the same cohorts [19,21]. These figures are consistent with dedicated rechallenge trials, including experience with dasatinib re-treatment after a failed first attempt [26]. Whether switching to a more potent agent before a second attempt improves the odds, and whether the optimal duration of consolidation before rechallenge differs from that before the first attempt, are unresolved questions. The observation in one included registry that duration of DMR before the second cessation showed a non-significant trend toward benefit suggests that the same biological principle may apply, but the sample sizes involved are too small to support firm recommendations [21].

Research Gaps

Several knowledge gaps follow directly from this synthesis. First, only one included study externally validated a multivariable prediction model, a transcriptomic signature whose discrimination declined from development (AUROC 0.83) to an independent real-world cohort of 70 patients (AUROC 0.71) [24], and no model has been prospectively validated; the clinical scoring systems that have been proposed [20] were developed and tested within the same or closely related populations and require independent validation before clinical adoption. Second, the relationship between molecular and immunological predictors has not been formally modeled; residual disease burden and immune competence may interact rather than act additively, but no included study was designed or powered to test such an interaction. Third, the field has concentrated almost exclusively on patients who reach the discontinuation threshold, a group that represents a minority of all patients who start therapy. Predicting who will become a candidate is a distinct and equally important question, and the evidence reviewed here suggests that different variables govern it: early response kinetics predict eligibility well but success poorly [11,17]. Fourth, the impact of newer agents is essentially unexamined. Asciminib and ponatinib are not represented in any included cohort, and the finding that imatinib and second-generation agents produced comparable success rates once discontinuation criteria were met cautions against assuming that deeper and faster responses will translate into higher TFR rates [21]. Fifth, residual disease measurement below the threshold of standard assays is not yet standardized. dPCR thresholds derived in one laboratory network require cross-platform harmonization before they can be embedded in guidelines, notwithstanding the encouraging external validation already achieved [18]; earlier work establishing residual disease and treatment duration as joint determinants of recurrence after imatinib cessation points to the same requirement [27]. Finally, geographical representation is uneven: the included studies originate almost entirely from high-income settings with mature molecular monitoring infrastructure, and the feasibility and safety of discontinuation where frequent high-sensitivity monitoring is unavailable is essentially unstudied, despite being the setting in which the economic argument for stopping therapy is strongest.

Clinical Implications

The following implications represent the authors' interpretation of the evidence rather than findings directly

demonstrated by the included studies. Candidate selection should be anchored on the duration and depth of DMR, with prior treatment duration as a supporting criterion, in line with current recommendations [2]. In the included studies, age, sex, the Sokal or ELTS score, and the generation of agent used were not consistently associated with TFR once discontinuation criteria were met. Where dPCR is available, a pre-cessation measurement may provide adjunctive prognostic information to inform the timing of an attempt but should not yet be used as a decision rule given the lack of cross-laboratory standardization. Monitoring should be intensive during the first six months, since both the hazard of recurrence and the informational value of each measurement are concentrated there, and retention of a deep response at three months can be regarded as a favorable prognostic development. Patients should be counseled that failure is common, that it is almost always reversible with appropriate monitoring and prompt re-treatment, and that a second attempt after a further period of consolidation remains possible, albeit with lower expected success. Above all, the decision should be framed as a shared one, weighing a quantifiable probability of durable freedom from therapy against the burdens of monitoring and the disappointment of relapse, in a disease where continuing treatment carries no threat to survival. The evidence assembled here permits that conversation to be conducted with greater precision than was possible five years ago, but it does not yet permit an individualized prediction of sufficient accuracy to be used as a decision rule.

Limitations

This review has several limitations. The five-year publication window yields an evidence base aligned with current monitoring standards but excludes the foundational discontinuation trials, which are therefore discussed only narratively, and the review was not prospectively registered. Most importantly, screening, data extraction, and risk-of-bias appraisal were performed by a single reviewer; although full-text decisions were repeated after a two-week interval, this does not substitute for independent duplicate assessment and may have introduced systematic selection, extraction, or appraisal bias. Initial screening was based on titles rather than titles and abstracts, which may have reduced sensitivity despite an inclusive approach, and two eligible reports were identified only during peer review. ROBINS-I, an intervention-oriented tool, was used in preference to a prognostic-factor tool such as QUIPS, which may have limited the appraisal of prognosis-specific biases. Exclusion of conference proceedings and grey literature may have introduced publication bias. Heterogeneity in predictor thresholds, outcome timepoints, definitions of DMR and molecular recurrence, assay sensitivity, and re-treatment triggers precluded meta-analysis, so the synthesis describes the direction and consistency of associations rather than pooled effects, and overlap between registry, trial-derived, and single-center reports cannot be fully excluded, so the total number of discontinuation attempts is approximate. Finally, all included studies were non-randomised; the strong intercorrelation of treatment duration, response duration, and response depth, together with clinician-directed selection of patients for discontinuation, means the findings identify reproducible associations for prospective testing rather than causal determinants of TFR.

