ISSN: 2583-5386
DOI: doi.org/10.51219/abel-getachew/91

URF PUBLISHERS
connect with research world

Journal of Integrated Health

https://urfpublishers.com/journal/integrated-health

Vol: 5 & Iss: 4

Review

Stepwise and Ridge Regression in Biomedical Research: A Narrative Methodological Review and Decision Framework

Abel Getachew[1]* and Ilker Etikan[2]

[1]PhD Candidate, Department of Biostatistics, Near East University, Nicosia, Cyprus

[2]Faculty of Medicine, Department of Biostatistics, Near East University, Nicosia, Cyprus

Citation: Getachew A, Etikan I. Stepwise and Ridge Regression in Biomedical Research: A Narrative Methodological Review and Decision Framework. *J Integrated Health* 2026;5(4): 565-578. DOI: doi.org/10.51219/abel-getachew/91

Received: 18 October, 2026; **Accepted:** 26 October, 2026; **Published:** 28 October, 2026

***Corresponding author:** Abel Getachew, PhD Candidate, Department of Biostatistics, Near East University, Nicosia, Cyprus, E-mail: 20244863@std.neu.edu.tr

Copyright: © 2026 Getachew A, et al., This is an open-access article published in J Integrated Health (JIH) and distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

A B S T R A C T

Background: The procedure of stepwise regression introduces variables while, in the case of ridge regression, the procedure is the one of shrinking coefficients. In this way, the two techniques approach the problem of model complexity in a slightly different way.

Objective: Combining the methodological approaches with their implications in practice and forming an objective-based decision-making matrix.

Methods: This narrative methodological review examined the principles, strengths, limitations and biomedical applications of stepwise regression and ridge regression. The available bibliographic records comprised 671 entries: 608 from PubMed and 63 from Google Scholar. Duplicate checking based on titles and DOIs identified 23 duplicate records, leaving 648 records for relevance screening. One researcher reviewed available titles and abstracts against the review's eligibility criteria. Fifty sources were selected for qualitative synthesis based on their relevance to three main areas: (1) fundamental statistical principles of the methods; (2) methodological and simulation research; and (3) biomedical applications and relevant methodological guidance. The selected literature was synthesized narratively, without quantitative pooling or meta-analysis.

Result: This review highlights methodological differences, instead of indicating comparative efficiency. Although stepwise methods yield simple models, selected variables can differ based on the sample, variable acquisitions and classification procedures. Ridge regression incorporates an L2 penalty to lessen estimator variability for correlated variables, but introducing bias is unavoidable; therefore, superior prediction outside of the dataset is not guaranteed. Direct assessment of the head-to-head methods is not available in the analyzed sources.

Conclusion: Use techniques that are applicable for the problem analyzed, determine the modeling procedure in advance and check all steps that rely on the data.

Keywords: Stepwise regression; Ridge regression; Variable selection; Multicollinearity; Penalized regression; Biomedical research; Prediction models; Validation

1. Introduction

Biomedical researchers often face a practical choice when candidate predictors are numerous or correlated: should they seek a compact set of variables or stabilize estimates while retaining a broader set? The answer matters because a model built for prediction has different requirements from an analysis intended to estimate an etiological or adjusted association. Clinical registries, electronic health records, laboratory measurements, imaging and molecular data make this choice common, while model validity also depends on sample information, missing-data handling, specification and generalizability[1,2].

The methods of stepwise regression differ from those of ridge regression in terms of handling complexity in a model. Stepwise methods give a model various variables as predictors based on an established rule of selection; on the other hand, ridge regression applies a loss function to be minimized with the use of the L_2 penalty that causes the coefficients to shrink in size and keeps candidate predictors in the model[3].

The practical consequences of these various technologies of modeling should be emphasized. For instance, while in the presence of correlated predictors, the use of the stepwise method may lead to different predictors in different samples, ridge regression is able to overcome the issue of correlation in general. It needs to be pointed out that there is no method capable of identifying the causal relationships within the analysis or improving the quality of predictions at the same time[4-7].

This narrative methodological review integrates the contributions of available medical and methodological publications and comes up with an objective-oriented framework for methodology selection and validation with regard to a task. The review presents a framework that connects particular research objectives with modeling considerations; however, it is not focused on evaluating performance of the methods.

2. Methods

2.1. Review design and sources

This research project is classified as the narrative methodological review, as opposed to systematic review or meta-analysis. The bibliography comprises 671 studies (from 2021 to 2026) retrieved from Google scholar (63 records) and PubMed (608 records). records were associated with an export dated 13 October 2026.

2.2. Relevance criteria

Potentially relevant records were determined when the titles or abstracts they contained were about either advanced or retrogressive regression, variable selection, ridge regression or any one of the penalized regression methods, coefficient shrinkage, multicollinearity, any one of the methods that deal with model overfitting and clinical or medical prediction models. The selected eligible types of sources included both statistical publications on methodology and simulation, method reviews, biomedical prediction methodologies and guides on reporting and prediction models. The documents were evaluated based on the review questions and the possible contribution of each document to finding out how the methods in question function, how to compare these methods and how to apply them in practice in the field of medicine.

Records were excluded from further analysis if they were identified as either duplicates, irrelevant records, ones referring to something non-biological without having any methodological contribution or lacking enough information to evaluate the relevance of each document.

2.3. Record selection and duplication

According to one reviewer, it was assessed that by looking at bibliographic titles and available abstracts 50 sources for focused narrative synthesis relevant to the review questions were selected. The 50 sources were chosen to fully cover basic statistical studies, methodological and simulation research, applied biomedical research as well as reporting/validation practices and to contain sources providing explanations of the underlying mechanisms in the models, shrinkage/selection processes, validation methods or decision-making suggestions. The result was purposive rather than exhaustive narrative sampling aimed at gathering an adequate representation of the literature.

2.4. Data extraction and narrative synthesis

The analysis of the selected 50 sources was a thorough examination of critical elements such as the research approach taken for the problem, statistical method and the techniques adopted for choosing the right variables, such as regression, parameter correlations, analysis goal and validation method.

2.5. Study-selection summary

A total of 671 records were analyzed, out of which 63 were found to be from Google Scholar and 608 from PubMed. The normalized-title/DOI check led to the identification of 23 duplications and as a result, 648 records were confirmed as the final records to be selected. Among them, 50 records were purposefully chosen for narrative synthesis, while 598 records went unselected.

These numbers provide an overview of the bibliographies and narrative-selection process executed as per the literature. It is important to note that the 598 records that were not chosen for synthesis were not excluded on the basis of the full text but were based on individual decisions.

The record-selection counts and the limits of the available screening documentation are summarized in **(Table 1)**.

Table 1: Summary of bibliographic record selection.

Item	Number
Google Scholar records	63
PubMed records	608
Total Records	671
Records flagged as duplicate candidates	23
Records remaining after provisional deduplication	648
Sources purposively selected for focused narrative synthesis	50
Records not selected for focused narrative synthesis	598

3. Statistical Foundations

3.1. Ordinary least squares

In the context of a continuous outcome, one can describe multiple linear regression as

$$Y_i = \beta_0 + \Sigma_{j=1}^{p} \beta_j X_{ij} + \varepsilon_i.$$

The estimation of ordinary least squares by minimizing the error sum of squares leads to estimates of the coefficients. If predictors are highly correlated, then estimates of coefficients will have large variances and be sensitive to small changes in the data. This leads to problems with interpretation as well as reduced ability to make predictions when the model that has been fitted is not stable[8,9].

3.2. Stepwise regression

Stepwise regression consists of successive processes of variable selection. Forward selection begins with an intercept-only or predetermined model and adds new predictors based on an entry criterion. Backward elimination begins with a candidate model and eliminates predictors based on a removal criterion. Bidirectional techniques do both entering and exiting of variables. The entry criterion can be based on p-values, likelihood methods or information criteria such as AIC; the criterion, set of candidates and stopping rules should be clearly defined[10-13].

The end model depends on the predictors used, sample used and selection method. Predictors can compete for entry if they are correlated, so different samples will have different sets of selected predictors. Selection can lead to obtaining overly optimistic estimates and p-values, since consideration of selection uncertainty is usually not taken into account. Yet, having a small model does not mean that it is valid, unbiased or transferable[3-17].

3.3. Ridge regression

For a continuous outcome, ridge regression estimates coefficients by minimizing the residual sum of squares with an L_2 penalty. The penalty parameter λ is nonnegative and controls the strength of shrinkage; $\lambda = 0$ gives ordinary least squares, while larger values generally shrink the slope coefficients more strongly toward zero. In predictive modeling, λ should be chosen using a prespecified tuning procedure, commonly cross-validation, performed entirely within the training data. The selected value is data- and procedure-dependent and should be reported:

$$RSS(\beta) = \Sigma_i [y_i - \beta_0 - \Sigma_{j=1}^{p} \beta_j x_{ij}]^2 + \lambda \Sigma_{j=1}^{p} \beta_j^2, \lambda \geq 0.$$

Here, β_0 is the intercept, which is conventionally not penalized and λ controls the degree of shrinkage. This convention should be confirmed in the software and implementation used.

