

Quantile Regression in Length-of-Stay Research: An Abstract-Based Methodological Scoping Review of PubMed Records

Bruna Alves

Universidade Estadual de Maringá

Josmar Mazucheli

jmazucheli@gmail.com

Universidade Estadual de Maringá

Research Article

Keywords: length of stay, LOS, hospital stay, quantile regression, median regression, health services research, scoping review, PubMed

Posted Date: September 24th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10157983/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Quantile Regression in Length-of-Stay Research: An Abstract-Based Methodological Scoping Review of PubMed Records

Bruna Alves and Josmar Mazucheli

June 11, 2026

Abstract

Background: Length of stay (LOS) is a central outcome in health services research, clinical epidemiology, perioperative medicine, emergency care, hospital management, and economic evaluation. Many LOS questions are explicitly distributional: capacity planning and cost analysis target the upper tail of long-stay patients, and the covariate associations governing typical stays need not coincide with those governing prolonged ones, so a model of the conditional mean alone cannot capture covariate effects that vary across the LOS distribution. Quantile regression, introduced by Koenker and Bassett, models conditional quantiles directly and is therefore well suited to characterizing this distributional heterogeneity in LOS-related outcomes. Previous LOS reviews have mapped broad prediction, risk-adjustment, machine-learning, emergency-department, and intervention literatures, but have not focused specifically on quantile regression as a distribution-sensitive strategy for LOS analysis.

Objective: To map and characterize how quantile-regression terminology is reported in association with LOS-related outcomes in PubMed records, with emphasis on the statistical-method terminology used (explicit quantile regression versus median regression) and on the abstract-level field in which the LOS outcome was positioned (title, abstract, keywords, or MeSH terms).

Methods: We conducted an abstract-based methodological scoping review of PubMed records. The PubMed query combined LOS-related terms in MeSH and title/abstract fields with *quantile regression*, *regression quantile(s)*, or *median regression* in title/abstract fields. The search was conducted on 10 June 2026. The exported dataset contained 210 unique PubMed records, all with abstracts. Rule-based post-processing classified records according to statistical-method terminology, source of LOS evidence, and analytic-corpus assignment. Records with explicit quantile-regression terminology and LOS-related terminology in the title or abstract were defined as the main analytic corpus. The expanded sensitivity corpus comprised all records mentioning median regression without explicit quantile-regression terminology, together with explicit quantile-regression records whose LOS evidence was confined to keywords or MeSH terms (i.e., with no LOS evidence in the title or abstract). All numerical summaries, percentages, methodological classifications, and corpus assignments were generated from the exported PubMed files

using custom R scripts. A complementary PubMed title search for LOS records containing the word *review* was used only to contextualize the methodological gap and was not merged with the main corpus.

Results: The search retrieved 210 unique records published from 2000 to 2026. All records contained abstracts, 206 (98.1%) contained a DOI, 83 (39.5%) contained a PubMed Central identifier, 144 (68.6%) contained keywords, and 173 (82.4%) contained MeSH terms. Methodological classification identified 168 records (80.0%) with explicit quantile-regression terminology and 42 records (20.0%) mentioning median regression without explicit quantile-regression terminology. Using the hierarchy title, abstract, keywords, and MeSH terms, LOS evidence was assigned to the title for 39 records (18.6%), the abstract for 135 records (64.3%), keywords for 5 records (2.4%), and MeSH terms for 31 records (14.8%). The main analytic corpus comprised 133 records (63.3%) with explicit quantile-regression terminology and LOS in the title or abstract. The expanded sensitivity corpus comprised 77 records: 41 median-regression-only records with LOS in the title or abstract, 30 explicit quantile-regression records with LOS evidence assigned to MeSH terms, 5 explicit quantile-regression records with LOS keyword evidence but no LOS title/abstract evidence, and 1 median-regression-only record with LOS evidence restricted to keywords or MeSH terms. A complementary PubMed title search for LOS records containing the word *review* yielded 200 analysable records (219 retrieved in the PubMed interface), none of which mentioned *quantile* or *quantile regression* in the title or abstract.

Conclusions: This review provides an abstract-level methodological map of explicit quantile-regression use in LOS-related research. The classification also shows that retrieval driven by MeSH indexing and the inclusion of median-regression terminology materially inflate the apparent corpus, so records mentioning median regression without explicit quantile-regression terminology, and records whose LOS evidence is limited to keywords or MeSH terms, should be treated as a sensitivity set rather than as direct quantile-regression applications to LOS. The complementary review search suggests that quantile regression remains underrepresented as an explicit methodological focus in LOS review abstracts, despite an extensive review literature on LOS prediction, determinants, interventions, and health-service outcomes. Because the search was restricted to PubMed, this characterization applies to the PubMed-indexed literature and may understate quantile-regression activity reported in statistics and econometrics venues that PubMed indexes less comprehensively.

Keywords: length of stay; LOS; hospital stay; quantile regression; median regression; health services research; scoping review; PubMed.

1 Introduction

Hospital length of stay (LOS) is one of the most frequently used outcomes in health services research, hospital epidemiology, perioperative medicine, emergency care, critical care, neonatology, and health economics. It is used as a proxy for resource utilization, bed occupancy, organizational efficiency, severity of illness, discharge readiness, complications, and direct or indirect costs of care. Across the wider LOS literature, the outcome appears in studies of all-admission prediction,

emergency-department flow, postoperative recovery, intensive-care use, neonatal care, and economic burden.

LOS is also statistically challenging. Its distribution is usually right-skewed, bounded below by zero or one day depending on the definition, heterogeneous across diagnoses and procedures, and influenced by a minority of long-stay observations. In many clinical settings, an exposure may have little effect among short-stay patients but a substantial effect among those in the upper tail of the LOS distribution. This feature is especially relevant for hospital planning, because long-stay patients account disproportionately for bed occupancy and resource use.

