Estado da publicação: O preprint não foi publicado em outro meio.

O que funciona? Por quê? Em que contexto? Uma proposta metodológica para avaliação de políticas através de revisão sistemática mista e sequencial

Bruno Marques Schaefer, Tomás Paixão Borges, Carlos Freitas, Frederico Augusto Auad de Gomes Filho, Ana Paula Lima dos Santos

https://doi.org/10.1590/SciELOPreprints.17986

Submetido em: 2026-09-08
Postado em: 2026-09-09 (versão 1)
(AAAA-MM-DD)

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.17986

O Que Funciona?

MAPE

Laboratório de Monitoramento e Avaliação de Políticas e Eleições do IESP-UERJ

O que funciona? Por quê? Em que contexto? Uma proposta metodológica para avaliação de políticas através de revisão sistemática mista e sequencial[1]

Bruno Marques Schaefer

Universidade do Estado do Rio de Janeiro (UERJ)

ORCID: https://orcid.org/0000-0002-5127-7240

E-mail: brunoschaefer@iesp.uerj.br

Tomás Paixão Borges

Universidade do Estado do Rio de Janeiro (UERJ)

ORCID: https://orcid.org/0000-0002-5276-6636

E-mail: tomasborges@iesp.uerj.br

Carlos Freitas

Universidade do Estado do Rio de Janeiro (UERJ)

ORCID: https://orcid.org/0000-0002-2493-8154

E-mail: carlosfreitas@iesp.uerj.br

Frederico Augusto Auad de Gomes Filho

Universidade do Estado do Rio de Janeiro (UERJ)

ORCID: https://orcid.org/0009-0004-2217-8077

E-mail: fredericofilho@iesp.uerj.br

Ana Paula Lima dos Santos

Universidade do Estado do Rio de Janeiro (UERJ)

ORCID: https://orcid.org/0000-0002-5566-423X

E-mail: anapaulalima@iesp.uerj.br

[1] Este texto é produto de uma série de discussões no âmbito do MAPE (Laboratório de Monitoramento e Avaliação de Políticas e Eleições), sediado no IESP-UERJ. Uma versão inicial foi apresentada no VI Encontro Nacional de Ensino e Pesquisa do Campo de Públicas (ENEPCP), em 2025. Após a apresentação e discussão no Laboratório, o texto foi reformulado e dois autores foram incluídos. O protocolo da revisão foi publicado no OSF (https://osf.io/b792u/overview) e as bases de dados estão disponíveis no OSF (https://osf.io/pb9xz/overview).

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.17986

Resumo: Este trabalho propõe um protocolo de revisão sistemática para a avaliação de políticas. A inovação do protocolo está na combinação dos aspectos misto e sequencial: misto por enfatizar a integração de diferentes tipos de estudo na revisão; sequencial por estruturar essa integração em etapas complementares, evitando a hierarquização convencional que tende a deixar aspectos importantes do funcionamento das políticas em segundo plano. O resultado é um guia para avaliações mais completas de políticas e intervenções, que podem ser apresentadas em diferentes graus de complexidade, seja na forma mais detalhada para a comunidade acadêmica ou sintetizada numa "caixa de ferramentas", mais adequada para informar a formulação e implementação das políticas ou a divulgação dos seus resultados. Além da discussão conceitual e descrição da metodologia, o trabalho contém ainda uma parte empírica, na qual aplicamos o protocolo ao caso concreto do Programa Bolsa Verde. Com isso, esperamos tornar a metodologia mais tangível e elucidar suas vantagens e limitações.

Palavras-chave: avaliação de políticas públicas; revisão sistemática; métodos mistos; síntese de evidências; Bolsa Verde.

What works? Why? In what context? A methodological proposal for policy evaluation through mixed and sequential systematic review.

Abstract: This paper proposes a systematic review protocol for policy evaluation. The protocol's innovation lies in combining mixed and sequential features: mixed, in that it emphasizes the integration of different study types within the review; sequential, in that it structures this integration into complementary stages, avoiding the conventional hierarchization that tends to leave important aspects of how policies function in the background. The result is a guide for more complete evaluations of policies and interventions, which can be presented at different levels of complexity, whether in a more detailed form for the academic community or synthesized into a "toolbox" better suited to informing policy formulation and implementation, or to disseminating results. Beyond the conceptual discussion and description of the methodology, the paper also includes an empirical section, in which we apply the protocol to the concrete case of the Bolsa Verde Program. In doing so, we hope to make the methodology more tangible and to elucidate its advantages and limitations.

Keywords: public policy evaluation; systematic review; mixed methods; evidence synthesis; Bolsa Verde.

¿Qué funciona? ¿Por qué? ¿En qué contexto? Una propuesta metodológica para la evaluación de políticas mediante una revisión sistemática mixta y secuencial.

Resumen: Este artículo propone un protocolo de revisión sistemática para la evaluación de políticas. La innovación del protocolo radica en la combinación de características mixtas y secuenciales: mixtas, porque enfatiza la integración de diferentes tipos de estudios dentro de la revisión; secuenciales, porque estructura esta integración en etapas complementarias, evitando la jerarquización convencional que tiende a dejar en

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.17986

segundo plano aspectos importantes del funcionamiento de las políticas. El resultado es una guía para evaluaciones más completas de políticas e intervenciones, que pueden presentarse en diferentes niveles de complejidad, ya sea de forma más detallada para la comunidad académica o sintetizadas en una "caja de herramientas" más adecuada para fundamentar la formulación e implementación de políticas, o para difundir los resultados. Además de la discusión conceptual y la descripción de la metodología, el artículo también incluye una sección empírica, en la que aplicamos el protocolo al caso concreto del Programa Bolsa Verde. De esta manera, esperamos hacer que la metodología sea más tangible y dilucidar sus ventajas y limitaciones.

Palabras clave: evaluación de políticas públicas; revisión sistemática; métodos mixtos; Síntesis de evidencia; Bolsa Verde.

1. INTRODUÇÃO

Avaliar políticas públicas é atribuir valor a uma intervenção, analisando seu desenho (ex ante), sua implementação (monitoramento) e/ou seus resultados (ex post) (Azeredo et al., 2012; Figueiredo & Figueiredo, 1986; Ramos & Schabbach, 2012). A pergunta que move esse campo — "quais políticas funcionam?" — ganhou destaque nas últimas décadas com o movimento de políticas baseadas em (ou informadas por) evidências, os incentivos de organizações nacionais e internacionais, as inovações metodológicas e as restrições fiscais sobre o gasto público (Faria, 2022; Koga et al., 2020; Nutley et al., 2013; Pawson, 2002). No Brasil, essa agenda ganhou ainda mais expressão recentemente: a discussão sobre a Reforma Administrativa, em especial a PEC 38/2025, dedica um eixo específico ao tema, propondo uma Lei de Responsabilidade por Resultados que exigiria planejamento estratégico, metas e avaliação para intervenções em todas as esferas da federação.

A avaliação, contudo, não segue fórmula única. Ela pode focar na eficiência do gasto, na execução do que era pretendido, nas percepções de gestores e usuários e, no exercício mais exigente, no impacto, a partir do contrafactual: "os resultados existiriam sem a política?". Este último caso impõe estabelecer uma relação de causalidade entre intervenção e resultado. Além disso, numa perspectiva otimista, essa relação também se sustenta em outros contextos e populações (Brady, 2011; Keele, 2015). É essa exigência que expõe o primeiro grande desafio de quem avalia: a ausência de evidência sobre a intervenção no próprio contexto. Um gestor que pretende implementar uma nova política num município, por exemplo, simplesmente não tem evidências sobre aquela intervenção naquele contexto específico. A saída, nesse caso, é recorrer a evidências produzidas em

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.17986

contextos diferentes, para intervenções semelhantes. Essa saída, no entanto, gera um segundo desafio, derivado do primeiro: como avaliar essa evidência externa e determinar, de fato, o que funciona? Reunidos estudos de contextos distintos, com desenhos, dados e definições variados, e não raro com conclusões que divergem entre si, como saber qual desses resultados deveria orientar a decisão em questão?