For a binary outcome, ridge logistic regression uses a different loss function.

Let $p_i = Pr(Y_i = 1 \mid x_i)$ and $logit(p_i) = \beta_0 + \Sigma_{j=1}^{p} \beta_j x_{ij}$. The penalized objective is the negative logistic log-likelihood plus an L_2 penalty:

$$-\Sigma_i \{y_i \log(p_i) + (1 - y_i)\log(1 - p_i)\} + \lambda \Sigma_{j=1}^{p} \beta_j^2, \lambda \geq 0.$$

The linear-regression RSS formulation should not be interpreted as the objective for a binary outcome. In both formulations, the penalty trades some coefficient bias for a potential reduction in variance. This can improve estimation stability or prediction when predictors are correlated, but better predictive performance is not guaranteed; it depends on the data, outcome, tuning procedure and validation design. The penalty parameter is commonly selected using cross-validation or another prespecified tuning procedure[18,19].

3.4. Scaling, preprocessing and tuning

The ridge penalty operates on the coefficients; hence numerical variables that are measured on different scales should be standardized and centered before any fitting procedure. Without standardization, the variable measured on larger scales will be affected for reasons that do not relate to their actual contribution. Categorical variables require a coding system, e.g., effect coding or reference coding. Changing the scaling of the coding variables will change the effective penalty imposed on the variable and this must be indicated. The intercept is usually not penalized, but there are different implementations in the literature, which should be checked.

All data-dependent activities including imputation, scaling and selection of variables need to be carried out only with respect to the already held training set. The processes that were selected will then be used for calculations in the test set sampled in the resampling procedure. If these steps were performed for full data before the CV, the information from the test set can leak into the modeling process, which will lead to bias in the performance of the model. Thus, using the whole procedure in every training sample is paramount for unbiased conclusions[7,9,20].

4. Assumptions and Diagnostic Considerations

The selection of a variable or regularization should not take the place of a properly defined model. Researchers should examine many aspects, including the form of functional relationships, the type of outcome distributions and residual behaviour and the locations of heterogeneity and dependence between repeated measures, clustering or missing data. For nonlinear relationships, transformations and splines might be considered. For clustered or longitudinal data, mixed models, generalized estimating equations or other approaches catering for dependence must be used[8,21,22].

The analysis of multicollinearity must be carried out in relation to the structure of predictors and their objectives in terms of science. Correlation matrices, variance inflation factors, condition indices and knowledge of the subject matter can help determine how predictors are dependent. Furthermore, correlated predictors may still find themselves useful for prediction purposes and thus not solely depend on correlation. In turn, ridge regression allows for less sensitivity of estimates of coefficients to correlated predictors by shrinking those down but does not eliminate correlation. These issues are to be taken into account at the stage of study design, processing of data, formulation of the model and analyses conducted[8,23,24].

The missing-data procedures and preprocessing techniques are necessary for validation of prediction models. The variables must be selected based on scientific estimand and proper confounding strategy in the case of etiological research instead of being guided solely by statistical variables or algorithm for their selection[4,6,25].

5. Stepwise Regression: Strengths and Limitations

Stepwise regression's main practical attraction is that it can produce a compact predictor subset that may be easier to communicate or implement. It can be used as an exploratory procedure when the candidate set, selection criterion and limitations are reported transparently[5,11].

The chosen set may be unpredictable, especially when

predictor variables are interrelated or effects are not strong enough. Continued testing can result in statistical estimates appearing overly optimistic. Exclusion of predictors doesn't mean that clinical relevance has been established and their inclusion doesn't mean that the predictor variable in question works in a causal manner. The stepwise selection approach may choose a variable from a correlated set while failing to resolve multicollinearity[12,14,16,17].

In the prediction setting, selection has to be carried out during every resample to make validation adequate. In the case of causal analysis, covariates have to be specified according to the theoretical knowledge and reasonable biases instead of being picked only by means of significant statistics[1,26,27].

6. Ridge Regression: Strengths and Limitations

Ridge regression is relevant when the aim is to stabilize coefficient estimates or prediction while retaining a broad candidate set, particularly when predictors are correlated. Shrinkage can reduce variance, but ridge does not eliminate predictor correlation and a nonzero coefficient is not evidence by itself of a causal effect or clinical importance.

Ridge keeps all probable predictors, but its estimates are biased towards zero and influenced by factors like scaling, coding and penalty selection. Also, it cannot fix issues like poor design, confounding, measurement errors in a model[9,28,29].

7. Comparative Analysis

7.1. Selection versus shrinkage

Stepwise regression changes predictor membership through discrete inclusion and exclusion. Ridge regression retains candidate predictors but continuously shrinks coefficient magnitudes. They therefore answer different questions and should not be compared only by the number of predictors in the final model[3,12,15,18].

7.2. Prediction versus explanation

Prediction focuses on accurate outcomes for new observations in a defined target population and setting. It requires evaluation of discrimination, calibration, overall prediction error and, when relevant, clinical utility. Explanatory or etiological analysis focuses on an interpretable scientific estimand and appropriate control of confounding. A predictor that improves prediction is not necessarily a causal factor and an algorithmically excluded variable may still be scientifically important[1,2,25,30,31].

7.3. Stability and interpretability

When describing stability, we should do so for the particular type of stability involved. Coefficient stability refers to the extent of variability in the coefficient estimates in terms of their values and signs in different samples or resamples. Selection stability indicates the frequency of selection of the same predictors in different samples and resamples. Tuning stability refers to the variability of the chosen penalty parameter λ across resamples or acceptable tuning perturbations. Predictive-performance stability refers to the variability of out-of-sample performance measures in resamples or validation datasets. The listed properties are distinct from each other and must not be treated as synonymous.

Compact models can be better communicated whereas ridge models preserve the information included in correlated predictors. Neither compactness nor shrinkage establish interpretability, causal relevance or transportability alone. Wherever stability is an important property, the researchers should specify the particular type of stability they measure and provide appropriate summaries based on resampling together with the validation results.

7.4. Validation and performance

In order to make a fair comparison, both stepwise and ridge model must work in similar development datasets, outcome definitions, target populations and predictions horizons. There should be an intentional focus on making outcome-appropriate performance measures that should also reflect uncertainty, not just point estimates. One has to consider the whole development pipeline within validation. The methodology used in developing models should be checked, including missing- data strategies, imputation, scaling, variable selection and penalty tuning. An internal validation proofed by bootstrap or cross-validation does not equal external validation of independent datasets or target population. For binary outcome analyses multiple aspects may be studied and in case of continuous outcomes, RMSE and MAE will fit. No one metric will cover all possible case.

7.5. Relation to LASSO and elastic net

LASSO applies an L_1 penalty and can set coefficients exactly to zero, combining shrinkage with variable selection. Elastic Net combines L_1 and L_2 penalties and may be useful when predictors are correlated. These methods are alternatives, not automatic replacements for stepwise or ridge regression. Selection should reflect the objective, predictor structure, sample information and validated performance[28,29,32].

7.6. Comparative evidence and strength of inference

Foundational and methodological evidence. Hoerl and Kennard[18] established the ridge-regression approach, in which an L_2 penalty shrinks coefficient estimates to address problems associated with unstable least-squares estimation. In contrast, stepwise regression uses sequential inclusion or exclusion rules to select a subset of candidate predictors. Methodological discussions of variable selection emphasize that the results of stepwise procedures depend on the candidate predictor set, selection criteria and observed data and that conventional inference after selection can be misleading when selection uncertainty is ignored[12,14]. These sources explain why the methods differ; they do not establish that one method will outperform the other in every biomedical dataset.

Simulation and predictive-performance evidence. Predictive performance depends on sample size, the number and correlation of candidate predictors, predictor-outcome associations, tuning and validation. Ridge regression can reduce coefficient variance through shrinkage, but the bias-variance trade-off means improved prediction is not guaranteed in every setting. Stepwise regression may yield a smaller model, although selected variables and apparent performance can be sensitive to sampling variation and the selection process. Studies of penalized-regression tuning provide contextual evidence about model development. Protocols like Kipruto and Sauerbrei[15] and Ullmann, et al.[13] exemplify the concept of planned research, which must never be equated with accomplished empirical discoveries.