Traditional approaches to LOS include linear regression, generalized linear models, survival models, Poisson or negative binomial regression, and logistic regression after dichotomizing LOS into prolonged versus non-prolonged stay. Each approach targets a different feature of the outcome. Mean regression describes conditional averages; survival analysis focuses on time to discharge or discharge hazards; count models treat LOS as a discrete non-negative outcome; and logistic regression reduces LOS to categories. These approaches can be useful, but they may obscure clinically meaningful heterogeneity when covariate effects differ between short, typical, and prolonged hospitalizations.

Several reviews have already demonstrated the importance and methodological heterogeneity of the broader LOS-prediction literature. Lu et al. reviewed risk-adjustment models for hospital LOS, emphasizing administrative-data applications in performance measurement and benchmarking [18]. Lequertier et al. reviewed LOS prediction methods based on hospital data and validation protocols [19]. Stone et al. reviewed hospital LOS prediction and proposed a unified framework for comparing approaches across hospital environments [20]. More recent reviews addressed prolonged-LOS prediction tools for general admissions and general-medicine populations [21] and emergency-department LOS prediction models [22]. Other reviews mapped the determinants of LOS across settings [23] and in the intensive care unit [24]. These reviews show that LOS research is clinically and operationally important, but they do not specifically foreground quantile regression as a distinct modeling strategy for LOS-related outcomes.

Quantile regression, introduced by Koenker and Bassett [1] and subsequently developed as a general framework for conditional quantile functions [2, 3], directly models conditional quantiles rather than only the conditional mean. It is therefore attractive for LOS because hospital managers and clinicians are often concerned not only with average stay but also with delayed discharge, long-stay patients, high-cost episodes, and upper-tail resource use. Median regression is the 0.5-quantile special case of quantile regression; however, in the context of an abstract-based review, records mentioning only median regression should be treated cautiously, because the abstract may not make clear whether the broader quantile-regression framework was used or whether only the median was modeled.

This article maps the quantile-regression literature on LOS using an explicit hierarchy of evidence. A PubMed search identifies a broad set of records related to LOS and quantile- or median-regression terminology. Within this set, the main analytic corpus is restricted to records with

explicit quantile-regression terminology and LOS evidence in the title or abstract; records mentioning only median regression, and records in which LOS appears only in MeSH terms or keywords, are retained as a sensitivity corpus. This structure preserves the focus on quantile regression for LOS while acknowledging the broader retrieval behavior of PubMed searches. To contextualize the methodological gap, a complementary PubMed search for records whose titles contain the word *review* was examined separately; it is reported in the Results and was not merged with the analytic corpus.

2 Methodological background: quantile regression for LOS

2.1 Conditional quantiles

Let Y denote hospital length of stay (LOS) and let $X \in \mathbb{R}^p$ be a vector of covariates, with $\{(Y_i, X_i)\}_{i=1}^n$ an observed sample of n independent hospitalization episodes. Depending on the clinical context, Y may represent total hospital LOS, intensive-care-unit LOS, emergency-department LOS, or another LOS-related duration. For a quantile level $\tau \in (0, 1)$, the conditional τ -th quantile of Y given $X = x$ is

$$Q_{Y|X}(\tau | x) = \inf\{y \in \mathbb{R} : F_{Y|X}(y | x) \geq \tau\},$$

where $F_{Y|X}(\cdot | x)$ is the conditional distribution function of LOS given $X = x$. Thus $Q_{Y|X}(\tau | x)$ is the value below which a proportion τ of the conditional LOS distribution lies.

The linear quantile-regression model specifies this conditional quantile as a linear function of the covariates,

$$Q_{Y|X}(\tau | x) = x^\top \beta(\tau), \tag{1}$$

where $\beta(\tau) \in \mathbb{R}^p$ is a vector of quantile-specific coefficients. Letting $\beta(\tau)$ depend on τ allows covariate effects to vary across the conditional LOS distribution; this is the essential difference from ordinary least squares, which targets the single conditional mean $\mathbb{E}(Y | x)$.

For the observed sample $\{(y_i, x_i)\}_{i=1}^n$ and a fixed τ , the quantile-regression estimator solves the convex program

$$\hat{\beta}(\tau) = \arg \min_{\beta \in \mathbb{R}^p} \sum_{i=1}^n \rho_\tau(y_i - x_i^\top \beta), \tag{2}$$

where

$$\rho_\tau(u) = u\{\tau - I(u < 0)\} \tag{3}$$

is the asymmetric *check* loss and $I(\cdot)$ the indicator function. The check loss weights positive and negative residuals asymmetrically, thereby targeting the τ -th conditional quantile; at $\tau = 0.5$ it reduces to $\frac{1}{2}|u|$, so (2) becomes least-absolute-deviations (median) regression. The optimization is a linear program and is available in standard software, including the **quantreg** package in R [13], the **QUANTREG** procedure in SAS [53], and the **qreg** command in Stata [52]. Introductory treatments aimed at applied audiences are also available [4]. Throughout, Y denotes the random response and

y_i its realized value for episode i .

2.2 Robustness, leverage, and the upper tail

Two properties of (2) are frequently conflated and should be kept distinct. First, for a *fixed* quantile level the check loss is piecewise linear rather than quadratic, so the fitted quantile is less sensitive to extreme values *of the response* than the least-squares fit is to outliers in Y ; the median fit in particular has bounded influence in the response direction. This robustness is confined to the y -direction, however: like ordinary least squares, quantile regression is *not* robust to high-leverage observations in covariate space, and a single aberrant x_i can exert substantial influence on $\hat{\beta}(\tau)$.