Parte da dificuldade em avaliar essa evidência decorre de sua própria heterogeneidade metodológica: nem todo desenho de pesquisa serve a toda pergunta avaliativa. Se o objetivo é estimar o efeito de um programa, desenhos experimentais, quase-experimentais, de coorte ou caso-controle são mais indicados; se o objetivo é compreender a satisfação dos beneficiários, a adequação da política ao contexto ou o processo pelo qual ela é entregue, estudos qualitativos são mais úteis (Nutley et al., 2013; Petticrew & Roberts, 2003). O que conta como boa evidência, portanto, depende do tipo de pergunta que se pretende responder. E um único estudo, por mais rigoroso que seja, raramente cobre mais de uma dessas perguntas ao mesmo tempo.

É diante desse quadro, evidência ausente no contexto local e, quando disponível em outros contextos, difícil de avaliar e por vezes conflitante, que a revisão sistemática (RS) se apresenta como alternativa. Em linhas gerais, a RS é um método de localizar, selecionar, sintetizar e analisar evidências (Dacombe, 2018; Petticrew & Roberts, 2006; Xiao & Watson, 2019), e o "sistemático" provém da adoção de um protocolo explícito para cada uma dessas etapas. O que a diferencia da revisão narrativa, torna o processo reprodutível e replicável, reduz vieses na escolha dos trabalhos e aumenta a transparência da pesquisa (Figueiredo Filho et al., 2019). Dessas características decorrem respostas diretas aos dois desafios apresentados. Diante da ausência de evidência local, a RS permite mobilizar estudos de outros contextos para subsidiar a construção do próprio desenho da política (Abreu et al., 2025; Lassance, 2026). Diante da dificuldade de avaliar essa evidência externa (heterogênea e, por vezes, divergente), os protocolos explícitos orientam o tomador de decisão sobre o que a literatura acumulada efetivamente demonstra, reduzindo a dependência de escolhas *ad hoc*; e quando estudos com dados e métodos distintos convergem para uma mesma conclusão, essa se torna mais robusta do que a de qualquer estudo isolado, o que incrementa também a validade externa, isto é, a capacidade de generalização dos achados para além do contexto de cada avaliação particular (Slough & Tyson, 2023; Vetter, 2019).

Apesar dessas vantagens, a RS convencional não está isenta de limitações. Uma crítica recorrente diz respeito ao predomínio de estudos quantitativos em seu escopo, em

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.17986

particular no formato de meta-análises (Dacombe, 2018; Nutley et al., 2013). Meta-análises estimam bem o efeito médio de uma política e respondem a perguntas como "funciona ou não?" e "quanto funciona?", mas dizem pouco a quem precisa saber por que ela funciona, como opera na prática e para quem seus efeitos são mais pronunciados (Bamberger, 2012; Batista & Domingos, 2017; Pawson et al., 2005; Spohn et al., 2026). Decidir sobre implementação exige compreender os mecanismos causais subjacentes (Dalkin et al., 2015; Ludwig et al., 2011), as condições de contexto que moderam os efeitos, os grupos para os quais a política é mais ou menos efetiva (Cintron et al., 2022) e as percepções de implementadoras e beneficiárias (Hunter & Sugiyama, 2014). O problema é agravado pela tendência dessas revisões a excluir metodologias qualitativas, justamente as mais adequadas para investigar processos de implementação: o que conta como boa evidência, como já observamos, depende do tipo de pergunta que se pretende responder (Nutley et al., 2013).

Uma resposta possível a essa limitação é a triangulação metodológica, que combina sistematicamente evidências quantitativas e qualitativas em uma mesma revisão, como propõem a revisão realista (Pawson et al., 2005), a revisão sistemática mista (Buller et al., 2018; Stern et al., 2020) e a revisão integrativa (Souza et al., 2010). É nessa direção que se situa a proposta deste artigo: um protocolo de revisão sistemática mista e sequencial. Mista porque integra, em um mesmo protocolo, evidências quantitativas e qualitativas; sequencial porque organiza essa integração em etapas progressivas e articuladas, evitando a hierarquização convencional que tende a deixar aspectos importantes do funcionamento das políticas em segundo plano (Small, 2011). O objetivo é essencialmente pedagógico: apresentar essa proposta a partir de suas seis etapas, oferecendo à leitora um guia orientado à aplicação prática. Não se trata de uma revisão exaustiva da literatura metodológica, mas de um esforço de organização, exemplificação e tradução de boas práticas para quem produz ou utiliza sínteses de evidências.

O artigo está organizado em quatro seções, além desta introdução e das considerações finais. A primeira seção discute os tipos de RS e o problema da integração entre diferentes tipos de evidência; a segunda apresenta a proposta e a caixa de ferramentas que a sintetiza; a terceira percorre as seis etapas do protocolo aplicadas a uma política concreta, o Programa Bolsa Verde; e a quarta discute vantagens e limites da abordagem.

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.17986

2. REVISÃO SISTEMÁTICA, TIPOS DE EVIDÊNCIA E AVALIAÇÃO DE POLÍTICAS

2.1 Definições de revisão sistemática

Em artigo de 1976, Glass cunha o termo meta-análise para designar um esforço de terceiro nível: não produzir novos dados, nem reanalisar dados primários, mas tomar as próprias análises como dados (Glass, 1976). Com o tempo, "meta-análise" passou a designar um tipo específico de sistematização, baseado em dados numéricos e métodos estatísticos, ao passo que a revisão sistemática se consolidou como o conceito mais amplo, guarda-chuva que abriga diferentes tipos de dados e formas de síntese, das numéricas às textuais (Dixon-Woods et al., 2005; Porta, 2016). A distinção importa: toda meta-análise é uma revisão sistemática, mas nem toda revisão sistemática é uma meta-análise.

Essa amplitude se reflete na diversidade de tipologias propostas. Xiao e Watson (2019) identificam dezesseis tipos de revisão, organizados em torno de quatro objetivos: descrever um campo, testar hipóteses, estender o conhecimento e criticar pressupostos (ver Anexo 2). Por sua vez, Sutton et al. (2019) mapeiam 48 tipos em sete famílias metodológicas (ver Anexo 3). Já Covidence (2025) aponta 12 tipos de revisão (ver Anexo 4). Por fim, a Campbell Collaboration adota classificação mais enxuta, com cinco tipos orientados por questões distintas (MacDonald et al., 2024) (ver Anexo 5). As diferenças entre eles dizem respeito a decisões que estruturam todo o processo: quais características dos estudos primários justificam sua inclusão? Que critérios definem a qualidade metodológica mínima? Qual o grau de especificidade da pergunta — do mais amplo ("o que funciona para melhorar a educação?") ao mais restrito ("o tamanho das turmas impacta o desempenho em contextos de baixa renda?")?

A multiplicidade pode ser organizada com o auxílio da literatura de formação de conceitos (Gerring, 1999; Sartori, 1970), que distingue níveis de abstração. No mais alto, a revisão sistemática se define pela existência de um protocolo explícito que a diferencia das narrativas; no intermediário situam-se os tipos de revisão, diferenciados pelo propósito e pelo desenho; e, no mais específico, as técnicas de síntese, que variam pela forma como os dados são combinados, da meta-análise frequentista à meta-etnografia (Snilstveit et al., 2012) (ver Anexo 6). Não existe, portanto, um tipo único adequado a todos os objetivos. A escolha deve ser guiada pela natureza da pergunta e pelo tipo de evidência disponível (Nutley et al., 2013; Xiao & Watson, 2019).

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.17986

2.2 Por que a revisão sistemática importa para a avaliação de políticas

Do ponto de vista metodológico, a principal contribuição da revisão sistemática para a avaliação dos efeitos de uma política ou intervenção é o incremento da validade externa. Um único estudo, ainda que rigoroso e com alta validade interna, não responde por si só se seus achados se aplicam a outros contextos, populações ou momentos. É esta limitação que a revisão endereça ao reunir evidências de intervenções semelhantes conduzidas em cenários distintos (Slough & Tyson, 2023; Vetter, 2019).

Do ponto de vista substantivo, ela ocupa posição singular no ciclo de políticas. A avaliação é um processo de investigação que analisa o desenho, a implementação e os resultados de intervenções governamentais, podendo ser conduzida antes, durante ou depois de sua implementação (Ramos & Schabbach, 2012). A revisão, por depender de estudos primários já realizados, é conduzida *ex post*, mas pode ser mobilizada de forma *ex ante*, como balizador para a adoção de novos programas (Abreu et al., 2025). Constitui, ademais, um produto final de pesquisa, o que facilita sua consulta por gestores e pode reduzir a assimetria de informação entre as comunidades acadêmica e decisória (Koga et al., 2024). Esse uso, contudo, não é necessariamente linear. As evidências podem ser aplicadas diretamente a uma decisão, reconfigurar de forma difusa a compreensão de um problema ou legitimar decisões já tomadas (Batista, 2025).