Biomedical applications and reporting recommendations. Biomedical applications demonstrate cases of applying particular regression techniques in specific research contexts; however,

in their final form, their conclusions rely on the outcome of interest, the form of predictors, the sample size and the goal of the analysis. As an example, Arashi, et al.[23] explain the use of ridge regression in genetics research. All other cases mentioned in the evidence map present different outcomes, techniques or comparators but do not provide indications of a universal triumph of one technique over another.

Recommendations provided by sources like TRIPOD and TRIPOD+AI aid in explaining model construction, predictor development and results[1,30]. As a result, it is possible to understand whether the obtained results can be generalized or how reliable those results are without knowing that one method is more correct than another.

In conclusion, the available literature suggests the interpretation depending on the method and not universal grading of the methods. The mechanisms work according to their mathematical definitions, while links between reliability and practical instability as well as predictive ability is mainly indirect. In order to claim superiority, a direct comparison is needed under relevant conditions for the future usage of the method in biomedicine.

Comparisons between methods must feature the definition of the target sample and outcomes beforehand, application of identical training data and predictors, repetition of data processing and tuning in accordance with sample data and assessment of out-of-sample results.

The principal differences in model-building mechanism, predictor retention, interpretability and validation are compared in **(Table 2)**.

Table 2: Methodological comparison of stepwise and ridge regression.

Characteristic	Stepwise regression	Ridge regression
Primary mechanism	Sequential inclusion/exclusion	L_2 coefficient shrinkage
Predictors retained	Selected subset	Generally, retains candidate predictors
Exact zero coefficients	Yes, by exclusion	Usually no
Correlated predictors	May switch among correlated variables[3,12]	Shrinks correlated coefficients jointly[7,18]
Main tuning choices	Selection criterion and entry/removal rules	Penalty parameter λ and tuning criterion
Interpretability	Often compact, but selection is sample-dependent	May retain more predictors with smaller coefficients
Main risk	Selection instability and post-selection optimism[12,14]	Bias; results depend on scaling and λ[7,19]
Validation	Repeat selection inside resampling[1,33]	Repeat preprocessing and λ tuning inside resampling[1,7]
Causal interpretation	Not established by selection	Not established by shrinkage

Note: Author synthesis of the cited methodological literature; the table summarizes conceptual differences and should not be interpreted as direct comparative-performance evidence.

The selected literature spans foundational methods, variable-selection and penalization studies, biomedical applications and reporting guidance; **(Table 3)** summarizes these evidence groups, while Supplementary Table S1 provides the source-by-source map.

Evidence groups are summarized by their role in the review and the limits on interpretation. The full source-by-source evidence map is provided in Supplementary Table S1.

Table 3: Summary of evidence types and their contribution to the review.

Evidence group	What it contributes	Interpretive boundary
Foundational statistical methods	Defines least-squares estimation, stepwise inclusion/exclusion and ridge coefficient shrinkage.	Mathematical definitions explain mechanisms; they do not establish comparative clinical performance.
Variable-selection and overfitting literature	Supports caution about selection instability, repeated testing, optimism and the need to validate the full modeling process.	Evidence from methods other than the exact stepwise-versus-ridge comparison is indirect.
Ridge and penalized-regression studies	Informs discussion of shrinkage, correlated predictors, tuning and potential bias–variance trade-offs.	Performance depends on data conditions, preprocessing, tuning and validation design.
Biomedical applications	Illustrates how regression procedures are used in specific clinical and biomedical settings.	A single application does not establish general superiority or transferability.
Prediction-model validation and reporting guidance	Informs transparent reporting, internal/external validation, calibration, discrimination and uncertainty assessment.	Guidance supports appraisal and reporting; it is not itself a head-to-head experiment.
Direct stepwise-versus-ridge evidence	No source in the currently available selected set was verified here as a direct head-to-head comparison under comparable conditions.	This describes the reviewed source set; it does not prove that no such study exists.

Note: This concise synthesis and Supplementary Table S1 are author-organized evidence maps, not formal risk-of-bias assessments, systematic evidence grading or certainty ratings.

8. Practical Framework for Biomedical Researchers

When using **(Table 4)** as a decision-making resource, it's important to first determine whether the objective is prediction or causal hypothesis testing. Based on this determination, one must consider whether to select a set of compact predictors or to conduct a shrinkage of the more extensive predictors. You must also pre-specify the evaluation in reporting results from your new knowledge. In etiological studies, the causal estimator and control for confounding seem to be influenced significantly by the subject matter and causal knowledge rather than statistical learning. Hence the choice of adjustment set is based on relevant expert knowledge rather than automatic method selection[4,6,8,26,34].

Table 4: Objective-led decision framework for choosing between stepwise and ridge regression.

Biomedical research situation	Methodological consideration	Recommended approach
Prediction with correlated candidate predictors	Shrinkage may improve coefficient estimation and predictive performance, but improvement is not guaranteed. Ridge regression does not remove correlation among predictors.	Consider ridge regression and tune the penalty using resampling; evaluate the complete modeling pipeline.
A small, explicitly required predictor subset	Stepwise regression can select a subset, but the selected variables may be unstable across samples.	Use cautiously; prespecify the candidate set and selection rule and validate the entire selection procedure.
Estimation of an etiological or adjusted association	Automated, prediction-oriented selection may omit important confounders or select variables inappropriate for the scientific estimand.	Prespecify the estimand and confounder strategy; do not use either algorithm alone to establish causality.
Model intended for use in a different hospital or population	Internal validation does not establish transportability to a different setting or population.	Evaluate the final model in an independent external population where feasible.

Note: This author-developed framework summarizes conditional methodological guidance; it is not a ranking based on a new empirical comparison.

The recommendations are conditional on the research objective, predictor structure and validation design; they do not rank the algorithms using a new empirical comparison.

For a prediction study, use the following implementation and reporting checklist to make comparisons fair and reproducible:

For both methods, report the outcome and target population, candidate predictor set, missing-data procedure, coding decisions and the complete modeling specification.

For stepwise regression, report the direction (forward, backward or bidirectional), selection criterion, entry/removal thresholds, stopping rule and whether the full selection process was repeated within each training resample.

Provide details about the scaling of continuous and categorical predictors, grid of λ values, selection criterion for λ, chosen value and whether intercept was penalized.

Identify the validation approach and the performance metrics appropriate to the problem, ensure all data-specific preprocessing and tuning is done based on training resamples and use these separated steps for assessing each validation set.

Present performance estimates along with their uncertainty specification, clarify internal validation vs. external validation and provide stability estimates when stability is of interest in the study.

If any comparison of predictive performance is made, use the same development data, outcome definition, population under study, predictor variables and prediction period used before. Use the same measures of performance that are applicable to the problem and its purpose.

This article is methodological and does not contain any empirical comparisons. There is no single illustrative dataset used, nor are there any estimates of AUROC, calibration, error rate and benefit in the clinic.

9. Results

9.1. Overview of the available evidence

The synthesis of the narrative was created on the basis of Studies of 671 records of which 63 were attributed to Google Scholar whereas 608 to PubMed. The first stage of the process of selection of records allowed the existence of 23 duplicate records. As a result, there were 648 records selected for further analysis. After that, 50 records were chosen by the qualitative criteria to represent the basic regression methods, the methods and simulations studies, the topics of biomedical research and the model development and report guidance issues.

9.2. Differences in modeling mechanisms

The literature examined differentiates stepwise regression from ridge regression depending on the treatment of model complexity. The stepwise regression process eliminates or adds predictors in stages, conforming to the selected method of description, so that the model becomes comprised of a subset of predictors selected from the pool of candidate predictors. In this approach, ridge regression is characterized by implementing L_2 penalty in the objective function, which makes the coefficients close to zero but preserves the predictor variables.

The different approaches give rise to different analytical approaches. Stepwise regression emphasizes the selection and simplicity of the model, while ridge regression highlights coefficient shrinking and stabilization potential. Ridge regression is not usually characterized by making predictors numerically zero and does not remove inter-correlation processes from the model.

9.3. Coefficient, Selection and Predictive stability

The analysis suggests that stability needs to be assessed based on the particular modeling variable of interest. The stepwise selection may give rise to different sets of predictors between various samples and resampling strategies, especially in the case of collinearity among predictor variables or presence of many predictors providing a similar level of information. Thus, a compact model that is selected may not necessarily suggest that the variables selected have been identified in a similar way across all samples.

One way Ridge regression helps to reduce variance of the coefficients is by means of shrinkage. This is true even if the predictors are correlated, but it is followed by the bias-variance tradeoff. How effective Ridge regression is depends on the particularities of the data used, sample information, penalty parameter and validation technique. Therefore, it cannot be concluded that the emergence of a smaller model or implementation of penalty would guarantee a good predictive performance. Coefficient stability, the stability of selection and tuning and predictive performance stability are different properties and should be measured and reported separately.