Second, and separately, modeling high quantile levels (e.g. $\tau = 0.90$ or 0.95) is a deliberate choice to describe the *upper tail* of the conditional LOS distribution, not a robustness property. Such models are by construction informative about long-stay episodes and must be interpreted as statements about the specific conditional quantile being estimated, rather than as outlier-resistant summaries of the whole distribution. In LOS research the distinction matters: long-stay patients drive bed occupancy and cost, and upper-tail quantile models address them directly.

2.3 Interpretation of coefficients

Under the linear specification (1), the coefficient $\beta_j(\tau)$ is the partial derivative $\partial Q_{Y|X}(\tau | x)/\partial x_j$: it measures the association between the j -th covariate and the τ -th *conditional* quantile of LOS, holding the remaining covariates fixed. It is not, in general, an effect on the *unconditional* (population) LOS quantile unless additional assumptions or a different estimand are invoked. Because $\beta(\tau)$ varies with τ , a covariate may be weakly associated with the conditional median LOS yet strongly associated with the conditional 0.90 quantile—a clinically meaningful pattern when a risk factor or intervention acts mainly on delayed discharge or long-stay patients.

2.4 Conditional versus unconditional quantile effects

The coefficients in (1)–(2) are *conditional*-quantile effects and must be distinguished from *unconditional* (marginal) quantile effects, which describe how a shift in the covariate distribution would move the unconditional LOS quantiles. Unconditional quantile methods—most notably recentered-influence-function (RIF) regression [8]—answer different, policy-oriented questions and should not be conflated with the conditional quantile regression of [1, 3]. In LOS applications, conflating the two can lead to misinterpreting a within-stratum gradient as a population-level shift in the LOS distribution.

2.5 Censored length of stay and competing risks

LOS is a duration outcome and is therefore frequently subject to incomplete observation, which classical quantile regression does not accommodate. Three situations are common. (i) *Administrative right-censoring*: a patient still hospitalized at the end of follow-up has a LOS known only to

exceed the observed value. (ii) *In-hospital death*: depending on the estimand, death either censors the time to live discharge or, more appropriately, acts as a *competing event* that precludes discharge alive. (iii) *Transfer* to another facility, which similarly truncates the observed episode. Because long stays are precisely those most likely to be censored, naive quantile regression on the observed LOS—treating censored episodes as complete, or discarding them—can bias $\widehat{\beta}(\tau)$, and the bias is typically most severe at the upper quantiles that are of greatest clinical and operational interest.

Several quantile-regression estimators address censoring. When the censoring point is fixed and known, the censored (Tobit-type) quantile-regression estimator of [9] applies. For random right-censoring, the recursively reweighted estimator of [10] and the martingale-based estimating-equation approach of [11] extend conditional quantile regression to survival-type data, playing a role analogous to that of the Kaplan–Meier and Nelson–Aalen estimators in mean- or hazard-based survival analysis. When in-hospital death precludes live discharge, a competing-risks formulation—rather than ordinary right-censoring—is required to keep the discharge-quantile estimand well defined; otherwise the modeled quantity confounds discharge with death. These distinctions are central to LOS, because the choice between a censoring and a competing-risks framing materially changes what an upper-tail LOS quantile represents.

2.6 Discrete length of stay, ties, and quantile crossing

Classical linear quantile regression was developed for continuous responses, whereas LOS is usually recorded as an integer number of days, producing many ties and flat regions in the empirical conditional distribution; the conditional quantile function is then a step function, which affects estimation and the asymptotic theory underlying standard inference. Applied studies often proceed with standard quantile regression because the fitted quantiles remain interpretable at clinically relevant levels, but this approximation should be acknowledged. Dedicated alternatives for discrete or count responses include the jittering approach of [6], which smooths the discrete response to recover a continuous quantile process; moment-based conditional-quantile estimators [7]; and distributional or generative models for counts, including Bayesian formulations [12].

A further issue arises when several quantiles are estimated by separate fits: the predicted quantile curves may *cross*, violating the monotonicity of $\tau \mapsto Q_{Y|X}(\tau | x)$. Monotonicity-constrained estimation or post-estimation rearrangement [5] restores a valid, non-crossing set of conditional quantiles.

2.7 Inference, reporting, and the limits of abstract-level mapping

Inference for quantile regression typically relies on asymptotic sandwich (Huber-type) covariance estimators, rank-score methods, or the bootstrap [3]. With clustered data—patients nested within hospitals, or repeated admissions for the same patient—cluster-robust or multilevel variants are required, since the independence assumption underlying naive standard errors fails. Applied LOS abstracts, however, seldom state which inferential approach was used.

For methodological transparency, LOS studies using quantile regression should report: the quantile levels modeled and the rationale for their selection; the model specification and covariates; whether the target estimand is conditional or unconditional; the treatment of discreteness, ties, and quantile crossing; the handling of censoring, in-hospital death, transfer, and competing discharge destinations; the uncertainty-quantification method and any clustering; and the software used. These same items delimit what the present review can establish. Because the analysis is abstract-based, most of these dimensions—inferential method, censoring treatment, estimand, quantile crossing—are not recoverable from titles, abstracts, keywords, and MeSH terms alone. This section therefore defines the methodological grid that a full-text review would need to extract: the abstract-level corpus mapped here characterizes how quantile regression is *named and positioned* in the LOS literature, not how it was *implemented*, and the contrast between the two is itself part of the rationale for this review.

3 Methods

3.1 Design

This was an abstract-based methodological scoping review of PubMed records. The unit of analysis was the PubMed record, not the full-text article. Quantitative summaries therefore refer to records and abstracts, not to fully adjudicated studies. The review was informed by PRISMA-ScR principles [15] and by the PRISMA 2020 reporting guidance [16], but it was not a conventional full-text systematic review. Full-text retrieval, formal risk-of-bias assessment, and methodological quality scoring were not performed.