Conclusions

Across 14 contemporary studies comprising more than 3,000 reported TKI discontinuation attempts in CP-CML, the duration of sustained DMR before cessation is the factor most consistently associated with TFR, supported by the duration of prior therapy. Residual disease below conventional detection limits, measured by dPCR before cessation, is a promising adjunctive marker that requires standardization, and the trajectory of BCR::ABL1 transcripts during the first three months off therapy is associated with outcome. Baseline prognostic scores, patient demographics, and the generation of agent used were not consistently associated with TFR once discontinuation criteria were met. Early on-treatment response kinetics, BCR::ABL1 transcript type, transcriptomic signatures, and immune effector profiles are biologically plausible but have not been consistently replicated and should currently be regarded as research parameters rather than selection criteria. Within the included studies, discontinuation conducted with structured, frequent molecular monitoring and prompt re-treatment was safe: virtually all patients who lost MMR regained it on re-treatment, and progression was not observed. Prospective external validation of composite models integrating molecular burden with immune competence is the principal priority for the field.

Appendices

Database	Search strategy	Records
MEDLINE (PubMed)	("leukemia, myelogenous, chronic, BCR-ABL positive"[MeSH] OR "chronic myeloid leukemia"[tiab] OR CML[tiab]) AND ("protein kinase inhibitors"[MeSH] OR imatinib[tiab] OR nilotinib[tiab] OR dasatinib[tiab] OR bosutinib[tiab] OR "tyrosine kinase inhibitor*"[tiab] OR TKI[tiab]) AND (discontinu*[tiab] OR cessation[tiab] OR stopping[tiab] OR withdraw*[tiab] OR "treatment-free remission"[tiab] OR TFR[tiab]) AND (predict*[tiab] OR prognostic[tiab] OR factor*[tiab] OR biomarker*[tiab] OR determinant*[tiab] OR relapse[tiab] OR recurrence[tiab])	296
Embase (Embase.com)	#1 'chronic myeloid leukemia'/exp OR 'chronic myeloid leukemia':ti,ab OR cml:ti,ab; #2 'protein tyrosine kinase inhibitor'/exp OR imatinib:ti,ab OR nilotinib:ti,ab OR dasatinib:ti,ab OR bosutinib:ti,ab OR 'tyrosine kinase inhibitor*':ti,ab OR tki:ti,ab; #3 discontinu*:ti,ab OR cessation:ti,ab OR stopping:ti,ab OR withdraw*:ti,ab OR 'treatment-free remission':ti,ab OR tfr:ti,ab; #4 predict*:ti,ab OR prognostic:ti,ab OR factor*:ti,ab OR biomarker*:ti,ab OR determinant*:ti,ab OR relapse:ti,ab OR recurrence:ti,ab; #5 #1 AND #2 AND #3 AND #4 AND (2021-2026)/py	102
Web of Science Core Collection	#1 TS=("chronic myeloid leukemia" OR "chronic myelogenous leukemia" OR CML); #2 TS=(imatinib OR nilotinib OR dasatinib OR bosutinib OR "tyrosine kinase inhibitor*" OR TKI); #3 TS=(discontinu* OR cessation OR stopping OR withdraw* OR "treatment-free remission" OR TFR); #4 TS=(predict* OR prognostic OR factor* OR biomarker* OR determinant* OR relapse OR recurrence); #5 #1 AND #2 AND #3 AND #4, with publication years 2021-2026	118
Cochrane Library	#1 MeSH descriptor: [Leukemia, Myelogenous, Chronic, BCR-ABL Positive] explode all trees OR ("chronic myeloid leukemia" OR CML):ti,ab,kw; #2 MeSH descriptor: [Protein Kinase Inhibitors] explode all trees OR (imatinib OR nilotinib OR dasatinib OR bosutinib OR "tyrosine kinase inhibitor*" OR TKI):ti,ab,kw; #3 (discontinu* OR cessation OR stopping OR withdraw* OR "treatment-free remission" OR TFR):ti,ab,kw; #4 (predict* OR prognostic OR factor* OR biomarker* OR determinant* OR relapse OR recurrence):ti,ab,kw; #5 #1 AND #2 AND #3 AND #4, with publication date 2021-2026	5

TABLE 5: Full search strategies

Additional Information

Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

Concept and design: Hanady ME M Osman, Nahla Mohammed Wahby, Sara Mohamed Abuelgasim Hassan, Islam Mustafa Hamid Mohammed, Samih Abdelmutalab Mohamed Abdalla