9.4. Implications for biomedical research objectives

According to the reviewed sources, it is important to clearly differentiate between prediction and etiological/explanatory

analysis before selecting a regression approach. For prediction, the key criterion is how well this approach performs in the targeted population. Variable selection, data preprocessing and penalty tuning should be part of evaluation to achieve realistic performance estimates and prevent information leakage. Correct internal validation and, if possible, external validation are important for assessing model's performance and generalizability[1,2].

In contrast, for etiological/explanatory analyses, the choice of regression technique should depend on the underlying research question, target estimand and confounder-adjustment method. A predictor being important for prediction is not necessarily a causal factor and the variable excluded from the model may also be important for scientific explanation. None of the selection methods by itself provides evidence of the existence of causal relationships, nor does it guarantee unbiased results.

9.5. Comparative evidence and decision framework

the evidences advocates for a methodological comparison identifying each of the methods mechanism, limitations and appropriate use. Basic studies can explain the mathematical basis of ridge regression while methodological as well as biomedical applications can provide the contextual evidence with regard to variable selection, shrinkage and validation and interpretation. However, materials on each of methods cannot reasonably provide evidence that one method is better than another generally.

The proposed guidelines suggest that the choice of method depends on the main analytics goal, the predictor structure, the data available and the plan for the results interpretation. Stepwise regression can be a good fit if a small number of predictors is needed, although the potential for contradicting selection decisions should be properly handled. Ridge regression can be useful in the presence of predictor correlation as long as the method is properly adjusted and tested. Transparent presentation and verification of all aspects of the modeling process are also key components of applying both methods.

In summary, the narrative synthesis presents evidence that suggests using an objective-dependent approach for choosing the method. However, it does not provide a definitive ranking for stepwise and ridge regression. To draw this conclusion a number of formally conducted experiments comparing the suggested methodologies would be required.

10. Discussion

It is vital to decide on the method before comparing effectiveness results from various algorithms. A compact model and a reduced model will have different objectives; therefore, it is inappropriate to judge their effectiveness based solely on their number of predictors. In case of prediction, credibility of the comparison relies on the validation of the entire development process. In case of estimation of an underlying cause of a certain condition, the choice of variables is mainly determined by the scientific estimand and the selected confounding approach. Unfortunately, based only on the data set involved in the present study, it is impossible to make a direct comparison between stepwise regression and ridge regression methods. There are evident mathematical differences between them, yet results of studies, which involve other methods of variable selection and penalized regression, do not enable making general conclusions about the superiority of one of these approaches.

While stepwise regression comes up with a group of the candidate predictors, ridge regression works by reducing the coefficient estimates, hence keeping the candidate predictors. Since the two methods have different objectives, this review does not prove which one method is superior to the other.

11.Conclusion

In choosing between the methods of stepwise regression and ridge regression, the issues of model complexity have to be dealt with and the method has to be appropriate for the purpose of the research. Stepwise regression leads to more parsimonious models since variable choices are based on certain criteria. However, what is obtained may depend heavily on the sample, the chosen candidates and the choice method. Ridge regression uses the penalty of L_2 shrinkage, which helps in obtaining more stable coefficients and introducing bias.

When selecting the method, its choice should be based on the research question, its nature, intended usage of the model, data type and estimates. In the case of predictions, it is necessary to check all the modeling pipeline and validate every step related to data, including methods of variable selection, processing and coupling. If etiological or explanatory studies should be carried out, it is very important to specify what is to be estimated and properly adjust for con-founders based on expert knowledge and not just select methods automatically.

The present review has presented a framework that emphasizes what the strengths and weaknesses of methods are, for what to use them. Nevertheless, since we cannot conduct a direct comparison of the two methods, the results do not provide a universal ranking of how the methods perform. Future work should focus on transparent and reproducible comparison of the two methods in realistic biomedical settings, looking into their predictive capabilities, stability of coefficients, selection of variables and how the results of both methods would affect scientific conclusion. In the end, the quality of the results of a regression analysis depends not only on the method itself but also on the design of the study, clarity of reporting and validation of the method.

12. Declarations

12.1. Author contributions

Abel Getachew contributed to the conception of the review, literature identification, methodological synthesis, interpretation, and manuscript preparation and revision. Prof. Dr. İlker Etikan contributed to the review's conceptualization, methodological guidance, critical review, and revision of the manuscript. Both authors reviewed and approved the final version of the manuscript.

12.2. Funding

No specific funding was received for this work.

12.3. Competing interests

The author declares no competing interests.

12.4. Ethics approval

Ethical approval was not required because this narrative review used previously published literature and did not involve identifiable individual-level data.

12.5. Consent for publication

Not applicable.

12.6. Data availability

The bibliographic workbook supplied for this revision was used as the source for the focused literature selection. No primary dataset was generated or analyzed.

12.7. Use of artificial intelligence

ChatGPT (OpenAI; GPT-6; source: https://chatgpt.com/) was used to assist with structural revision, language editing and consistency checks during manuscript revision in October 2026.

13. References

1. Collins GS, Reitsma JB, Altman DG, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD statement. Annals of Internal Medicine, 2015;162: 55-63.
2. Steyerberg EW. Clinical prediction models: A practical approach to development, validation and updating (2nd ed.). Springer, 2019.
3. Hanke M, Dijkstra L, Foraita R, et al. Variable selection in linear regression models: Choosing the best subset is not always the best choice. Biometrical Journal, 2024;66: 2200209.
4. Dyer BP. Variable selection for causal inference, prediction and descriptive research: A narrative review of recommendations. European Heart Journal Open, 2025;5(3).
5. Kaneko H. Examining variable selection methods for the predictive performance of regression models and the proportion of selected variables and selected random variables. Heliyon, 2021;7: e07356.
6. Staerk C, Byrd A, Mayr A. Recent methodological trends in epidemiology: No need for data-driven variable selection? American Journal of Epidemiology, 2024;193: 370-376.
7. Van Calster B, van Smeden M, Steyerberg EW, et al. On the variability of regression shrinkage methods for clinical prediction models: Simulation study on predictive performance. Diagnostic and Prognostic Research, 2019;3: 8.
8. Harrell FE. Regression modeling strategies: With applications to linear models, logistic and ordinal regression and survival analysis (2nd ed.). Springer, 2015.
9. James G, Witten D, Hastie T, et al. An introduction to statistical learning: With applications in R (2nd ed.). Springer, 2021.
10. Bag S, Gupta K, Deb S. A review and recommendations on variable selection methods in regression models for binary data. arXiv, 2022.
11. Chowdhury MZI, Turin TC. Variable selection strategies and its importance in clinical prediction modelling. Family Medicine and Community Health, 2020;8: e000262.
12. Heinze G, Wallisch C, Dunkler D. Variable selection-A review and recommendations for the practicing statistician. Biometrical Journal, 2018;60: 431-449.
13. Ullmann T, Heinze G, Hafermann L, et al. Evaluating variable selection methods for multivariable regression models: A simulation study protocol. PLOS ONE, 2024;19: e0308543.
14. Babyak MA. What you see may not be what you get: A brief, nontechnical introduction to overfitting in regression-type models. Psychosomatic Medicine, 2004;66: 411-421.
15. Kipruto E, Sauerbrei W. Comparison of variable selection procedures and investigation of the role of shrinkage in linear regression: Protocol of a simulation study in low-dimensional data. PLOS ONE, 2022;17: e0271240.
16. Mahmoudian M, Venäläinen MS, Klén R, et al. Stable iterative variable selection. Bioinformatics, 2021;37: 4810-4817.
17. Pressat-Laffouilhère T, Jouffroy R, Leguillou A, et al. Variable selection methods were poorly reported but rarely misused in major medical journals: Literature review. Journal of Clinical Epidemiology, 2021;139: 12-19.
18. Hoerl AE, Kennard RW. Ridge regression: Biased estimation for nonorthogonal problems. Technometrics, 1970;12: 55-67.
19. Pavlou M, Omar RZ, Ambler G. Penalized regression methods with modified cross-validation and bootstrap tuning produce better prediction models. Biometrical Journal, 2024;66: e202300245.
20. Kuhn M, Johnson K. Feature engineering and selection: A practical approach for predictive models. CRC Press, 2019.
21. Iwagami M, Matsui H. Introduction to clinical prediction models. Annals of Clinical Epidemiology, 2022;4: 72-80.
22. Zhao Y, Kohli N, Lock EF. Evaluating Bayesian variable selection approaches for nonlinear random effects models. Multivariate Behavioral Research, 2021;1-20.
23. Arashi M, Roozbeh M, Hamzah NA, et al. Ridge regression and its applications in genetic studies. PLOS ONE, 2021;16: e0245376.
24. Davino C, Romano R, Vistocco D. Handling multicollinearity in quantile regression through the use of principal component regression. METRON, 2022;80: 153-174.
25. Webster-Clark M, Ross RK, Keil AP, et al. Variable selection when estimating effects in external target populations. American Journal of Epidemiology, 2024;193: 1176-1181.
26. Efthimiou O, Seo M, Chalkou K, et al. Developing clinical prediction models: A step-by-step guide. BMJ, 2024;386: e078276.
27. Riley RD, Ensor J, Snell KIE, et al. Calculating the sample size required for developing a clinical prediction model. BMJ, 2020;368: 441.
28. Tibshirani R. Regression shrinkage and selection via the LASSO. Journal of the Royal Statistical Society: Series B (Methodological), 1996;58: 267-288.
29. Zou H, Hastie T. Regularization and variable selection via the Elastic Net. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 2005;67: 301-320.
30. Collins GS, Moons KGM, Dhiman P, et al. TRIPOD+AI statement: Updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ, 2024;385: e078378.
31. Vickers AJ, van Calster B, Steyerberg EW. A simple, step-by-step guide to interpreting decision curve analysis. Diagnostic and Prognostic Research, 2019;3: 18.
32. Tay JK, Narasimhan B, Hastie T. Elastic net regularization paths for all generalized linear models. Journal of Statistical Software, 2023;106: 1-31.
33. Wolff RF, Moons KGM, Riley RD, et al. PROBAST: A tool to assess the risk of bias and applicability of prediction model studies. Annals of Internal Medicine, 2019;170: 51-58.
34. Moons KGM, Wolff RF, Riley RD, et al. PROBAST: A tool to assess risk of bias and applicability of prediction model studies: Explanation and elaboration. Annals of Internal Medicine, 2019;170: W1-W33.
35. Gilholm P, Lister P, Irwin A, et al. Comparison of random forest and stepwise regression for variable selection using low prevalence predictors: A case study in paediatric sepsis. Maternal and Child Health Journal, 2025;29: 604-613.
36. Huang Y, Cho M, Chakraborty S, et al. Variable selection for prediction in clinical research. WIREs Computational Statistics, 2025;17: e70030.
37. Montesinos-López A, Montesinos-López OA, Lecumberry F, et al. Refining penalized ridge regression: A novel method for optimizing the regularization parameter in genomic prediction. G3: Genes, Genomes, Genetics, 2024;14.