3.2 Data source and search strategy

The search was conducted in PubMed [17] on 10 June 2026. It combined a LOS block with a statistical-method block. The LOS block used the MeSH descriptor *Length of Stay* and title/abstract variants of common length-of-stay terminology. The statistical-method block required *quantile regression*, *regression quantile(s)*, or *median regression* terminology in title/abstract fields.

```
("Length of Stay"[MeSH Terms]
OR "length of stay"[Title/Abstract]
OR "lengths of stay"[Title/Abstract]
OR "length of hospital stay"[Title/Abstract])
AND
("quantile regression"[Title/Abstract]
OR "regression quantile"[Title/Abstract]
OR "regression quantiles"[Title/Abstract]
OR "median regression"[Title/Abstract])
```

The acronym LOS was not used as a standalone title/abstract search term because of its limited specificity outside the length-of-stay context. Instead, retrieval relied on the MeSH descriptor *Length of Stay* and on spelled-out title/abstract expressions. Therefore, the search could have missed records that used only the acronym LOS in the title or abstract and were not indexed with the *Length of Stay* MeSH descriptor. This conservative choice was treated as a limitation.

The inclusion of median regression was motivated statistically because median regression is the ($\tau = 0.5$) special case of quantile regression. Nevertheless, because abstracts mentioning only median regression may not describe the broader quantile-regression framework, these records were separated during post-processing and were not automatically included in the main analytic corpus. This decision preserved the main focus on explicit quantile regression while allowing sensitivity assessment of median-regression-only records.

3.3 Complementary contextual search of LOS reviews

A complementary PubMed search was conducted using the query (`length of stay[Title]`) AND (`review[Title]`). This search was not used to define or expand the quantile-regression corpus. Its sole purpose was contextual: to assess whether quantile regression was explicitly visible in PubMed-indexed LOS review abstracts. The PubMed interface count recorded by the investigators was 219 records; the export available for analysis contained 200 unique records. The complementary search retrieves heterogeneous records because `review[Title]` captures systematic reviews, scoping reviews, review protocols, narrative reviews, and titles containing expressions such as retrospective chart review or medical-record review.

3.4 Data extraction

The following fields were extracted from PubMed XML output: PMID, PMCID, DOI, title, journal, journal abbreviation, year, publication date, volume, issue, pages, authors, language, publication types, keywords, MeSH terms, and abstract. The dataset contained 210 records and 27 columns after extraction. All 210 records contained abstracts.

3.5 Post-processing and corpus definition

A rule-based post-processing procedure was applied to the exported records. The procedure classified records according to three dimensions:

- (i) statistical-method terminology: explicit quantile-regression terminology (the phrases *quantile regression*, *regression quantile*, or *regression quantiles*, including annotated forms such as *quantile (median) regression*) versus median-regression-only terminology (*median regression* in the absence of any explicit quantile-regression phrase);
- (ii) source of LOS evidence, assigned to a single field using the priority order title > abstract > keywords > MeSH terms (author-origin fields before indexer-assigned MeSH);

(iii) screening category: main analytic corpus versus sensitivity corpus.

The main analytic corpus was defined as records satisfying both conditions: explicit quantile-regression terminology and LOS-related terminology in the title or abstract. Records were retained in the expanded sensitivity corpus if they met one of the following conditions: median-regression-only terminology (regardless of where LOS appeared); or explicit quantile-regression terminology with LOS evidence confined to keywords or MeSH terms. Text matching used the same spelled-out LOS expressions as the title/abstract block of the search query, after lowercasing and normalizing hyphens and dash variants. This ensured consistency between retrieval and post-processing while preserving a conservative approach that did not rely on the standalone acronym LOS.

3.6 Synthesis and software

Results were summarized using counts and percentages. The primary synthesis focused on the size and composition of the main analytic corpus and the sensitivity corpus. No meta-analysis, clinical effect-size pooling, or full-text risk-of-bias assessment was performed. Methodological interpretation was restricted to information available in titles, abstracts, keywords, and MeSH terms. All record retrieval, rule-based classification, and numerical summaries were carried out in base R [14] using the packages `rentrez`, `xml2`, `tibble`, `dplyr`, `stringr`, and `readr`. The classification script applies the rules above and regenerates every count reported below, so that the tables and text derive from a single documented computation.

4 Results

4.1 Record characteristics

The PubMed query retrieved 210 unique records. All records contained an abstract. DOI information was available for 206 records (98.1%), PMCID information for 83 (39.5%), keywords for 144 (68.6%), and MeSH terms for 173 (82.4%). Publication years ranged from 2000 to 2026. That only 173 of 210 records (82.4%) carried MeSH terms reflects that a non-negligible fraction of retrieved records were not (yet) MEDLINE-indexed, which is why the corpus is described as a set of PubMed, rather than MEDLINE, records. The temporal distribution shows a marked concentration in recent years: 144 of the 210 records (68.6%) were published from 2020 onward, and the main corpus follows the same trend. Counts for 2026 are partial, covering records indexed up to the search date (10 June 2026).

4.2 Statistical-method terminology

Among the 210 records, 168 (80.0%) contained explicit quantile-regression terminology. The remaining 42 records (20.0%) mentioned median regression without explicit quantile-regression terminology. These 42 records were retained for sensitivity assessment because median regression

Table 1: General characteristics of the PubMed record set.

Characteristic	Count	Percentage
Unique PubMed records	210	100.0
Records with abstract	210	100.0
Records with DOI	206	98.1
Records with PubMed Central identifier	83	39.5
Records with keywords	144	68.6
Records with MeSH terms	173	82.4
Publication period	2000–2026	–

Table 2: Distribution of records by publication year.