2.3 Da hierarquia à tipologia de evidências

O movimento de políticas informadas por evidências é fortemente orientado à inferência causal e à pesquisa quantitativa (Faria, 2022). Os guias para classificar evidências recorrem usualmente à hierarquia de evidência, que ordena os estudos pelo desenho e sua relação com a validade interna: revisões sistemáticas e meta-análises no nível 1, ensaios randomizados no nível 2, estudos de coorte e comparações observacionais no nível 3, estudos de caso no nível 4 e evidências anedóticas no nível 5 (Kano & Hayashi, 2021). A *Maryland Scale* opera sob uma lógica semelhante, ranqueando estudos conforme sua capacidade de demonstrar relações causais e classificando a intervenção como eficaz, promissora ou ineficaz conforme o número de estudos acima de determinado nível (Farrington et al., 2002; Kopittke & Ramos, 2021).

Perguntas fundamentais permanecem sem resposta nesse modelo: como a política funciona? Por que (não) funciona? Em que contextos o impacto é maior? Qual a percepção de beneficiários, gestores e implementadores? Essas questões podem ser respondidas por pesquisas quantitativas, via análise de mecanismos, moderadores e

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.17986

efeitos heterogêneos (Cintron et al., 2022; Ludwig et al., 2011), mas também pela pesquisa qualitativa. Avaliações mais completas integram, portanto, evidências de diferentes abordagens (Bamberger, 2012; Batista & Domingos, 2017), razão pela qual diversos autores defendem o uso de uma tipologia de evidências, mais do que de uma hierarquia (Nutley et al., 2013; Petticrew & Roberts, 2003).

O Quadro 1 sintetiza a relação entre desenhos de pesquisa e tipos de perguntas. Por exemplo: se o objetivo é estimar o efeito de um programa, desenhos experimentais, quase-experimentais, de coorte ou caso-controle são mais indicados do que pesquisas qualitativas. Por sua vez, se o objetivo é observar a satisfação dos beneficiários ou a adequação da política ao contexto, estudos qualitativos serão mais úteis. Hunter e Sugiyama (2014) recorrem a grupos focais com beneficiárias do mesmo programa para compreender se seu desenho institucional produz sentimento de inclusão e cidadania, dimensão que estimativas de efeito médio não captam.

Quadro 1: Tipos de pergunta avaliativa e desenhos de pesquisa adequados

Pergunta de pesquisa	**Qualitativa**	**Survey**	**Caso-controle**	**Coorte**	**ERC**	**Quase-exp.**	**Não exp.**	**RS**
Efetividade — isso funciona? Funciona melhor do que aquilo?			+	++	+	+		+++
Processo de entrega — como funciona?	++	+					+	+++
Relevância — isso importa?	++	++						+++
Segurança — faz mais bem do que mal?	+		+	+	++	+		+++
Aceitabilidade — o público aderirá ao serviço?	++	+			+	+	+	+++
Custo-efetividade — vale a pena investir?					++			+++
Adequação — é o serviço certo para este público?	++	++						++
Satisfação — usuários e provedores estão satisfeitos?	++	++	+	+			+	

Fonte: elaborado pelos autores a partir de Petticrew e Roberts (2003). Nota: ERC = ensaio randomizado controlado; RS = revisão sistemática. O número de sinais indica a adequação relativa do desenho à pergunta.

2.4 Uma tipologia transversal dos estudos incluídos

O Quadro 1 não esgota, contudo, a tarefa de integração. Um mesmo desenho pode cumprir finalidades distintas dentro de uma mesma revisão, ora descrevendo um fenômeno, ora testando uma relação causal. Para operacionalizá-la, propomos uma segunda tipologia, transversal, que cruza duas dimensões.

A primeira é o objetivo do estudo. Os de interpretação/descrição buscam caracterizar um fenômeno (o que é, como se distribui, que sentidos lhe são atribuídos) sem testar explicitamente uma relação causal; os de teste/explicação buscam estabelecer

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.17986

essa relação, estimando a magnitude de um impacto ou explicando o mecanismo pelo qual a intervenção produz um resultado. A segunda dimensão é a abordagem metodológica, que remete a concepções epistemológicas distintas sobre o que conta como conhecimento válido (Goertz & Mahoney, 2012; Silva, 2015). A quantitativa parte de uma epistemologia objetivista, operacionalizada tanto por dados numéricos quanto por dados textuais; a qualitativa, de uma epistemologia construtivista, em que o conhecimento é coproduzido na interação entre pesquisador e objeto.

O cruzamento das duas dimensões gera quatro tipos de estudos (Quadro 2). Os estudos a1 (qualitativo interpretativo) voltam-se aos sentidos, percepções e dinâmicas associadas à intervenção, iluminando o ponto de vista de beneficiários, implementadores e lideranças comunitárias. Os estudos b1 (qualitativo explanatório) têm pretensão explicativa e buscam padrões causais entre condições e resultados; usualmente através da comparação implícita ou explícita (Nunes, 2020). Os estudos a2 (quantitativo descritivo) descrevem características da população beneficiária, da implementação ou do alcance da política, sem estimar efeitos causais. Os estudos b2 (quantitativo explanatório) estimam o efeito causal por estratégias como ensaios randomizados, regressão descontínua, diferenças em diferenças e seleção em observáveis; constituem o núcleo da síntese quantitativa e, por análises de mediação e de subgrupos, também contribuem para identificar mecanismos e moderadores.

Quadro 2: Tipologia transversal dos estudos incluídos na revisão

Objetivo do estudo	**Abordagem qualitativa (1)**	**Abordagem quantitativa (2)**
Interpretação/descrição (a)	a1	a2
Teste/explicação (b)	b1	b2

Fonte: elaborado pelos autores.

A tipologia é um recurso heurístico. Qualquer desenho listado no Quadro 1 pode ser reclassificado em um dos quatro quadrantes, o que permite tratar de forma padronizada a contribuição de cada estudo. A integração entre eles segue lógica sequencial: os estudos b2 respondem se e quanto a política funciona; os demais auxiliam na resposta de como, por que, para quem e em que contexto.

3. A PROPOSTA: REVISÃO SISTEMÁTICA MISTA E SEQUENCIAL

Nossa proposta segue uma lógica sequencial de integração de evidências (Small, 2011; Stern et al., 2020), refletida na estrutura da caixa de ferramentas de avaliação (Quadro 3). Determina-se primeiro se a política funciona; depois, quanto funciona; em

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.17986

seguida, como funciona; em que contextos funciona melhor; qual a percepção dos atores envolvidos; como se dá sua implementação; e, por fim, qual o seu custo.

A sequência funciona como modelo encadeado. Cada dimensão depende logicamente da anterior e adiciona uma camada que a precedente, isoladamente, não poderia oferecer. Saber que a política funciona é pouco útil sem saber o quanto; saber o grau de funcionamento é insuficiente sem entender por qual caminho; e conhecer o mecanismo pouco informa se ele não for qualificado pelas condições em que se manifesta. As dimensões de percepção, implementação e custo ancoram a análise na experiência de quem vivencia, executa e financia a política. A vantagem do desenho é evitar a fragmentação típica de sínteses que respondem apenas se a política funciona. Cada dimensão corresponde a uma pergunta prática que gestores precisam responder para adaptar, e não apenas validar, uma intervenção.

Quadro 3: Caixa de ferramentas de avaliação

Dimensão	O que sintetiza	Tipo de evidência	Tipo de estudo
Escala do efeito	Direção do efeito sobre o resultado: positivo, misto ou negativo.	Quantitativa	b2
Força do efeito	Magnitude do efeito a partir da evidência disponível.	Quantitativa	b2
Mecanismo (como funciona?)	Caminho causal pelo qual a política (X) produz seu efeito (Y).	Quanti. e/ou quali.	a1, b1, b2
Moderador (onde, quando e com quem funciona?)	Condições de contexto que ampliam ou inibem o efeito.	Quanti. e/ou quali.	a1, a2, b1, b2
Implementação (como fazer?)	Como a política é executada, incluindo barreiras e adaptações.	Quanti. e/ou quali.	a1, a2, b1, b2
Percepção (como os atores interpretam a política?)	Sentido atribuído à intervenção por beneficiários, gestores e implementadores.	Qualitativa	a1, a2
Custo (qual o valor?)	Relação entre recursos investidos e resultados produzidos.	Quanti. e/ou quali.	a1, a2, b1, b2

Fonte: elaborado pelos autores.