38. Moons KGM, Altman DG, Reitsma JB, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): Explanation and elaboration. Annals of Internal Medicine, 2015;162: W1-W73.

39. Riley RD, Debray TPA, Collins GS, et al. Minimum sample size for external validation of a clinical prediction model with a binary outcome. Statistics in Medicine, 2021;40: 4230-4251.

40. Riley RD, Snell KIE, Ensor J, et al. Minimum sample size for developing a multivariable prediction model: Part I-Continuous outcomes. Statistics in Medicine, 2019;38: 1262-1275.

41. Riley RD, Snell KIE, Ensor J, et al. Minimum sample size for developing a multivariable prediction model: Part II-Binary and time-to-event outcomes. Statistics in Medicine, 2019;38: 1276-1296.

42. Rubin J, Mariani L, Smith A, et al. Ridge regression for functional form identification of continuous predictors of clinical outcomes in glomerular disease. Glomerular Diseases, 2023;3: 47-55.

43. Snell KIE, Archer L, Ensor J, et al. External validation of clinical prediction models: Simulation-based sample size calculations were more reliable than rules-of-thumb. Journal of Clinical Epidemiology, 2021;135: 79-89.

44. Snigurska UA, Ser SE, Solberg LM, et al. Application of a practice-based approach in variable selection for a prediction model development study of hospital-induced delirium. BMC Medical Informatics and Decision Making, 2023;23: 181.

45. Sperrin M, Riley RD, Collins GS, et al. Targeted validation: Validating clinical prediction models in their intended population and setting. Diagnostic and Prognostic Research, 2022;6: 24.

46. Su X, Ambler G, Green N, et al. The use of variable selection in clinical prediction modeling for binary outcomes: A systematic review. Journal of Clinical Epidemiology, 2026;197: 112371.

47. Sztepanacz JL, Houle D. Regularized regression can improve estimates of multivariate selection in the face of multicollinearity and limited data. Evolution Letters, 2024;8: 361-373.

48. van Smeden M, Moons KGM, de Groot JAH, et al. Sample size for binary logistic prediction models: Beyond events per variable criteria. Statistical Methods in Medical Research, 2019;28: 2455-2474.

49. Wan B, Caffo B, Vedula SS. A unified framework on generalizability of clinical prediction models. Frontiers in Artificial Intelligence, 2022;5: 872720.

50. Wong KY, Zeng D, Lin DY. Penalized regression for multiple types of many features with missing data. Statistica Sinica, 2023;33: 633-662.

Supplementary Table S1: Detailed evidence map of the 50 selected sources

Supplementary Table S1 records the source-level evidence type, topic, contribution and interpretation used in the narrative synthesis. These descriptions are not a formal quality rating.

No. / source	Evidence type	Methods examined / topic	Source-specific contribution to this review	Comparative-evidence interpretation
1. Arashi, M., Roozbeh, M., Hamzah, N. A., & Gasparini, M. (2021). Ridge regression and its applications in genetic studies. PLOS ONE, 16(1), e0245376.	Methodological study or applied regularization study	Ridge regression in genetic studies; shrinkage with correlated/high-dimensional predictors.	Provides the foundational ridge-regression formulation for nonorthogonal predictors and the rationale for accepting some coefficient bias to reduce variance; the genetic-study context illustrates relevance to correlated features.	Indirect evidence on ridge/regularization, coefficient shrinkage or correlated predictors; no direct stepwise–ridge result is claimed here.
2. Babyak, M. A. (2004). What you see may not be what you get: A brief, nontechnical introduction to overfitting in regression-type models. Psychosomatic Medicine, 66(3), 411–421.	Methodological / simulation / penalized-regression study	Regression model development and selection; overfitting and optimism.	Explains overfitting in regression through a practical, nontechnical discussion, supporting the review's caution that data-driven model building can produce optimistic or unstable results.	Contextual methodological or biomedical evidence; it informs model development or interpretation but does not directly compare stepwise and ridge regression.
3. Bag, S., Gupta, K., & Deb, S. (2022). A review and recommendations on variable selection methods in regression models for binary data. arXiv. https://arxiv.org/abs/2201.06063	Review / methodological recommendations	Variable-selection approaches for binary regression; review of alternatives.	Reviews variable-selection approaches for binary-response regression and offers methodological recommendations; used as secondary context, with claims limited to what can be confirmed from the available record.	Indirect evidence on variable-selection practice or recommendations; useful for context, not a direct stepwise–ridge comparison.
4. Chowdhury, M. Z. I., & Turin, T. C. (2020). Variable selection strategies and its importance in clinical prediction modelling. Family Medicine and Community Health, 8(1), e000262.	Review / methodological recommendations	Variable-selection strategies used in clinical prediction modeling.	Discusses variable-selection strategies in clinical prediction modeling and why selection decisions matter; supports the practical discussion of choosing and reporting selection procedures.	Indirect evidence on variable-selection practice or recommendations; useful for context, not a direct stepwise–ridge comparison.
5. Collins, G. S., Moons, K. G. M., Dhiman, P., Riley, R. D., Beam, A. L., Van Calster, B., et al. (2024). TRIPOD+AI statement: Updated guidance for reporting clinical prediction models that use regression or machine learning methods. BMJ, 385, e078378.	Reporting, appraisal, validation or sample-size guidance	Regression and machine-learning prediction models; reporting standards.	TRIPOD+AI reporting guidance for prediction models using regression or machine learning; informs transparency in reporting model development, validation and evaluation, not which algorithm is superior.	Indirect evidence for reporting, validation, appraisal, sample-size planning or clinical utility; it does not establish comparative superiority of stepwise or ridge regression.
6. Collins, G. S., Reitsma, J. B., Altman, D. G., & Moons, K. G. M. (2015). Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD statement. Annals of Internal Medicine, 162(1), 55–63.	Reporting, appraisal, validation or sample-size guidance	Multivariable prediction models; transparent reporting.	Introduces the original TRIPOD reporting framework for multivariable diagnostic/prognostic prediction models; supports completeness and transparency of reporting rather than methodological performance claims.	Indirect evidence for reporting, validation, appraisal, sample-size planning or clinical utility; it does not establish comparative superiority of stepwise or ridge regression.
7. Davino, C., Romano, R., & Vistocco, D. (2022). Handling multicollinearity in quantile regression through the use of principal component regression. METRON, 80, 153–174.	Methodological / simulation / penalized-regression study	Principal-component regression in quantile regression to address multicollinearity.	Examines principal-component regression as an approach to handling multicollinearity in quantile regression; provides a contrast showing that correlated predictors can be addressed by methods other than ridge.	Contextual methodological or biomedical evidence; it informs model development or interpretation but does not directly compare stepwise and ridge regression.
8. Dyer, B. P. (2025). Variable selection for causal inference, prediction and descriptive research: A narrative review of recommendations. European Heart Journal Open, 5(3), oeaf070.	Review / methodological recommendations	Variable selection for causal/effect estimation, prediction and descriptive objectives.	Narrative review distinguishing variable-selection goals in causal inference, prediction and descriptive research; supports the argument that selection strategy must follow the scientific estimand and purpose.	Indirect evidence on variable-selection practice or recommendations; useful for context, not a direct stepwise–ridge comparison.
9. Efthimiou, O., Seo, M., Chalkou, K., Debray, T. P. A., Egger, M., & Salanti, G. (2024). Developing clinical prediction models: A step-by-step guide. BMJ, 386, e078276.	Biomedical application / applied prediction study	Clinical prediction-model development, validation and interpretation.	Step-by-step guidance on developing clinical prediction models; supports the workflow discussion from defining the prediction task through model development and evaluation.	Contextual methodological or biomedical evidence; it informs model development or interpretation but does not directly compare stepwise and ridge regression.
10. Gilholm, P., Lister, P., Irwin, A., Harley, A., Raman, S., Schlapbach, L. J., & Gibbons, K. S. (2025). Comparison of random forest and stepwise regression for variable selection using low prevalence predictors: A case study in paediatric sepsis. Maternal and Child Health Journal, 29(5), 604–613.	Biomedical application / applied prediction study	Random forest versus stepwise regression for variable selection in low-prevalence predictors.	Paediatric-sepsis case study comparing random forest and stepwise regression for selecting low-prevalence predictors; illustrates an applied comparison in a specific clinical setting, not universal superiority.	Direct comparison of random forest and stepwise regression only; it does not compare stepwise regression with ridge regression. The reported results are specific to low-prevalence predictors in paediatric sepsis.