Year	All records	Main corpus	Cumulative
2000–2009	8	4	8
2010	5	4	13
2011	3	0	16
2012	3	1	19
2013	8	5	27
2014	5	2	32
2015	8	2	40
2016	6	4	46
2017	6	4	52
2018	6	3	58
2019	8	5	66
2020	13	8	79
2021	19	14	98
2022	14	12	112
2023	26	17	138
2024	28	18	166
2025	28	20	194
2026	16	10	210

Table 3: Record flow and corpus definition.

Flow item	Records
Records retrieved by the PubMed query	210
Duplicate PMIDs identified in the export	0
Records classified using title, abstract, keywords, and MeSH terms	210
Main analytic corpus: explicit quantile regression plus LOS in title/abstract	133
Expanded sensitivity corpus	77
Full-text eligibility assessment	Not performed

is statistically related to quantile regression, but they were not automatically interpreted as explicit quantile-regression applications.

Table 4: Classification by statistical-method terminology.

Method category	Records	Percentage
Explicit quantile-regression terminology	168	80.0
Median-regression-only terminology	42	20.0

4.3 Source of LOS evidence

Using the priority order title > abstract > keywords > MeSH terms, LOS evidence was assigned to the title for 39 records (18.6%), to the abstract for 135 records (64.3%), to keywords for 5 records (2.4%), and to MeSH terms for 31 records (14.8%). Thus most records had LOS evidence in the title or abstract (174 records), but a non-negligible fraction was retrieved because of indexing or keywords rather than explicit title/abstract wording.

Table 5: Source of LOS evidence (highest-priority field).

LOS evidence (highest field)	Records	Percentage
Title	39	18.6
Abstract (not title)	135	64.3
Keywords (not title/abstract)	5	2.4
MeSH terms (not title/abstract/keywords)	31	14.8

4.4 Main corpus and sensitivity corpus

The main analytic corpus comprised 133 records (63.3% of all retrieved records) with explicit quantile-regression terminology and LOS-related terminology in the title or abstract. This group provides the most defensible abstract-level basis for describing quantile-regression applications in LOS-related research.

The sensitivity corpus comprised 77 records (36.7%). This included 41 median-regression-only records with LOS in the title or abstract, 30 explicit quantile-regression records with LOS evidence assigned to MeSH terms, 5 explicit quantile-regression records with LOS in keywords only, and 1 median-regression-only record with LOS evidence restricted to keywords or MeSH terms.

Table 6: Screening categories produced by rule-based post-processing.

Screening category	Records	Percentage
Core: explicit quantile regression with LOS in title or abstract	133	63.3
Sensitivity: median regression with LOS in title or abstract	41	19.5
Sensitivity: explicit quantile regression, LOS via MeSH only	30	14.3
Sensitivity: explicit quantile regression, LOS via keywords only	5	2.4
Sensitivity: median regression, LOS via keywords/MeSH only	1	0.5

4.5 Illustrative applications in the main corpus

The main corpus spans a broad range of clinical and economic settings, which is consistent with quantile regression being used wherever distributional heterogeneity in LOS or related durations is of interest. Methodological primers written for clinical audiences introduce quantile regression and median-based comparisons as tools for skewed outcomes [25, 29, 30]. In emergency and acute care, quantile regression has been used to model waiting time and emergency-department flow [26, 27, 40]. In surgery and perioperative medicine, applications include colorectal cancer surgery [31], rib-fracture stabilization [38], hip fracture [35], paediatric surgical procedures [41], and burns [42]. Hospitalization cost and spending analyses frequently adopt quantile regression to characterize the upper tail of expenditure and stay [34, 36, 37, 44, 48]. Further applications cover frailty and older surgical patients [47, 33], neonatal and preterm care [39, 45], cardiovascular care [43, 46], hospice stay [28], formal home care and inpatient stay [32], cross-country health-system mortality [49], and analyses published in languages other than English [50, 51]. These records are intended as illustrative of the corpus rather than as an exhaustive enumeration.

4.6 Complementary LOS-review context

The complementary PubMed search for (length of stay[Title]) AND (review[Title]) returned 219 records in the PubMed interface according to the investigators’ search log. The export available for analysis contained 200 unique records with 200 unique PMIDs. None of these titles or abstracts contained the term *quantile* or *quantile regression*. This result supports a circumscribed interpretation: quantile regression is not absent from empirical LOS-related research, but it appears underrepresented as an explicit methodological focus in PubMed-indexed LOS review abstracts.

Table 7: Contextual summary of the complementary PubMed search for LOS review records.

Item	Records
PubMed interface count recorded by investigators	219
Export analyzed	200
Unique PMIDs in export	200
Records mentioning <i>quantile</i> in title/abstract	0
Records mentioning <i>quantile regression</i> in title/abstract	0

5 Discussion

This review maps quantile regression for LOS while incorporating the broader retrieval behavior of a PubMed query. The most important methodological decision was to separate the record set into a main analytic corpus and a sensitivity corpus. This separation is preferable to treating all 210 records as equivalent. It recognizes that explicit quantile-regression terminology and LOS title/abstract evidence provide the strongest abstract-level support for inclusion, whereas

median-regression-only records and MeSH-only or keyword-only LOS records require more cautious interpretation.

5.1 Interpretation of the main corpus

The main analytic corpus contained 133 records with explicit quantile-regression terminology and LOS-related terminology in the title or abstract. This group should be treated as the primary basis for claims about quantile regression in LOS-related literature. Because both the method and the LOS concept are visible in title/abstract fields, these records are less likely to represent incidental indexing or peripheral use of LOS. The corpus supports the premise that quantile regression has been used in LOS-related research, especially when investigators are interested in distributional heterogeneity, long-stay patients, conditional medians, upper-tail outcomes, or resource-use gradients. The method is particularly relevant when the scientific question concerns whether covariate associations differ across the LOS distribution rather than only at the mean.