A escala do efeito identifica a direção dos resultados em cada estudo (positivo, negativo ou nulo) e, combinada à significância estatística, permite uma primeira sistematização dos achados; testes como os de Fisher, Cooper ou Winer verificam se os estudos convergem para um efeito comum (Figueiredo et al., 2014). Saber que a maioria aponta na mesma direção, contudo, é insuficiente para estimar a magnitude. Para isso avança-se à força do efeito, extraindo de cada trabalho o coeficiente do modelo principal, o erro padrão e o valor de T (Harrer et al., 2021). A métrica mais difundida é o d de Cohen; como ele tende a superestimar o efeito populacional em amostras pequenas, o g de Hedges corrige esse viés e constitui o estimador preferencial na literatura de meta-análise. O objetivo é produzir um valor comparável entre estudos e estimar um efeito

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.17986

combinado, cuja heterogeneidade sinaliza em que medida contextos, populações e desenhos distintos produzem resultados diferentes, características que aparecem nos moderadores.

Um mecanismo é o processo ou caminho causal pelo qual a intervenção (X) produz o resultado (Y): uma ligação entre variável causal e resultado, de natureza causal e estrutura identificável (Hedström & Ylikoski, 2010). A distinção entre a existência de um efeito e a explicação de seu mecanismo é fundamental. Um estudo pode demonstrar, com alta validade interna, que o Bolsa Família reduz a mortalidade infantil sem explicar como: pela renda adicional que permite comprar alimentos? Pelo acesso às consultas exigidas como condicionalidade? Pela mudança no status da mulher no domicílio? Cada resposta implica um mecanismo distinto, com implicações diferentes para o desenho da política. Em estudos quantitativos, mecanismos são estimados por análises de mediação (Imai et al., 2011) ou testes com subgrupos; em qualitativos, inferidos de entrevistas, observação participante e triangulação de fontes.

Moderadores são fatores que afetam a magnitude ou a direção do efeito. Diferentemente do mecanismo, que está no caminho causal, o moderador condiciona esse caminho: responde a/para quem, onde e quando a política funciona melhor ou pior (Cintron et al., 2022). Sua identificação informa decisões de focalização e adaptação.

Às quatro dimensões que formam o núcleo analítico somam-se três complementares. A implementação descreve como a política é executada. A literatura sobre *implementation gap* documenta que políticas bem desenhadas podem falhar por problemas de execução (Matland, 1995), mas o hiato não se resume a falhas. Implementadores também modificam, adaptam e reinterpretam a política em curso, discricionariedade típica da burocracia de nível de rua (Lotta, 2018), produzindo variação de efeito entre unidades que nenhum modelo de efeito médio captura. A percepção dos atores busca compreender como a intervenção é interpretada por beneficiários, gestores e implementadores. Como argumentam Batista e Domingos (2017), o padrão de significação dos atores faz com que a percepção de um mesmo efeito varie: o que para um gestor é condicionalidade razoável pode ser vivenciado pelo beneficiário como controle, estigma ou reconhecimento de um direito. O custo, por fim, examina a relação entre recursos investidos e resultados, registrando custos financeiros diretos, de implementação e não monetários, como custos sociais e políticos.

Tomadas em conjunto, as sete dimensões oferecem uma síntese estruturada e abrangente da evidência disponível. É a transparência sobre o que se sabe, o que não se

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.17986

sabe e com que grau de confiança que confere à caixa de ferramentas seu valor como instrumento de apoio à decisão.

4. APLICAÇÃO: O PROGRAMA BOLSA VERDE

A proposta se operacionaliza em um protocolo de seis etapas: planejamento, busca, seleção, sistematização, análise e divulgação, podendo ainda ser precedidas por uma etapa exploratória opcional (ver Anexo 9). Percorremos essas etapas acompanhando cada uma de sua execução em um caso concreto: a avaliação do Programa de Apoio à Conservação Ambiental, denominado como Bolsa Verde (Lei nº 12.512/2011), que transferia renda a famílias em situação de extrema pobreza residentes em áreas protegidas e territórios tradicionais, mediante compromisso de conservação ambiental. O passo-a-passo detalhado, com o código e as bases intermediárias, está disponível publicamente[2].

4.1 Planejamento

O planejamento condiciona todas as demais etapas: é nele que definimos a pergunta e declaramos o que estamos dispostos a considerar evidência. Uma pergunta mal delimitada compromete a revisão, pois nenhuma sofisticação analítica posterior corrige um escopo mal definido desde o início (Xiao & Watson, 2019). Formulamos a nossa como: o Bolsa Verde (X) reduz a extrema pobreza (Y1) e/ou o desmatamento (Y2)? O programa foi escolhido por englobar três fatores principais: 1) combinar transferência de renda condicionada com obrigações de conservação ambiental, caso relativamente raro de política socioambiental integrada; 2) ter sido descontinuado, o que permite análises contrafactuais *ex post* sobre um período de vigência bem delimitado; e 3) pela escassez de revisões que tratassem simultaneamente dos dois resultados de interesse. Uma busca exploratória preliminar, em ferramenta assistida por inteligência artificial[3] (Lassance, 2026), confirmou que já existe um campo de avaliação minimamente maduro, sobretudo quanto ao desmatamento.

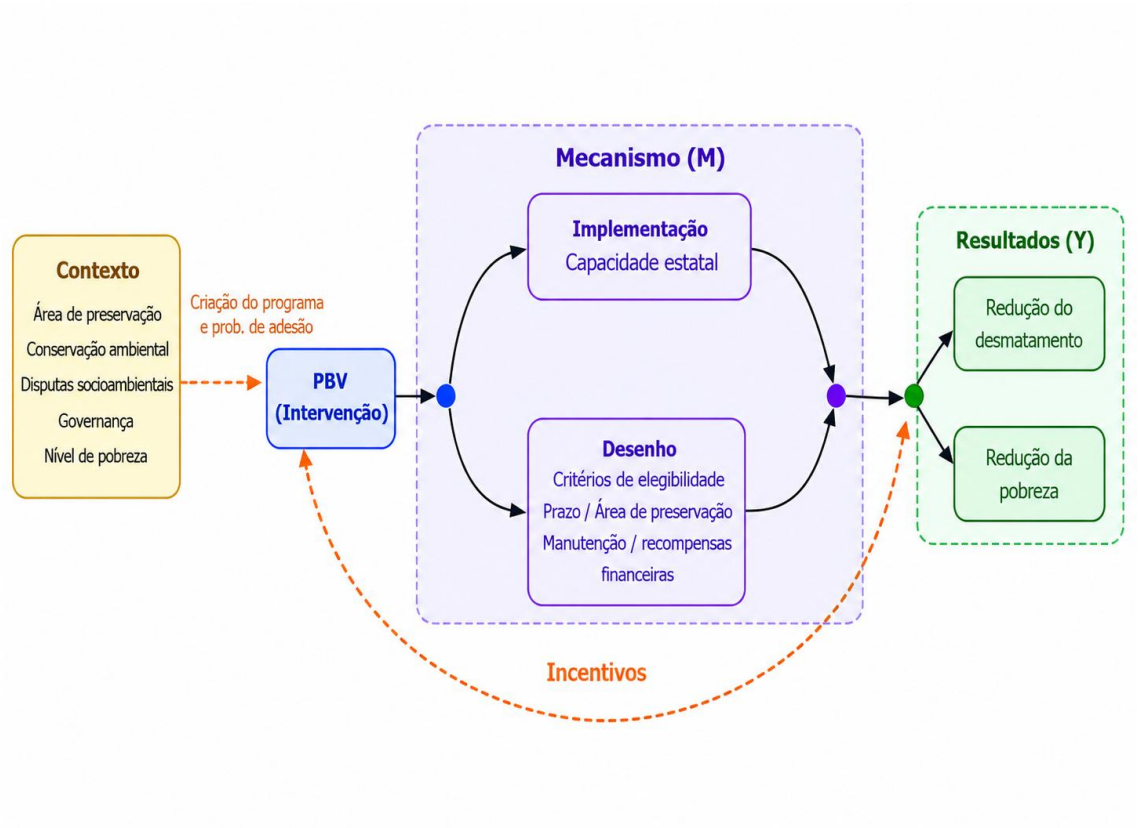
A partir daí construímos uma teoria do programa. Ao explicitar quais mecanismos e moderadores esperamos encontrar, definimos quais variáveis o formulário de extração precisa capturar. Uma boa teoria do programa descreve como a política deveria funcionar, formula hipóteses concorrentes de por que ela pode não funcionar como pretendido, e

[2] https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/1MVLGR
[3] Usamos a ferramenta Consensus.app.