No. / source	Evidence type	Methods examined / topic	Source-specific contribution to this review	Comparative-evidence interpretation
11. Hanke, M., Dijkstra, L., Foraita, R., & Didelez, V. (2024). Variable selection in linear regression models: Choosing the best subset is not always the best choice. Biometrical Journal, 66(1), 2200209.	Methodological / simulation / penalized-regression study	Best-subset variable selection versus broader modeling strategies.	Examines the premise that selecting the single "best" subset is not always optimal in linear regression; informs the discussion of alternatives to subset selection and the importance of the modeling objective.	Indirect evidence on variable-selection practice or recommendations; useful for context, not a direct stepwise–ridge comparison.
12. Harrell, F. E., Jr. (2015). Regression modeling strategies: With applications to linear models, logistic and ordinal regression and survival analysis (2nd ed.). Springer.	Textbook / methodological reference	Regression modeling across linear, logistic ordinal and survival outcomes.	Provides broad regression-modeling guidance across linear, logistic ordinal and survival settings; supports explanations of model specification, overfitting, interpretation and validation.	Contextual methodological or biomedical evidence; it informs model development or interpretation but does not directly compare stepwise and ridge regression.
13. Heinze, G., Wallisch, C., & Dunkler, D. (2018). Variable selection—A review and recommendations for the practicing statistician. Biometrical Journal, 60(3), 431–449.	Review / methodological recommendations	Variable-selection methods; comparator and outcomes should be confirmed in full text.	Reviews variable-selection methods and gives recommendations for applied statisticians; supports the synthesis of selection limitations, uncertainty and cautious interpretation after data-driven selection.	Indirect evidence on variable-selection practice or recommendations; useful for context, not a direct stepwise–ridge comparison.
14. Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. Technometrics, 12(1), 55–67.	Methodological / simulation / penalized-regression study	Ridge regression (L□ penalty) versus ordinary least squares under nonorthogonality.	Original mathematical paper introducing ridge regression for nonorthogonal design problems; anchors the L2-penalized least-squares formulation and bias–variance rationale.	Direct evidence about ridge regression versus ordinary least squares; indirect for the stepwise–ridge question.
15. Huang, Y., Cho, M., Chakraborty, S., & Dey, T. (2025). Variable selection for prediction in clinical research. WIREs Computational Statistics, 17(2), e70030.	Biomedical application / applied prediction study	Variable-selection methods; comparator and outcomes should be confirmed in full text.	Reviews variable selection for prediction in clinical research; supports the distinction between selecting variables for predictive accuracy and selecting variables for explanatory aims.	Indirect evidence on variable-selection practice or recommendations; useful for context, not a direct stepwise–ridge comparison.
16. Iwagami, M., & Matsui, H. (2022). Introduction to clinical prediction models. Annals of Clinical Epidemiology, 4(3), 72–80.	Biomedical application / applied prediction study	Clinical prediction-model development/evaluation; exact estimators depend on the cited study.	Introductory overview of clinical prediction models; supports basic terminology and the sequence of development, validation and interpretation in clinical settings.	Contextual methodological or biomedical evidence; it informs model development or interpretation but does not directly compare stepwise and ridge regression.
17. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). An introduction to statistical learning: With applications in R (2nd ed.). Springer.	Textbook / methodological reference	General statistical learning, including regression, regularization and model assessment.	Textbook background on statistical learning, including regression, model assessment, resampling and preprocessing; supports general explanations rather than serving as a biomedical head-to-head study.	Contextual methodological or biomedical evidence; it informs model development or interpretation but does not directly compare stepwise and ridge regression.
18. Kaneko, H. (2021). Examining variable selection methods for the predictive performance of regression models and the proportion of selected variables and selected random variables. Heliyon, 7(6), e07356.	Methodological / simulation / penalized-regression study	Variable-selection methods; comparator and outcomes should be confirmed in full text.	Evaluates variable-selection methods in relation to regression predictive performance and the frequency of selected variables, including random variables; informs discussion of both prediction and selection behavior.	Indirect evidence on variable-selection practice or recommendations; useful for context, not a direct stepwise–ridge comparison.
19. Kipruto, E., & Sauerbrei, W. (2022). Comparison of variable selection procedures and investigation of the role of shrinkage in linear regression: Protocol of a simulation study in low-dimensional data. PLOS ONE, 17(10), e0271240.	Study protocol (planned methods; not completed results)	Penalized regression/ shrinkage methods; details and comparator depend on study design.	Protocol for a planned simulation comparing variable-selection procedures and investigating shrinkage in low-dimensional linear regression; describes intended methods, not completed simulation findings.	Indirect evidence on ridge/regularization, coefficient shrinkage or correlated predictors; no direct stepwise–ridge result is claimed here.
20. Kuhn, M., & Johnson, K. (2019). Feature engineering and selection: A practical approach for predictive models. CRC Press.	Textbook / methodological reference	Method(s) indicated by the source title; verify exact analysis and comparator in full text.	Practical text on feature engineering and selection for predictive models; supports the discussion of preprocessing, resampling and keeping data-dependent steps inside training folds.	Contextual methodological or biomedical evidence; it informs model development or interpretation but does not directly compare stepwise and ridge regression.
21. Mahmoudian, M., Venäläinen, M. S., Klén, R., & Elo, L. L. (2021). Stable iterative variable selection. Bioinformatics, 37(24), 4810–4817.	Methodological / simulation / penalized-regression study	Iterative variable selection with emphasis on selection stability.	Presents a stable iterative variable-selection method; relevant to the question of reproducibility of selected predictors, but does not by itself establish that ridge and stepwise regression are interchangeable.	Indirect evidence on variable-selection practice or recommendations; useful for context, not a direct stepwise–ridge comparison.