5.2 Role of median-regression-only records

The query retrieved 42 records that mentioned median regression only. These records should not be discarded automatically because median regression is the $\tau = 0.5$ case of quantile regression. However, they should not be merged uncritically with explicit quantile-regression records. In an abstract-based review, median regression may refer to a single robust location model rather than to a broader quantile-regression analysis across multiple quantiles. For this reason, the 42 median-regression-only records are best interpreted as a sensitivity set. They can be used to evaluate how much the corpus expands when the search includes the most common special case of quantile regression, but they should be reported separately from records explicitly mentioning quantile regression.

5.3 Role of MeSH-only and keyword-only LOS retrieval

Thirty explicit quantile-regression records were retrieved through LOS MeSH terms only, and five through LOS keywords only. These records are useful because MeSH indexing can capture articles related to LOS even when the title or abstract does not contain the exact phrase. At the same time, MeSH-only retrieval weakens direct interpretability at the abstract level. A record indexed with *Length of Stay* may use LOS as a secondary endpoint, an example, a covariate, or a contextual variable. Therefore, MeSH-only and keyword-only records should be retained for sensitivity analysis or manual verification, but not automatically counted as direct LOS modeling applications.

5.4 Relationship with previous LOS reviews

The complementary LOS-review search reinforces the article’s rationale without altering the main corpus. The review-title export contained 200 records and no mention of quantile or quantile

regression in titles or abstracts. This is consistent with the observation that previous LOS reviews have focused on prediction models, risk adjustment, interventions, emergency-department flow, ICU outcomes, machine learning, and clinical determinants [18, 19, 20, 21, 22, 23, 24]. The present review complements this literature by focusing on a narrower statistical question: how quantile-regression terminology appears in the PubMed-indexed LOS literature, and how to define a defensible abstract-level corpus for that purpose.

5.5 Implications for LOS research

For LOS applications, quantile regression is most valuable when the target is heterogeneity across the LOS distribution. Researchers should avoid presenting quantile regression merely as a robust substitute for mean regression. Its main contribution is that it allows different covariate associations at different points of the conditional LOS distribution. This is particularly relevant for upper-tail quantiles, where long-stay patients may drive bed pressure, delayed discharge, and high costs. Future LOS studies using quantile regression should report which quantiles were modeled, why those quantiles were selected, whether the focus was conditional or marginal, how discrete LOS and ties were handled, how uncertainty was quantified, whether clustering or repeated admissions were considered, and how death, transfer, discharge destination, censoring, and competing outcomes were addressed. Abstracts should be more explicit about these features because abstract-level indexing and review depend heavily on title and abstract terminology.

5.6 Strengths and limitations

A strength of this review is its transparent corpus definition. Rather than relying only on a PubMed count, the review separates explicit quantile-regression records from median-regression-only records and distinguishes LOS evidence in titles/abstracts from MeSH-only or keyword-only evidence; the classification rules are documented and the reported counts are regenerated from a single script. The main limitation is that the review remains abstract-based. No full-text eligibility assessment was performed, and the classification cannot determine with certainty how quantile regression was implemented in each article. The analysis also depends on PubMed indexing, title/abstract wording, and the exported fields available to the investigators; in particular, restricting the search to PubMed may understate quantile-regression activity reported in statistics and econometrics venues that PubMed indexes less comprehensively. The main corpus of 133 records should therefore be interpreted as a defensible abstract-level corpus, not as a final set of full-text-confirmed studies. The complementary review search also has limitations: the term `review[Title]` retrieves heterogeneous records, including systematic reviews, protocols, narrative reviews, scoping reviews, and chart or medical-record reviews.

6 Conclusion

The PubMed search retrieved 210 LOS-related records mentioning quantile-regression, regression-quantile, or median-regression terminology. A rule-based post-processing strategy identified 133 records as the main analytic corpus because they contained explicit quantile-regression terminology and LOS evidence in the title or abstract. An additional 77 records were retained as a sensitivity corpus, mainly because they mentioned median regression only or because LOS appeared only in MeSH terms or keywords. This structure provides a more rigorous basis for an abstract-based methodological scoping review: quantile regression is present in the LOS literature, but the most defensible synthesis should distinguish explicit quantile-regression applications from median-regression-only and indexing-driven records. The complementary LOS-review search suggests that quantile regression remains underrepresented as an explicit methodological focus in LOS review abstracts. Because the search was restricted to PubMed, this conclusion applies to the PubMed-indexed literature and motivates a future multi-database, full-text review.

Data and code availability

The R scripts used for record extraction and rule-based classification, together with the exported PubMed records and the derived classification files, are available from the author on reasonable request. All counts reported in this article can be regenerated from these scripts. No individual patient-level data were used.

Funding

No specific funding was received for this work.

Conflicts of interest

The authors declare no conflicts of interest.

References

- [1] Koenker R, Bassett G. Regression quantiles. *Econometrica*. 1978;46(1):33–50. doi:10.2307/1913643.
- [2] Koenker R, Hallock KF. Quantile regression. *Journal of Economic Perspectives*. 2001;15(4):143–156. doi:10.1257/jep.15.4.143.
- [3] Koenker R. *Quantile Regression*. Cambridge: Cambridge University Press; 2005. doi:10.1017/CBO9780511754098.