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.17986

realiza a comparação entre essas hipóteses, o que confere à revisão sua pretensão explicativa (Dalkin et al., 2015; Pawson et al., 2005). O passo seguinte é a construção de um diagrama causal (Figura 1), que articula a intervenção aos dois resultados por meio dos mecanismos previstos na legislação e dos moderadores contextuais identificados na literatura de pagamentos por serviços ambientais, sobretudo a segurança de posse da terra e a institucionalidade local. Separamos dois componentes do mecanismo que julgamos analiticamente distintos: a dimensão de desenho, o incentivo formal estabelecido pela legislação, e a de implementação, o *enforcement* efetivamente exercido no território.

Figura 1: Teoria do programa do Bolsa Verde em diagrama causal



Fonte: elaborado pelos autores a partir da Lei nº 12.512/2011 e do Decreto nº 7.572/2011.

Traduzimos a teoria do programa em estruturas PICOC e CMO (ver Anexo 10) e fixamos a *priori* os critérios de inclusão e exclusão, a saber: estudos empíricos sobre a intervenção e seus resultados no contexto brasileiro, entre 2011 e 2025, excluindo trabalhos normativos, teóricos ou baseados exclusivamente em análise de jurisprudência. Consolidamos o protocolo resultante e o pré-registramos no Open Science Framework[4] (Schaefer et al., 2026).

[4] O pré-registro pode ser acessado em: http://osf.io/b792u.

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.17986

4.2 Busca

A busca converte a pergunta estruturada em estratégia de recuperação de evidências. Pequenas variações na *string* podem incluir ou excluir dezenas de estudos; por isso tratamos cada refinamento como evento a ser registrado, e não apenas a versão final. Testamos e refinamos a *string* em quatro rodadas progressivas na base OpenAlex até equilibrarmos sensibilidade e especificidade (Priem et al., 2022). Percebemos a necessidade de um filtro de país apenas depois que os primeiros testes mostraram alto volume de ruído vindo de estudos sobre pagamentos por serviços ambientais em outros países. Conduzimos as buscas formais em seis fontes (Tabela 1) e normalizamos os campos bibliográficos entre bases. É importante ressaltar que algumas bases de dados não aceitam operadores booleanos complexos, o que torna necessário simplificar as expressões para garantir a recuperação adequada dos resultados. O volume relativamente pequeno de registros brutos é coerente com a especificidade do programa: escopo geográfico restrito, população-alvo pequena e janela temporal curta.

Tabela 1: Bases consultadas, termos de busca e registros recuperados

Base	Termos de busca	Filtros	N
Web of Science	String completa	Nenhum	1
Web of Science/SciELO	String completa	Nenhum	1
SciELO	"bolsa verde"	Nenhum	8
BDTD/CAPES	"Bolsa Verde"	Área temática (humanidades e sociais aplicadas); ano (2011–2022)	12
OpenAlex	String completa	Nenhum	15
Ipea	"bolsa verde"	Nenhum	7
Total			44

Fonte: elaborado pelos autores. Nota: a string completa combina três blocos ligados por AND — intervenção ("Bolsa Verde" OR "Green Grant" OR "Green Basic Income" OR "Green Cash Transfer"), resultados (termos de pobreza e de desmatamento em português, inglês e espanhol) e país ("Brazil" OR "Brasil"). A string integral e as adaptações por base estão no Anexo 11. Buscas realizadas em abril de 2026.

4.3 Seleção

Na seleção, a leitura substantiva passa a determinar quais trabalhos entram na revisão, e a realizamos em dois momentos. Na filtragem formal, aplicamos um funil de filtros objetivos: remoção de duplicatas por DOI normalizado e título limpo, ano (a partir de 2011), idioma e menção a método no resumo, título ou palavras-chave, este último por dicionários em três idiomas. Documentamos o número de registros remanescentes a cada filtro, informação que alimenta o diagrama PRISMA (Figura 2).

Sobre a base resultante aplicamos a filtragem substantiva, que depende da leitura de títulos e resumos e é, por isso, a etapa mais suscetível ao viés de um único avaliador. Nela, incorporamos uma inovação presente na literatura recente sobre triagem assistida

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.17986

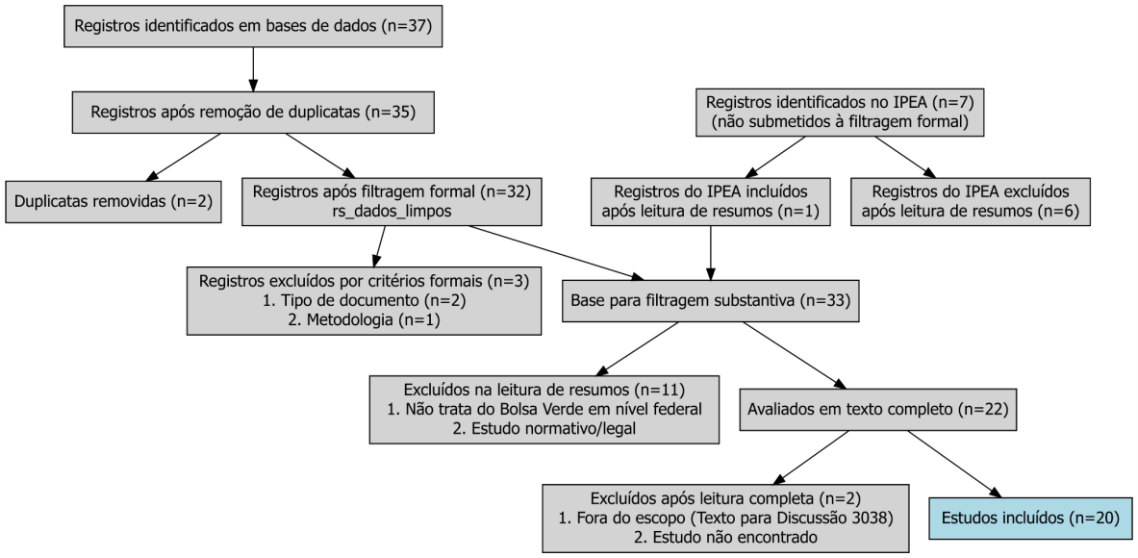
por inteligência artificial: empregar modelos de linguagem como agentes de validação, submetendo cada resumo a múltiplos modelos independentes cujas decisões são combinadas e confrontadas com a leitura humana (Moens et al., 2025; Wagner et al., 2026). Submetemos os 32 registros a três modelos em paralelo, cada um recebendo o mesmo *prompt* padronizado com os critérios de inclusão e exclusão e produzindo decisão binária com justificativa textual. Combinamos os pareceres por regra de maioria e encaminhamos à revisão manual os casos de empate ou de ausência de resposta.

Verificamos a validade do procedimento por leitura humana independente, realizada por dois avaliadores, calculando o Kappa de Cohen entre pares. Considerando os sete pares, o acordo teve média de 75% (de 50% a 91%); o Kappa de Cohen médio foi de 0,50 (moderado, de 0,04 a 0,82). Entre os dois avaliadores humanos, o Kappa foi de 0,72, e o modelo de melhor desempenho alcançou 0,63 (Sonnet 4.6). O erro mais comum dos modelos foi considerar, mesmo com a especificação do *prompt*, estudos normativos como válidos para leitura completa, bem como a seleção de textos sobre políticas estaduais e não o PBV federal (Bolsa Verde de Minas Gerais, Lei Estadual nº 17.727/2008 e Decreto nº 45.113/2009).