Drafting of the manuscript: Hanady ME M Osman, Ebtihal Abdalglil Mohammed Elhaj, Riham Elgaali Mohamed Abdelfattah, Islam Mustafa Hamid Mohammed, Samih Abdelmutalab Mohamed Abdalla, Alaa Khidir Attaelmanan Mahgoub

Critical review of the manuscript for important intellectual content: Hanady ME M Osman, Ebtihal Abdalglil Mohammed Elhaj, Riham Elgaali Mohamed Abdelfattah, Nahla Mohammed Wahby, Sara Mohamed Abuelgasim Hassan, Alaa Khidir Attaelmanan Mahgoub

Acquisition, analysis, or interpretation of data: Ebtihal Abdalglil Mohammed Elhaj, Riham Elgaali Mohamed Abdelfattah, Sara Mohamed Abuelgasim Hassan, Islam Mustafa Hamid Mohammed, Samih Abdelmutalab Mohamed Abdalla, Alaa Khidir Attaelmanan Mahgoub

Supervision: Nahla Mohammed Wahby

Disclosures

Conflicts of interest: In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work. **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

Acknowledgements

Riham Elgaali Mohamed Abdelfattah contributed equally to the work and should be considered a joint first author. All authors contributed equally to this work and met the ICMJE authorship criteria. Although we are based at different institutions, we share similar research interests and often work together on research projects, discussing ideas, planning studies, and analyzing the results as a team. ChatGPT (OpenAI, Inc., San Francisco, United States) was used to assist with language editing and manuscript preparation. The authors take full responsibility for the accuracy, integrity, and final content of the manuscript.