- [4] Hao L, Naiman DQ. Quantile Regression. Thousand Oaks, CA: Sage Publications; 2007. doi:10.4135/9781412985550.
- [5] Chernozhukov V, Fernandez-Val I, Galichon A. Quantile and probability curves without crossing. *Econometrica*. 2010;78(3):1093–1125. doi:10.3982/ECTA7880.
- [6] Machado JAF, Santos Silva JMC. Quantiles for counts. *Journal of the American Statistical Association*. 2005;100(472):1226–1237. doi:10.1198/016214505000000330.
- [7] Machado JAF, Santos Silva JMC. Quantiles via moments. *Journal of Econometrics*. 2019;213(1):145–173. doi:10.1016/j.jeconom.2019.04.009.
- [8] Firpo S, Fortin NM, Lemieux T. Unconditional quantile regressions. *Econometrica*. 2009;77(3):953–973. doi:10.3982/ECTA6822.
- [9] Powell JL. Censored regression quantiles. *Journal of Econometrics*. 1986;32(1):143–155. doi:10.1016/0304-4076(86)90016-3.
- [10] Portnoy S. Censored regression quantiles. *Journal of the American Statistical Association*. 2003;98(464):1001–1012. doi:10.1198/016214503000000954.
- [11] Peng L, Huang Y. Survival analysis with quantile regression models. *Journal of the American Statistical Association*. 2008;103(482):637–649. doi:10.1198/016214508000000355.
- [12] Yamauchi Y, Kobayashi G, Sugawara S. Flexible Bayesian quantile regression for counts via generative modeling. *Biometrics*. 2025;81(4):ujaf152. doi:10.1093/biomtc/ujaf152.
- [13] Koenker R. quantreg: Quantile Regression. R package version 5.97.1. Vienna: R Foundation for Statistical Computing; 2023. Available from: <https://CRAN.R-project.org/package=quantreg>.
- [14] R Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria: R Foundation for Statistical Computing, 2024. Available from: <https://www.R-project.org/>.
- [15] Tricco AC, Lillie E, Zarin W, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. *Annals of Internal Medicine*. 2018;169(7):467–473. doi:10.7326/M18-0850.
- [16] Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021;372:n71. doi:10.1136/bmj.n71.
- [17] National Library of Medicine. PubMed. Bethesda (MD): National Center for Biotechnology Information, U.S. National Library of Medicine. Available from: <https://pubmed.ncbi.nlm.nih.gov/>.

- [18] Lu M, Sajobi TT, Lucyk K, Lorenzetti DL, Quan H. Systematic review of risk adjustment models of hospital length of stay (LOS). *Medical Care*. 2015;53(4):355–365. doi:10.1097/MLR.0000000000000317.
- [19] Lequertier V, Wang T, Fondrevelle J, Augusto V, Duclos A. Hospital length of stay prediction methods: a systematic review. *Medical Care*. 2021;59(10):929–938. doi:10.1097/MLR.0000000000001596.
- [20] Stone K, Zwiggelaar R, Jones P, Mac Parthaláin N. A systematic review of the prediction of hospital length of stay: towards a unified framework. *PLOS Digital Health*. 2022;1(4):e0000017. doi:10.1371/journal.pdig.0000017.
- [21] Gokhale S, Taylor D, Gill J, Hu Y, Zeps N, Lequertier V, Prado L, Teede H, Enticott J. Hospital length of stay prediction tools for all hospital admissions and general medicine populations: systematic review and meta-analysis. *Frontiers in Medicine*. 2023;10:1192969. doi:10.3389/fmed.2023.1192969.
- [22] Farimani RM, Karim H, Atashi A, Tohidinezhad F, Bahaadini K, Abu-Hanna A, Eslami S. Models to predict length of stay in the emergency department: a systematic literature review and appraisal. *BMC Emergency Medicine*. 2024;24:54. doi:10.1186/s12873-024-00965-4.
- [23] Buttigieg SC, Abela L, Pace A. Variables affecting hospital length of stay: a scoping review. *Journal of Health Organization and Management*. 2018;32(3):463–493. doi:10.1108/JHOM-10-2017-0275.
- [24] Peres IT, Hamacher S, Oliveira FLC, Thomé AMT, Bozza FA. What factors predict length of stay in the intensive care unit? Systematic review and meta-analysis. *Journal of Critical Care*. 2020;60:183–194. doi:10.1016/j.jcrc.2020.08.003.
- [25] Marrie RA, Dawson NV, Garland A. Quantile regression and restricted cubic splines are useful for exploring relationships between continuous variables. *Journal of Clinical Epidemiology*. 2009;62(5):511–517.e1. doi:10.1016/j.jclinepi.2008.05.015.
- [26] Sun Y, Teow KL, Heng BH, Ooi CK, Tay SY. Real-time prediction of waiting time in the emergency department, using quantile regression. *Annals of Emergency Medicine*. 2012;60(3):299–308. doi:10.1016/j.annemergmed.2012.03.011.
- [27] Kanzaria HK, Probst MA, Ponce NA, Hsia RY. The association between advanced diagnostic imaging and ED length of stay. *American Journal of Emergency Medicine*. 2014;32(10):1253–1258. doi:10.1016/j.ajem.2014.07.038.
- [28] Kaufman BG, Klemish D, Kassner CT, et al. Predicting length of hospice stay: an application of quantile regression. *Journal of Palliative Medicine*. 2018;21(8):1131–1136. doi:10.1089/jpm.2018.0039.