Dos 32 registros triados, os avaliadores humanos convergiram pela inclusão de 19 e pela exclusão de 9, restando 4 divergências, resolvidas por discussão e encaminhadas à exclusão. A esse conjunto, somaram-se três documentos do repositório do Ipea que atendiam aos critérios e não haviam entrado na base normalizada. Dos 22 trabalhos levados à leitura completa, um foi excluído por estar fora do escopo e outro não foi localizado, restando 20 estudos (ver Anexo 8).

Figura 2: Fluxo de seleção dos estudos (PRISMA)

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.17986



Fonte: elaborado pelos autores.

4.4 Sistematização

A sistematização operacionaliza a principal inovação da proposta. Ao invés de aplicarmos um formulário único a todos os estudos, diferenciamos a extração conforme o tipo metodológico de cada trabalho. Evitamos, assim, dois riscos: perder informação qualitativa relevante ao tentar quantificar o que não é quantificável, ou descartar estudos qualitativos por não se encaixarem em uma métrica de efeito.

Uma seção comum de identificação é preenchida para todos os estudos; a partir daí o formulário se ramifica. Para os estudos a1, os campos se voltam ao mecanismo identificado (a pergunta orientadora é sempre "qual é o caminho de X para Y na análise do autor?"), às condições contextuais que o ativam ou inibem, à percepção dos atores e às barreiras de implementação; para os a2, à fonte e ao tipo de amostragem, à população-alvo e ao percentual atendido; para os b1, à formalização da análise comparativa e à seleção de casos; e, para os b2, à estratégia de identificação, ao estimador, à direção e magnitude do efeito e aos testes de mecanismo e heterogeneidade. O instrumento completo, com as 115 variáveis distribuídas nos cinco blocos, está no Anexo 7.

Dos 20 trabalhos, classificamos 11 (55%) como a1, 5 (25%) como a2 e 4 (20%) como b2; nenhum foi classificado como b1. Os estudos que avaliaram o programa, portanto, majoritariamente descrevem a política e os sentidos que os stakeholders lhe atribuem.

4.5 Análise

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.17986

A análise transforma os dados extraídos em síntese, organizada segundo as sete dimensões da caixa de ferramentas (Tabela 3), na ordem em que recomendamos que sejam lidas. Reportamos aqui os achados principais; a extração completa, por estudo e por dimensão, está no material suplementar.

Quatro estudos testam explicitamente o efeito causal do programa sobre o desmatamento; nenhum estuda seu efeito sobre a extrema pobreza. Três reportam efeito significativo a 5% (Wong et al. (2026): $z = -3{,}49$; Costedoat et al. (2022): $z = -2{,}36$; Nascimento (2021): $z = 5{,}43$) e Andrade (2018) não rejeita a hipótese nula. Entre os três significativos, a direção diverge. Wong et al. (2026) e Costedoat et al. (2022) encontram redução do desmatamento; Nascimento (2021) encontra aumento. Optamos por não submeter essa divergência a testes formais de combinação: com quatro estudos, tanto as estatísticas de combinação quanto a contagem de votos têm poder muito baixo para distinguir um padrão real de uma configuração ao acaso. Recorremos, em vez disso, à Maryland Scale para ponderar os estudos por robustez de identificação (Kopittke & Ramos, 2021). Os dois com desenho mais robusto convergem na direção prevista pela teoria do programa. Nascimento (2021) emprega painel de grande N, mas sem pareamento, grupo de comparação deliberado ou comparação temporal, limitações que a própria autora reconhece; e Andrade (2018), com robustez comparável, não encontra efeito, mas trata de desfecho distinto do desmatamento propriamente dito. Ponderando por robustez, e não apenas pela convergência de sinais, diagnosticamos que o programa é efetivo na redução do desmatamento (Y2).

Com relação à força do efeito, o modelo de efeitos aleatórios produz $g = 0{,}125$ (IC 95%: −0,261; 0,511), não significativo, com $I^2 = 91{,}0\%$. Isto sugere que praticamente toda a dispersão observada decorre de os estudos estimarem quantidades distintas, e não de erro amostral. Tratar os quatro como estimativas ruidosas de um mesmo efeito é pressuposto que os próprios dados rejeitam, e por isso não apresentamos o efeito agrupado como estimativa central. Reportamos, em vez disso, o Common Language Effect Size de cada estudo (Tabela 2), que converte o g de Hedges na probabilidade de que uma observação sorteada do grupo tratado apresente resultado mais favorável à conservação do que uma do grupo controle. Assim se evita comunicar um intervalo agregado que cruza zero como se fosse efeito nulo com muita incerteza. Os efeitos são de magnitude pequena a moderada e sua direção depende de onde e como é medido. Além disso, os dois estudos com identificação mais robusta são também os que apresentam CLES mais distantes de 50% na direção favorável à conservação.

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.17986

Tabela 2: Tamanho de efeito e probabilidade de superioridade (CLES) por estudo

Estudo	Estratégia de identificação	g de Hedges	CLES
Wong et al. (2026)	Efeitos fixos de duas vias	0,423	61,7%
Costedoat et al. (2022)	Diferenças em diferenças	0,272	57,6%
Andrade (2018)	Seleção em observáveis com pareamento	0,014	50,4%
Nascimento (2021)	Painel administrativo com efeitos fixos	−0,116	46,7%

Fonte: elaborado pelos autores. Nota: o CLES indica a probabilidade de que uma observação sorteada do grupo tratado apresente resultado mais favorável à conservação do que uma do grupo controle. Valores acima de 50% indicam vantagem do grupo tratado. Modelo de efeitos aleatórios: g = 0,125 (IC 95%: −0,261; 0,511); I^2 = 91,0%.

O mecanismo previsto pela teoria previa o alívio de restrição financeira e a mediação pela institucionalidade local. O alívio financeiro aparece de modo recorrente nos relatos de beneficiários, que descrevem o benefício como complemento de uma renda instável (Cantanhêde et al., 2018; Graciano et al., 2018; MacNeill & Drummond, 2025). O segundo mecanismo mostrou-se mais heterogêneo do que a teoria antecipava. Em Kasanoski (2018) ele opera por reforço normativo, quando há liderança comunitária atuante, e se dissolve quando a associação está desmobilizada. Já o único estudo b2, que testa mecanismo formalmente (Wong et al., 2026), encontra um canal informacional, com autuações ambientais mais distantes de estradas e de alertas de satélite, o que sugere que as comunidades funcionam como sistema informal de informação sobre desmatamento ilegal. A teoria previa a institucionalidade local como constructo unificado. No entanto, a evidência indica ao menos dois canais independentes. Um terceiro achado, em Andrade (2018), reforça esse limite, uma vez que o programa aumenta a probabilidade de o beneficiário declarar conhecimento das regras territoriais, sem efeito sobre a extração madeireira. Este mesmo padrão é descrito qualitativamente por Novak (2018), o qual retrata beneficiários que recebem o benefício sem compreender sua condicionalidade. A ausência de cartão próprio e o pagamento processado junto ao do Bolsa Família sugerem que a separação entre desenho e implementação é acessória. A própria decisão de desenho compromete, desde a origem, a legibilidade da condicionalidade ambiental.

Quanto aos moderadores do efeito, a teoria antecipava a relevância da segurança de posse da terra e a pressão de arrendatários, e a leitura conjunta confirma parcialmente a expectativa. Costedoat et al. (2022) encontram efeito significativo apenas em assentamentos na Amazônia, contexto de risco de desmatamento moderado. Os estudos do tipo a1 de Novak (2018) e de Novak e Gonzaga (2024) identificam posse da terra e pressão de arrendatários como moderadores não da resposta durante a vigência, mas da velocidade de reconversão das áreas depois que o benefício acaba: onde há alta rotatividade e posse fraca, a reconversão acelera; onde há coesão comunitária, é freada mesmo após o fim do pagamento. A teoria foi desenhada para explicar o efeito do

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.17986

programa em vigência e não previu esse componente de sustentabilidade pós-descontinuação, considerado um limite relevante, já que o programa foi descontinuado. Dois estudos a2 (Viana, 2013, 2014) revelam ainda uma regularidade não incorporada. A cobertura nunca se aproximou da população-alvo, alcançando 44.082 das 73 mil famílias previstas (60,4%) em dezembro de 2013. Os efeitos estimados dizem respeito, portanto, a uma fração pequena e possivelmente não aleatória dessa população, o que é, em si, um moderador de escala.