No. / source	Evidence type	Methods examined / topic	Source-specific contribution to this review	Comparative-evidence interpretation
22. Montesinos-López, A., Montesinos-López, O. A., Lecumberry, F., Fariello, M. I., Montesinos-López, J. C., & Crossa, J. (2024). Refining penalized ridge regression: A novel method for optimizing the regularization parameter in genomic prediction. G3: Genes, Genomes, Genetics, 14(12), jkae246.	Methodological study or applied regularization study	Ridge regression and tuning of regularization for genomic prediction.	Proposes a method for optimizing the ridge regularization parameter in genomic prediction; informs the discussion that predictive performance depends on penalty tuning and the application context.	Indirect evidence on ridge/regularization, coefficient shrinkage or correlated predictors; no direct stepwise–ridge result is claimed here.
23. Moons, K. G. M., Altman, D. G., Reitsma, J. B., Ioannidis, J. P. A., Macaskill, P., Steyerberg, E. W., et al. (2015). Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): Explanation and elaboration. Annals of Internal Medicine, 162(1), W1–W73.	Reporting, appraisal, validation or sample-size guidance	Multivariable prediction models; transparent reporting.	TRIPOD explanation-and-elaboration document; supplies rationale and item-level detail for reporting prediction-model studies, complementing the shorter TRIPOD statement.	Indirect evidence for reporting, validation, appraisal, sample-size planning or clinical utility; it does not establish comparative superiority of stepwise or ridge regression.
24. Moons, K. G. M., Wolff, R. F., Riley, R. D., Whiting, P. F., Westwood, M., Collins, G. S., Reitsma, J. B., Kleijnen, J., & Mallett, S. (2019). PROBAST: A tool to assess risk of bias and applicability of prediction model studies: Explanation and elaboration. Annals of Internal Medicine, 170(1), W1–W33.	Reporting, appraisal, validation or sample-size guidance	Clinical prediction-model development/evaluation; exact estimators depend on the cited study.	Explains PROBAST domains for judging risk of bias and applicability in prediction-model studies; supports appraisal of model development and validation evidence rather than ranking regression methods.	Indirect evidence for reporting, validation, appraisal, sample-size planning or clinical utility; it does not establish comparative superiority of stepwise or ridge regression.
25. Pavlou, M., Omar, R. Z., & Ambler, G. (2024). Penalized regression methods with modified cross-validation and bootstrap tuning produce better prediction models. Biometrical Journal, 66(5), e202300245.	Reporting, appraisal, validation or sample-size guidance	Clinical prediction-model development/evaluation; exact estimators depend on the cited study.	Studies modified cross-validation and bootstrap approaches for tuning penalized regression; informs how tuning and resampling choices may affect predictive-model development.	Indirect evidence for reporting, validation, appraisal, sample-size planning or clinical utility; it does not establish comparative superiority of stepwise or ridge regression.
26. Pressat-Laffouilhère, T., Jouffroy, R., Leguillou, A., Kerdelhue, G., Benichou, J., & Gillibert, A. (2021). Variable selection methods were poorly reported but rarely misused in major medical journals: Literature review. Journal of Clinical Epidemiology, 139, 12–19.	Review / methodological recommendations	Variable-selection methods; comparator and outcomes should be confirmed in full text.	Literature review of reporting and use of variable-selection methods in major medical journals; supports the concern that methods need to be described transparently and reproducibly.	Indirect evidence on variable-selection practice or recommendations; useful for context, not a direct stepwise–ridge comparison.
27. Riley, R. D., Debray, T. P. A., Collins, G. S., Archer, L., Ensor, J., van Smeden, M., et al. (2021). Minimum sample size for external validation of a clinical prediction model with a binary outcome. Statistics in Medicine, 40(19), 4230–4251.	Reporting, appraisal, validation or sample-size guidance	Clinical prediction-model development/evaluation; exact estimators depend on the cited study.	Develops sample-size considerations for external validation of binary-outcome clinical prediction models; informs validation planning rather than initial predictor-selection choices.	Indirect evidence for reporting, validation, appraisal, sample-size planning or clinical utility; it does not establish comparative superiority of stepwise or ridge regression.
28. Riley, R. D., Ensor, J., Snell, K. I. E., Debray, T. P. A., Altman, D. G., Moons, K. G. M., et al. (2020). Calculating the sample size required for developing a clinical prediction model. BMJ, 368, m441.	Reporting, appraisal, validation or sample-size guidance	Clinical prediction-model development/evaluation; exact estimators depend on the cited study.	Provides sample-size guidance for developing clinical prediction models; supports the caution that model complexity and sample size influence overfitting risk.	Indirect evidence for reporting, validation, appraisal, sample-size planning or clinical utility; it does not establish comparative superiority of stepwise or ridge regression.
29. Riley, R. D., Snell, K. I. E., Ensor, J., Burke, D. L., Harrell, F. E., Moons, K. G. M., et al. (2019a). Minimum sample size for developing a multivariable prediction model: Part I—Continuous outcomes. Statistics in Medicine, 38(7), 1262–1275.	Reporting, appraisal, validation or sample-size guidance	Clinical prediction-model development/evaluation; exact estimators depend on the cited study.	Part I of a sample-size series addressing multivariable prediction models with continuous outcomes; informs planning for model development and the risk of overfitting.	Indirect evidence for reporting, validation, appraisal, sample-size planning or clinical utility; it does not establish comparative superiority of stepwise or ridge regression.
30. Riley, R. D., Snell, K. I. E., Ensor, J., Burke, D. L., Harrell, F. E., Moons, K. G. M., et al. (2019b). Minimum sample size for developing a multivariable prediction model: Part II—Binary and time-to-event outcomes. Statistics in Medicine, 38(7), 1276–1296.	Reporting, appraisal, validation or sample-size guidance	Clinical prediction-model development/evaluation; exact estimators depend on the cited study.	Part II of a sample-size series addressing binary and time-to-event outcomes; extends sample-size planning considerations beyond continuous-outcome models.	Indirect evidence for reporting, validation, appraisal, sample-size planning or clinical utility; it does not establish comparative superiority of stepwise or ridge regression.
31. Rubin, J., Mariani, L., Smith, A., & Zee, J. (2023). Ridge regression for functional form identification of continuous predictors of clinical outcomes in glomerular disease. Glomerular Diseases, 3(1), 47–55.	Methodological study or applied regularization study	Method(s) indicated by the source title; verify exact analysis and comparator in full text.	Clinical application of ridge regression to identify functional forms for continuous predictors of outcomes in glomerular disease; illustrates a specific biomedical use case rather than a general comparison with stepwise selection.	Indirect evidence on ridge/regularization, coefficient shrinkage or correlated predictors; no direct stepwise–ridge result is claimed here.