- [29] Staffa SJ, Kohane DS, Zurakowski D. Quantile regression and its applications: a primer for anesthesiologists. *Anesthesia and Analgesia*. 2019;128(4):820–830. doi:10.1213/ANE.0000000000004017.
- [30] Staffa SJ, Zurakowski D. Calculation of confidence intervals for differences in medians between groups and comparison of methods. *Anesthesia and Analgesia*. 2020;130(2):542–546. doi:10.1213/ANE.0000000000004535.
- [31] Vicendese D, Te Marvelde L, McNair PD, et al. Hospital characteristics, rather than surgical volume, predict length of stay following colorectal cancer surgery. *Australian and New Zealand Journal of Public Health*. 2020;44(1):73–82. doi:10.1111/1753-6405.12932.
- [32] Walsh B, Lyons S, Smith S, Wren MA, Eighan J, Morgenroth E. Does formal home care reduce inpatient length of stay? *Health Economics*. 2020;29(12):1620–1636. doi:10.1002/hec.4158.
- [33] Kilkenny MF, Phan HT, Lindley RI, et al. Utility of the Hospital Frailty Risk Score derived from administrative data and the association with stroke outcomes. *Stroke*. 2021;52(9):2874–2881. doi:10.1161/STROKEAHA.120.033648.
- [34] Ke X, Lin W, Li D, et al. Spending and hospital stay for melanoma in Hunan, China. *Frontiers in Public Health*. 2022;10:917119. doi:10.3389/fpubh.2022.917119.
- [35] Wong BLL, Chan YH, O’Neill GK, Murphy D, Merchant RA. Frailty, length of stay and cost in hip fracture patients. *Osteoporosis International*. 2023;34(1):59–68. doi:10.1007/s00198-022-06553-1.
- [36] Xiao Y, Yu Y, Zhou Z, et al. Hospital spending and length of hospital stay for mental disorders in Hunan, China. *Heliyon*. 2023;9(4):e14968. doi:10.1016/j.heliyon.2023.e14968.
- [37] Yang P, Yang R, Luo Y, Zhang Y, Hu M. Hospitalization costs of road traffic injuries in Hunan, China: a quantile regression analysis. *Accident Analysis and Prevention*. 2024;194:107368. doi:10.1016/j.aap.2023.107368.
- [38] Alvarado F, Kaban J, Chao E, Meltzer JA. Surgical stabilization of rib fractures in patients with pulmonary comorbidities. *Injury*. 2023;54(5):1287–1291. doi:10.1016/j.injury.2023.02.009.
- [39] Frostig T, Benjamini Y, Kehat O, et al. Developing a length of stay prediction model for newborns, achieving better accuracy with greater usability. *International Journal of Medical Informatics*. 2023;180:105267. doi:10.1016/j.ijmedinf.2023.105267.
- [40] Pak A, Gannon B, Staib A. Predicting waiting time to treatment for emergency department patients. *International Journal of Medical Informatics*. 2021;145:104303. doi:10.1016/j.ijmedinf.2020.104303.

- [41] Berry JG, Staffa SJ, Hong P, et al. Factors affecting length of stay for children hospitalized after pediatric surgical procedures. *Hospital Pediatrics*. 2024;14(10):799–808. doi:10.1542/hpeds.2023-007622.
- [42] Brekke RL, Almeland SK, Hufthammer KO, Guttormsen AB, Hansson E, Onarheim H. Critical appraisal of the one-day-per-percent rule: a register-based observational study. *Burns*. 2024;50(8):1991–1998. doi:10.1016/j.burns.2024.07.033.
- [43] Chung M, Almarzooq ZI, Tale A, et al. Days at home after transcatheter mitral valve repair versus medical therapy alone in heart failure. *Journal of the American Heart Association*. 2025;14(1):e038401. doi:10.1161/JAHA.124.038401.
- [44] Liu X, Ma Q, Tong D, Shen Y. Analysis of hospitalization costs in adult inguinal hernia: based on quantile regression model. *Hernia*. 2024;28(5):1969–1978. doi:10.1007/s10029-024-03138-1.
- [45] Murthy P, Soraisham A, Momin S, et al. Neonatal level of care and length of stay of moderate and late preterm infants. *Journal of Neonatal-Perinatal Medicine*. 2025;18(4):369–376. doi:10.1177/19345798251339648.
- [46] Tang X, Gong Y, Chen Y, Zhou Y, Wang Y. Impact of treatment management on the hospital stay in patients with acute coronary syndrome. *BMC Cardiovascular Disorders*. 2024;24:630. doi:10.1186/s12872-024-04304-0.
- [47] Swarbrick CJ, Williams K, Evans B, et al. Postoperative outcomes in older patients living with frailty and multimorbidity in the UK: SNAP-3, a snapshot observational study. *British Journal of Anaesthesia*. 2025;135(1):166–176. doi:10.1016/j.bja.2025.04.026.
- [48] Bao X, Huang W, Lin Y, et al. Hospitalization costs of oral cancer patients in South-east of China: a quantile regression analysis. *BMC Health Services Research*. 2025;25:532. doi:10.1186/s12913-025-12675-z.
- [49] Tavares AI. Treatable mortality and health care related factors across European countries. *Frontiers in Public Health*. 2024;12:1301825. doi:10.3389/fpubh.2024.1301825.
- [50] Nacher M, Deungoue S, Brousse P, et al. [The interplay between isolation and precariousness, and hospitalization duration in French Guiana]. *Revue d'épidémiologie et de santé publique*. 2020;68(1):3–9. doi:10.1016/j.respe.2019.09.004.
- [51] Abbas S, Ihle P, Hein R, et al. [Rehabilitation in geriatric patients after ischemic stroke: a comparison of 2 organizational systems in Germany using claims data of a statutory health insurance fund]. *Die Rehabilitation*. 2013;52(4):228–234. doi:10.1055/s-0033-1334914.
- [52] Stata Press. **qreg: quantile regression**. In: *Stata Base Reference Manual: Release 18*. College Station: Stata Press, 2023.

- [53] SAS Institute Inc. **The QUANTREG Procedure.** In: *SAS/STAT® 15.1 User's Guide*. Cary, NC: SAS Institute Inc., 2023.