Por fim, entramos nas dimensões complementares da avaliação: a percepção, implementação e custos da política. A regularidade mais forte entre os estudos a1 é que o programa é percebido como complemento de renda, não como incentivo ambiental. Em Kasanoski (2018), 64% dos beneficiários o descrevem como ajuda extra vinculada ao Bolsa Família (Y1) e apenas 21,5% o associam categoricamente à conservação (Y2). A cadeia prevista pela teoria (o termo de adesão formaliza o compromisso e este, uma vez compreendido, orienta comportamento) é interrompida no estágio de informação. Há, contudo, uma nuance. Em Moreira (2017), boa parte das beneficiárias entende o benefício como direito, e não como favor do governo, dimensão de legitimidade que a teoria não contemplava.

Quanto à implementação, os estudos convergem no diagnóstico de fragilidade institucional que apontam para a ausência de identidade própria, pagamento processado junto ao sistema do Bolsa Família, monitoramento raro e dificuldade de acesso a áreas remotas, com a adesão mediada pelo órgão gestor e operacionalizada por associações comunitárias que assumem sobrecarga não prevista (Queiroz e Silva, 2016). A institucionalidade local, que a teoria tratou como condição de contorno estável, é na prática o próprio canal pelo qual a implementação se processa.

Em relação aos custos, o orçamento decresce ano a ano (Novak & Gonzaga, 2024): R$ 106,1 milhões em 2015, R$ 78 milhões em 2016 e R$ 72,2 milhões em 2017. Apenas os dois estudos b2 mais robustos trazem análise de custo-efetividade, na forma de custo por tonelada de CO_2 evitada. Wong et al. (2026) estimam abatimento a cerca de US$ 8,6/tCO_2, retorno ambiental de cerca de 2,8 vezes o custo do programa; Costedoat et al. (2022), entre US$ 1,73 e US$ 3,84 por tonelada. Custos sociais, atrito intrafamiliar gerado pela individualização do benefício, insatisfação quanto ao valor e à periodicidade, perda de serviços ecossistêmicos com a descontinuação, também são citados como custos, mas descritos apenas qualitativamente, não em valores monetários.

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.17986

4.6 Escrita e divulgação

A última etapa trata da escrita e da divulgação. Uma característica da proposta é que a síntese pode assumir diferentes formatos conforme o público, num esforço deliberado de tradução do conhecimento: para a comunidade acadêmica, o artigo de revisão acompanhado do diagrama PRISMA; para gestores e formuladores, a própria caixa de ferramentas, que organiza as evidências de forma resumida, padronizada e acionável e pode ser desdobrada em produtos mais curtos, como *policy briefs*. Neste artigo, resumimos as principais conclusões através da caixa de ferramentas.

Tabela 3: Caixa de ferramentas

Dimensão	**Pergunta**	**Base**	**O que a evidência permite afirmar**
Escala do efeito	A política funciona?	4 de 20 (b2)	O programa é efetivo na redução do desmatamento.
Força do efeito	Quanto funciona?	1 de 4 (b2)	O efeito é moderado sobre o desmatamento.
Mecanismo	Como funciona?	13 de 20	O mecanismo predominante identificado é a condicionalidade monetária operando como norma coletiva de conservação, com fiscalização mútua no território, e não como incentivo individual.
Moderador	Onde, quando e com quem funciona?	14 de 20	Moderadores recorrentes: coesão social e qualidade da liderança comunitária; proximidade física da sede da associação; bioma (Amazônia vs. fora) e regime de posse da terra (assentamentos vs. áreas protegidas); distância de estradas; e registro no CAR. O efeito sobre desmatamento é maior em áreas não povoadas e fora do CAR. Inibidores: baixa comunicação institucional, ausência de capacitação ambiental e de monitoramento técnico.
Implementação	Como fazer?	10 de 20	Barreiras recorrentes: isolamento geográfico dos beneficiários, que dificulta tanto o acesso ao benefício quanto a própria avaliação; linguagem pouco acessível do Termo de Adesão; ausência de visitas técnicas e de monitoramento; critérios de elegibilidade por renda per capita que excluem parte das famílias elegíveis pela unidade de conservação, gerando conflito comunitário; e vulnerabilidade a mudança política, que culminou na descontinuação em janeiro de 2018 por ausência de previsão na LOA.
Percepção	Como os stakeholders interpretam a política?	9 de 20	Achado convergente entre os estudos qualitativos: os beneficiários interpretam o Bolsa Verde como transferência de renda, não como pagamento por serviço ambiental. Quando questionados sobre os incentivos, mencionam o componente econômico e não o ambiental. O recebimento simultâneo com o Bolsa Família, de valor maior, reforça essa leitura.
Custo	Qual o valor?	14 de 20	Benefício de R$ 300,00 por trimestre (R$ 100,00/mês) por família. Orçamento federal decrescente: R$ 91,2 milhões (2013), R$ 106,2 milhões (2014), R$ 106,1 milhões (2015), R$ 78 milhões (2016), R$ 72,2 milhões (2017, último ano). Custo total do programa estimado em USD 135 milhões (USD constantes de 2016). O programa foi custo-efetivo em emissões de CO2 evitadas.

Fonte: Elaboração própria.

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.17986

Ao tornar público os achados, o protocolo e o código que os produziram, a revisão contribui para reduzir a assimetria de informação entre as comunidades acadêmica e decisória.

5. DISCUSSÃO: VANTAGENS E LIMITES

A política surge em um contexto político e institucional específico (durante o primeiro mandato de Dilma Rousseff) e sofre descontinuidades a partir do governo Michel Temer, consolidando o seu desmantelamento no governo Jair Bolsonaro. Em 2023, o programa foi reativado com mudanças relevantes de desenho (ampliação da elegibilidade, redirecionamento territorial, aumento do benefício trimestral e digitalização do cadastro (Brasil, 2024)), alterações que ficam fora do escopo temporal desta revisão, mas que abrem uma janela para aplicação prospectiva do protocolo a essa nova fase do programa.

A aplicação da revisão sistemática mista e sequencial ao Programa Bolsa Verde ilustra o que a proposta permite fazer e o que ela não resolve. Uma das vantagens é tornar visível a distribuição desigual da evidência entre as dimensões da avaliação. Dos 20 estudos revisados, apenas 4 estimam efeito causal, todos sobre o desmatamento (Y2), enquanto 14 informam custos e 11 tratam de percepções e implementação. Uma meta-análise convencional descartaria dezesseis desses trabalhos e produziria uma estimativa agregada não significativa e altamente heterogênea. A revisão mista permite afirmar que o programa reduziu o desmatamento (Y2) nos contextos em que foi mais bem avaliado, que essa redução foi custo-efetiva, que o mecanismo previsto pela legislação não operou como esperado porque a condicionalidade ambiental nunca se tornou legível ao beneficiário, e que a cobertura efetiva jamais se aproximou da população-alvo.

Um achado importante da aplicação, contudo, é uma ausência. Não há, na literatura recuperada, nenhum estudo com estratégia de identificação causal sobre o efeito do programa na extrema pobreza (Y1), que era um dos dois objetivos declarados da política. Identificar lacunas dessa natureza é função da revisão sistemática tão relevante quanto sintetizar o que já se sabe, e só se cumpre quando o protocolo distingue o que não foi encontrado do que não foi procurado.

A proposta, no entanto, não é isenta de limitações. Em primeiro lugar, revisões sistemáticas são tão boas quanto os estudos que revisam. Se a literatura primária é escassa, metodologicamente frágil ou concentrada em contextos muito específicos, a revisão não

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.17986

superará essas deficiências. Ela não substitui avaliações originais; pelo contrário, depende delas para existir. Em segundo lugar, o tempo de realização varia com o tamanho da equipe, a amplitude da pergunta e o volume de literatura, o que as torna instrumentos mais adequados ao planejamento de políticas do que a demandas urgentes de gestão. Em terceiro, sua qualidade depende criticamente da formulação da pergunta. Perguntas amplas demais geram revisões heterogêneas e de difícil síntese; restritas demais, corpus insuficiente.

Uma quarta limitação diz respeito ao uso de modelos de linguagem na triagem. Os níveis moderados de concordância que observamos indicam que a triagem assistida por inteligência artificial reduz o custo da leitura de resumos e a dependência das idiossincrasias de um avaliador isolado, mas não dispensa a leitura humana nem a documentação das divergências. O erro mais frequente dos modelos (aceitar estudos normativos e trabalhos sobre políticas homônimas) exige familiaridade substantiva com o campo, e não apenas com os critérios formais do protocolo. Tratamo-los, por isso, como avaliadores adicionais cujo desempenho é medido e reportado, e não como substitutos do julgamento da equipe.