References

1. Bower H, Björkholm M, Dickman PW, Höglund M, Lambert PC, Andersson TM: Life expectancy of patients with chronic myeloid leukemia approaches the life expectancy of the general population. J Clin Oncol. 2016, 34:2851-7. 10.1200/JCO.2015.66.2866
2. Apperley JF, Milojkovic D, Cross NC, et al.: 2025 European LeukemiaNet recommendations for the management of chronic myeloid leukemia. Leukemia. 2025, 39:1797-813. 10.1038/s41375-025-02664-w
3. Mahon FX, Réa D, Guilhot J, et al.: Discontinuation of imatinib in patients with chronic myeloid leukaemia who have maintained complete molecular remission for at least 2 years: the prospective, multicentre Stop Imatinib (STIM) trial. Lancet Oncol. 2010, 11:1029-35. 10.1016/S1470-2045(10)70233-3
4. Saussele S, Richter J, Guilhot J, et al.: Discontinuation of tyrosine kinase inhibitor therapy in chronic myeloid leukaemia (EURO-SKI): a prespecified interim analysis of a prospective, multicentre, non-randomised, trial. Lancet Oncol. 2018, 19:747-57. 10.1016/S1470-2045(18)30192-X
5. Rousselot P, Charbonnier A, Cony-Makhoul P, et al.: Loss of major molecular response as a trigger for restarting tyrosine kinase inhibitor therapy in patients with chronic-phase chronic myelogenous leukemia who have stopped imatinib after durable undetectable disease. J Clin Oncol. 2014, 32:424-30. 10.1200/JCO.2012.48.5797
6. Rea D, Henry G, Khaznadar Z, et al.: Natural killer-cell counts are associated with molecular relapse-free survival after imatinib discontinuation in chronic myeloid leukemia: the IMMUNOSTIM study. Haematologica. 2017, 102:1368-77. 10.3324/haematol.2017.165001
7. Irani YD, Hughes A, Clarson J, et al.: Successful treatment-free remission in chronic myeloid leukaemia and its association with reduced immune suppressors and increased natural killer cells. Br J Haematol. 2020, 191:433-41. 10.1111/bjh.16718
8. Ilander M, Olsson-Strömberg U, Schlums H, et al.: Increased proportion of mature NK cells is associated with successful imatinib discontinuation in chronic myeloid leukemia. Leukemia. 2017, 31:1108-16. 10.1038/leu.2016.360
9. Page MJ, McKenzie JE, Bossuyt PM, et al.: The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021, 372:n71. 10.1136/bmj.n71
10. Methley AM, Campbell S, Chew-Graham C, McNally R, Cheraghi-Sohi S: PICO, PICOS and SPIDER: a comparison study of specificity and sensitivity in three search tools for qualitative systematic reviews. BMC Health Serv Res. 2014, 14:579. 10.1186/s12913-014-0579-0
11. Shanmuganathan N, Pagani IS, Ross DM, et al.: Early BCR-ABL1 kinetics are predictive of subsequent achievement of treatment-free remission in chronic myeloid leukemia. Blood. 2021, 137:1196-207. 10.1182/blood.2020005514
12. Atallah E, Schiffer CA, Radich JP, et al.: Assessment of outcomes after stopping tyrosine kinase inhibitors among patients with chronic myeloid leukemia: a nonrandomized clinical trial. JAMA Oncol. 2021, 7:42-50. 10.1001/jamaoncol.2020.5774
13. Dulucq S, Nicolini FE, Rea D, et al.: Kinetics of early and late molecular recurrences after first-line imatinib cessation in chronic myeloid leukemia: updated results from the STIM2 trial. Haematologica. 2022, 107:2859-69. 10.3324/haematol.2022.280811
14. Park H, Kim HJ, Sohn SK, et al.: Effect of BCR::ABL1 transcript type and droplet digital polymerase chain reaction on successful treatment-free remission in chronic myeloid leukemia patients who discontinued tyrosine kinase inhibitor. Ther Adv Hematol. 2023, 14:20406207231205637. 10.1177/20406207231205637
15. Marcé S, Méndez A, Xicoy B, et al.: e14a2 transcript favors treatment-free remission in chronic myeloid leukemia when associated with longer treatment with tyrosine kinase inhibitors and sustained deep molecular response. J Clin Med. 2024, 13:779. 10.3390/jcm13030779
16. Mahon FX, Pfirrmann M, Dulucq S, et al.: European Stop Tyrosine Kinase Inhibitor Trial (EURO-SKI) in chronic myeloid leukemia: final analysis and novel prognostic factors for treatment-free remission. J Clin Oncol. 2024, 42:1875-80. 10.1200/JCO.23.01647
17. Saugues S, Lambert C, Daguenet E, et al.: The initial molecular response predicts the deep molecular response but not treatment-free remission maintenance in a real-world chronic myeloid leukemia cohort. Haematologica. 2024, 109:2893-907. 10.3324/haematol.2023.284860
18. Kockerols C, Valk PJ, Janssen JJ, et al.: Prediction of sustained remission after tyrosine kinase inhibitor discontinuation with BCR::ABL1 digital PCR in chronic myeloid leukemia patients. Eur J Haematol. 2024, 113:606-13. 10.1111/ejh.14271
19. Ureshino H, Nakamura Y, Budirahardja Y, Nozawa I, Kidoguchi K, Kamachi K, Kimura S: Clinical features and T-cell repertoire of chronic myeloid leukemia patients who attempt discontinuation of tyrosine kinase inhibitors: the ISAC-TFR study. Cancer Med. 2025, 14:e71142. 10.1002/cam4.71142
20. Laganà A, Scalzulli E, Carmosino I, et al.: Treatment-free remission (TFR) in chronic phase-chronic myeloid leukemia (CP-CML): analysis of predictive factors and novel baseline scoring system to predict molecular relapse. Eur J Haematol. 2025, 115:367-79. 10.1111/ejh.70006
21. Takahashi N, Kimura S, Matsuki E, et al.: Five-year interim analysis of J-SKI: an observational study of TKI discontinuation in patients with CML in Japan. Int J Hematol. 2026, 124:96-103. 10.1007/s12185-026-04184-4
22. Jo T, Saburi Y, Masunari T, et al.: Activated memory cytotoxic T-lymphocytes and T-cell receptor Vβ clonality predict treatment-free remission after tyrosine kinase inhibitor discontinuation in chronic-phase chronic myeloid leukemia: a 1-year prospective immuno-monitoring study. Int J Mol Sci. 2026, 27:2713.

10.3390/ijms27062713

23. Radich JP, Chaturvedi S, Sadek I, Obourn V: Predictive biomarkers of sustained treatment-free remission in chronic myeloid leukemia: gene expression analyses from the ENESTfreedom and ENESTop studies. Haematologica. 2026, 111:3318-29. 10.3324/haematol.2025.300215
24. Alcazer V, Dulucq S, Mosnier I, et al.: Development and external validation of a transcriptome-based multivariable prediction model for treatment-free remission in chronic myeloid leukemia. J Clin Oncol. 2026, 44:1992-2005. 10.1200/JCO-25-02948
25. Etienne G, Guilhot J, Rea D, et al.: Long-term follow-up of the French Stop Imatinib (STIM1) study in patients with chronic myeloid leukemia. J Clin Oncol. 2017, 35:298-305. 10.1200/JCO.2016.68.2914
26. Flygt H, Söderlund S, Richter J, et al.: Treatment-free remission after a second TKI discontinuation attempt in patients with chronic myeloid leukemia re-treated with dasatinib - interim results from the DAstop2 trial. Leukemia. 2024, 38:781-7. 10.1038/s41375-024-02145-6
27. Nicolini FE, Dulucq S, Boureau L, et al.: Evaluation of residual disease and TKI duration are critical predictive factors for molecular recurrence after stopping imatinib first-line in chronic phase CML patients. Clin Cancer Res. 2019, 25:6606-13. 10.1158/1078-0432.CCR-18-3373