No. / source	Evidence type	Methods examined / topic	Source-specific contribution to this review	Comparative-evidence interpretation
32. Snell, K. I. E., Archer, L., Ensor, J., Bonnett, L. J., Debray, T. P. A., Phillips, B., et al. (2021). External validation of clinical prediction models: Simulation-based sample size calculations were more reliable than rules-of-thumb. Journal of Clinical Epidemiology, 135, 79–89.	Reporting, appraisal, validation or sample-size guidance	Clinical prediction-model development/evaluation; exact estimators depend on the cited study.	Uses simulation-based sample-size calculations for external validation and compares them with rules of thumb; supports context-specific validation planning.	Indirect evidence for reporting, validation, appraisal, sample-size planning or clinical utility; it does not establish comparative superiority of stepwise or ridge regression.
33. Snigurska, U. A., Ser, S. E., Solberg, L. M., Prosperi, M., Magoc, T., Chen, Z., Bian, J., Bjarnadottir, R. I., & Lucero, R. J. (2023). Application of a practice-based approach in variable selection for a prediction model development study of hospital-induced delirium. BMC Medical Informatics and Decision Making, 23, 181.	Methodological / simulation / penalized-regression study	Clinical prediction-model development/evaluation; exact estimators depend on the cited study.	Describes a practice-based approach to variable selection in developing a hospital-induced delirium prediction model; illustrates how clinical/practice considerations can inform candidate predictor choices.	Indirect evidence on variable-selection practice or recommendations; useful for context, not a direct stepwise–ridge comparison.
34. Sperrin, M., Riley, R. D., Collins, G. S., & Martin, G. P. (2022). Targeted validation: Validating clinical prediction models in their intended population and setting. Diagnostic and Prognostic Research, 6, 24.	Reporting, appraisal, validation or sample-size guidance	Clinical prediction-model development/evaluation; exact estimators depend on the cited study.	Discusses validation in the population and setting where a clinical prediction model is intended to be used; supports target-population relevance and generalizability considerations.	Indirect evidence for reporting, validation, appraisal, sample-size planning or clinical utility; it does not establish comparative superiority of stepwise or ridge regression.
35. Staerk, C., Byrd, A., & Mayr, A. (2024). Recent methodological trends in epidemiology: No need for data-driven variable selection? American Journal of Epidemiology, 193(2), 370–376.	Methodological review / causal or explanatory analysis	Variable-selection methods; comparator and outcomes should be confirmed in full text.	Reviews recent epidemiologic arguments about whether data-driven variable selection is needed; supports distinguishing prediction objectives from etiologic or causal research questions.	Indirect evidence on variable-selection practice or recommendations; useful for context, not a direct stepwise–ridge comparison.
36. Steyerberg, E. W. (2019). Clinical prediction models: A practical approach to development, validation and updating (2nd ed.). Springer.	Reporting, appraisal, validation or sample-size guidance	Clinical prediction-model development/evaluation; exact estimators depend on the cited study.	Comprehensive practical text on clinical prediction-model development, validation and updating; supports the review's workflow and the need for external validation and recalibration where appropriate.	Indirect evidence for reporting, validation, appraisal, sample-size planning or clinical utility; it does not establish comparative superiority of stepwise or ridge regression.
37. Su, X., Ambler, G., Green, N., & Pavlou, M. (2026). The use of variable selection in clinical prediction modelling for binary outcomes: A systematic review. Journal of Clinical Epidemiology, 197, 112371.	Review / methodological recommendations	Clinical prediction-model development/evaluation; exact estimators depend on the cited study.	Systematic review focused on use of variable selection in binary-outcome clinical prediction modeling; provides a field-level overview of reported practice, with conclusions interpreted according to the review's own methods and scope.	Indirect evidence on variable-selection practice or recommendations; useful for context, not a direct stepwise–ridge comparison.
38. Sztepanacz, J. L., & Houle, D. (2024). Regularized regression can improve estimates of multivariate selection in the face of multicollinearity and limited data. Evolution Letters, 8(3), 361–373.	Methodological study or applied regularization study	Method(s) indicated by the source title; verify exact analysis and comparator in full text.	Studies regularized regression for estimating multivariate selection under multicollinearity and limited data in an evolutionary context; informs the potential value and context-dependence of regularization, not a direct clinical algorithm ranking.	Contextual methodological or biomedical evidence; it informs model development or interpretation but does not directly compare stepwise and ridge regression.
39. Tay, J. K., Narasimhan, B., & Hastie, T. (2023). Elastic net regularization paths for all generalized linear models. Journal of Statistical Software, 106(1), 1–31.	Methodological / simulation / penalized-regression study	Penalized regression/ shrinkage methods; details and comparator depend on study design.	Describes elastic-net regularization paths for generalized linear models; provides technical context for combined L1/ L2 penalties and contrasts elastic net with ridge's coefficient shrinkage without automatic zeroing.	Indirect methodological evidence about elastic-net regularization; useful for distinguishing L1/L2 penalties, not a direct stepwise–ridge comparison.
40. Tibshirani, R. (1996). Regression shrinkage and selection via the LASSO. Journal of the Royal Statistical Society: Series B (Methodological), 58(1), 267–288.	Methodological / simulation / penalized-regression study	Penalized regression/ shrinkage methods; details and comparator depend on study design.	Introduces the LASSO, which combines regression fitting with an L1 penalty that can set some coefficients exactly to zero; provides a key contrast with ridge, which generally shrinks rather than eliminates coefficients.	Indirect methodological evidence about LASSO/penalization and sparsity; not a direct stepwise–ridge comparison.
41. Ullmann, T., Heinze, G., Hafermann, L., Schilhart-Wallisch, C., & Dunkler, D. (2024). Evaluating variable selection methods for multivariable regression models: A simulation study protocol. PLOS ONE, 19(8), e0308543.	Study protocol (planned methods; not completed results)	Variable-selection methods; comparator and outcomes should be confirmed in full text.	Protocol for a planned simulation study evaluating variable-selection methods in multivariable regression; identifies a planned comparison and must not be cited as reporting completed simulation results.	Indirect evidence on variable-selection practice or recommendations; useful for context, not a direct stepwise–ridge comparison.

No. / source	Evidence type	Methods examined / topic	Source-specific contribution to this review	Comparative-evidence interpretation
42. Van Calster, B., van Smeden, M., Steyerberg, E. W., & Collins, G. S. (2019). On the variability of regression shrinkage methods for clinical prediction models: Simulation study on predictive performance. Diagnostic and Prognostic Research, 3, 8.	Methodological / simulation / penalized-regression study	Clinical prediction-model development/evaluation; exact estimators depend on the cited study.	Simulation study of variability in regression-shrinkage methods for clinical prediction and their predictive performance; supports the point that performance and stability vary across methods and data conditions.	Indirect evidence on ridge/regularization, coefficient shrinkage or correlated predictors; no direct stepwise–ridge result is claimed here.
43. van Smeden, M., Moons, K. G. M., de Groot, J. A. H., Collins, G. S., Altman, D. G., Eijkemans, M. J. C., & Reitsma, J. B. (2019). Sample size for binary logistic prediction models: Beyond events per variable criteria. Statistical Methods in Medical Research, 28(8), 2455–2474.	Reporting, appraisal, validation or sample-size guidance	Clinical prediction-model development/evaluation; exact estimators depend on the cited study.	Examines sample-size considerations for binary logistic prediction models beyond a simple events-per-variable rule; supports avoiding overly simplistic sample-size thresholds when assessing overfitting risk.	Indirect evidence for reporting, validation, appraisal, sample-size planning or clinical utility; it does not establish comparative superiority of stepwise or ridge regression.
44. Vickers, A. J., van Calster, B., & Steyerberg, E. W. (2019). A simple, step-by-step guide to interpreting decision curve analysis. Diagnostic and Prognostic Research, 3, 18.	Reporting, appraisal, validation or sample-size guidance	Method(s) indicated by the source title; verify exact analysis and comparator in full text.	Practical guide to interpreting decision-curve analysis; supports evaluating clinical net benefit alongside conventional discrimination and calibration metrics.	Indirect evidence for reporting, validation, appraisal, sample-size planning or clinical utility; it does not establish comparative superiority of stepwise or ridge regression.
45. Wan, B., Caffo, B., & Vedula, S. S. (2022). A unified framework on generalizability of clinical prediction models. Frontiers in Artificial Intelligence, 5, 872720.	Reporting, appraisal, validation or sample-size guidance	Clinical prediction-model development/evaluation; exact estimators depend on the cited study.	Presents a framework for generalizability of clinical prediction models; supports the distinction between performance in development data and usefulness in intended external populations or settings.	Indirect evidence for reporting, validation, appraisal, sample-size planning or clinical utility; it does not establish comparative superiority of stepwise or ridge regression.
46. Webster-Clark, M., Ross, R. K., Keil, A. P., & Platt, R. W. (2024). Variable selection when estimating effects in external target populations. American Journal of Epidemiology, 193(8), 1176–1181.	Methodological review / causal or explanatory analysis	Variable selection for causal/effect estimation, prediction and descriptive objectives.	Examines variable selection when estimating effects in external target populations; supports caution that predictors useful for prediction need not be appropriate choices for causal or effect-estimation goals.	Indirect evidence on variable-selection practice or recommendations; useful for context, not a direct stepwise–ridge comparison.
47. Wolff, R. F., Moons, K. G. M., Riley, R. D., Whiting, P. F., Westwood, M., Collins, G. S., et al. (2019). PROBAST: A tool to assess the risk of bias and applicability of prediction model studies. Annals of Internal Medicine, 170(1), 51–58.	Reporting, appraisal, validation or sample-size guidance	Clinical prediction-model development/evaluation; exact estimators depend on the cited study.	Introduces PROBAST for assessing risk of bias and applicability of prediction-model studies; supports structured appraisal of model development and validation evidence.	Indirect evidence for reporting, validation, appraisal, sample-size planning or clinical utility; it does not establish comparative superiority of stepwise or ridge regression.
48. Wong, K. Y., Zeng, D., & Lin, D. Y. (2023). Penalized regression for multiple types of many features with missing data. Statistica Sinica, 33(2), 633–662.	Methodological / simulation / penalized-regression study	Penalized regression/shrinkage methods; details and comparator depend on study design.	Develops penalized-regression methods for settings with many features and missing data; relevant to the interaction between missing-data handling and regularized modeling, while implementation details should be checked in the full article.	Contextual methodological or biomedical evidence; it informs model development or interpretation but does not directly compare stepwise and ridge regression.
49. Zhao, Y., Kohli, N., & Lock, E. F. (2026). Evaluating Bayesian variable selection approaches for nonlinear random effects models. Multivariate Behavioral Research.	Methodological or applied study; verify design in full text	Variable-selection methods; comparator and outcomes should be confirmed in full text.	Evaluates Bayesian variable-selection approaches for nonlinear random-effects models; broadens the methodological context beyond stepwise and ridge regression and is not direct evidence comparing those two methods.	Indirect evidence about an alternative variable-selection/modeling family; do not interpret as a stepwise–ridge comparison.
50. Zou, H., & Hastie, T. (2005). Regularization and variable selection via the Elastic Net. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 67(2), 301–320.	Methodological / simulation / penalized-regression study	Penalized regression/shrinkage methods; details and comparator depend on study design.	Introduces elastic net as a combination of L1 and L2 regularization; supports the contrast between ridge shrinkage, LASSO-style sparsity and penalties that can combine both properties.	Indirect methodological evidence about elastic-net regularization; useful for distinguishing L1/L2 penalties, not a direct stepwise–ridge comparison.