Por fim, é importante ter em mente que a classificação proposta em quatro tipos não é uma taxonomia exaustiva. Trabalhos que combinam desenhos ou não explicitam sua estratégia analítica exigem decisão do codificador, que deve ser documentada. A saída mais adequada é a classificação do estudo no tipo que mais se aproxima das suas características, com estudos com desenhos e métodos mistos podendo ser classificados em mais de um tipo simultaneamente. O critério deve ser sempre o de extrair o máximo de variáveis relevantes que o estudo tem a oferecer.

6. CONSIDERAÇÕES FINAIS

Este artigo apresentou uma proposta de revisão sistemática mista e sequencial para a avaliação de políticas públicas, que parte de um diagnóstico simples: a pergunta "quais políticas funcionam?" é necessária, mas insuficiente para quem precisa decidir (Spohn et al, 2026). Gestores e formuladores não escolhem entre implementar ou não uma intervenção em abstrato; escolhem como desenhá-la, onde priorizá-la, com que recursos e a partir de que arranjos institucionais. Responder a isso exige integrar, num mesmo protocolo, evidências que a hierarquia convencional trata como pertencentes a patamares distintos de credibilidade.

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.17986

Nesse sentido, a contribuição do artigo reside na operacionalização da metodologia. Propusemos três instrumentos articulados: uma tipologia transversal que classifica os estudos incluídos pelo cruzamento entre objetivo e abordagem; um formulário de extração que se ramifica conforme essa classificação, em vez de aplicar o mesmo instrumento a trabalhos incomparáveis; e uma caixa de ferramentas que organiza os achados em sete dimensões e explicita, para cada uma, quantos estudos a sustentam. É essa última informação que transforma a síntese em instrumento de decisão. Dizer que uma dimensão se apoia em quatro estudos e outra em catorze é dizer, com precisão, onde a recomendação é firme e onde é provisória.

A aplicação ao Bolsa Verde não pretende ser avaliação definitiva do programa, porém fornece a demonstração do protocolo em condições reais, inclusive nas suas dificuldades. Ela mostrou um programa efetivo na redução do desmatamento nos contextos mais bem avaliados e custo-efetivo em termos de carbono evitado (Y2), mas cujo mecanismo central não operou como a legislação previa e cuja cobertura nunca alcançou a população-alvo. Mostrou, sobretudo, que um dos dois objetivos declarados da política jamais foi avaliado com estratégia de identificação causal. Um programa descontinuado sem que se soubesse se cumpriu metade daquilo a que se propunha é, em si, argumento a favor de protocolos de síntese que registrem tanto o que a evidência mostra quanto o que ela silencia.

Duas agendas se abrem a partir daqui. A primeira é metodológica, visto que o uso de modelos de linguagem (LLM) na triagem e na extração avança rapidamente, e a comunidade precisa de parâmetros compartilhados sobre como documentar versões, *prompts*, *datas* e níveis de concordância, sob pena de tornar irreproduzível justamente a etapa que a revisão existe para tornar auditável. A segunda é de nível institucional, caracterizada pela discussão em curso sobre avaliação obrigatória de programas federais, oferecendo uma janela para que a síntese de evidências deixe de ser exercício acadêmico e passe a integrar o ciclo ordinário de decisão sobre políticas públicas no país.

REFERÊNCIAS

Abreu, L. A. R. de, et al. (2025). Bolsas de estudo e evasão: avaliação de impacto ex-ante. Autografia.

Andrade, B. T. (2018). Has the Green Grant Program had an impact on conservation in Brazil? [Dissertação de mestrado]. Universidade de São Paulo.

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.17986

Azeredo, T. B., Luiza, V. L., & Baptista, T. W. de F. (2012). Políticas públicas e avaliação de implementação. BIB — Revista Brasileira de Informação Bibliográfica em Ciências Sociais, (74), 7–25.

Bamberger, M. (2012). Introduction to mixed methods in impact evaluation. InterAction; The Rockefeller Foundation.

Batista, M., & Domingos, A. (2017). Mais que boas intenções: técnicas quantitativas e qualitativas na avaliação de impacto de políticas públicas. Revista Brasileira de Ciências Sociais, 32(94), e329414. https://doi.org/10.17666/329414/2017.

Batista, M. (2025). Evidências e tomada de decisão em políticas públicas: uma análise qualitativa a partir da perspectiva da burocracia dirigente. Revista de Administração Pública, 59(2), e2024-0202.

Brady, H. E. (2011). Causation and explanation in social science. In R. E. Goodin (Ed.), The Oxford handbook of political science (pp. 1054–1107). Oxford University Press.

Brasil (2024). *Relatório de avaliação ex ante - Plano de Monitoramento e Avaliação do Programa Bolsa Verde*. Brasília: Governo Federal, Ministério do Planejamento e Orçamento, Secretaria de Monitoramento e Avaliação de Políticas Públicas e Assuntos Econômicos. Disponível em: https://www.gov.br/planejamento/pt-br/assuntos/avaliacao-de-politicas-publicas/arquivos/avaliacoes-ex-ante/plano-de-monitoramento-e-avaliacao-do-programa-bolsa-verde.

Buller, A. M., Peterman, A., Ranganathan, M., Bleile, A., Hidrobo, M., & Heise, L. (2018). A mixed-method review of cash transfers and intimate partner violence in low-and middle-income countries. The World Bank Research Observer, 33(2), 218-258.

Cantanhêde, L. G., et al. (2018). Environmental perception of fishermen: Use and conservation of fisheries resources. Biota Neotropica, 18, e20170510.

Cintron, D. W., Adler, N. E., Gottlieb, L. M., Hagan, E., Tan, M. L., Vlahov, D., Glymour, M. M., & Matthay, E. C. (2022). Heterogeneous treatment effects in social policy studies: An assessment of contemporary articles in the health and social sciences. Annals of Epidemiology, 70, 79–88. https://doi.org/10.1016/j.annepidem.2022.01.001

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.17986

Costedoat, S., et al. (2022). Cost-effective climate mitigation via conservation incentives targeting poverty: Bolsa Verde, Brazil [Pré-print]. SSRN.

Covidence. (2025). Understanding review types in research (1a ed.). Veritas Health Innovation.

Dacombe, R. (2018). Systematic reviews in political science: What can the approach contribute to political research? Political Studies Review, 16(2), 148–157. https://doi.org/10.1177/1478929917695340

Dalkin, S. M., Greenhalgh, J., Jones, D., Cunningham, B., & Lhussier, M. (2015). What's in a mechanism? Development of a key concept in realist evaluation. Implementation Science, 10, 49. https://doi.org/10.1186/s13012-015-0237-x

Dixon-Woods, M., Agarwal, S., Jones, D., Young, B., & Sutton, A. (2005). Synthesising qualitative and quantitative evidence: A review of possible methods. Journal of Health Services Research & Policy, 10(1), 45–53. https://doi.org/10.1177/135581960501000110

Faria, C. A. P. de. (2022). O movimento das políticas públicas baseadas em evidências: uma radiografia crítica. Revista Brasileira de Ciência Política, (39), e256196.

Farrington, D. P., Gottfredson, D. C., Sherman, L. W., & Welsh, B. C. (2002). The Maryland Scientific Methods Scale. In L. W. Sherman, D. P. Farrington, B. C. Welsh, & D. L. MacKenzie (Eds.), Evidence-based crime prevention (pp. 13–21). Routledge.

Figueiredo Filho, D., Paranhos, R., Silva, J. A. da, Rocha, E. C. da, & Alvarez, R. C. (2019). Seven reasons why: A user's guide to transparency and reproducibility. Brazilian Political Science Review, 13(2), e0001. https://doi.org/10.1590/1981-3821201900020001

Figueiredo, D., et al. (2014). O que é, para que serve e como se faz uma meta-análise? Teoria e Pesquisa, 23(2), 205–228.

Figueiredo, M. F., & Figueiredo, A. C. (1986). Avaliação política e avaliação de políticas: um quadro de referência teórica. Análise & Conjuntura, 1(3), 107–127.

Gerring, J. (1999). What makes a concept good? A criterial framework for understanding concept formation in the social sciences. Polity, 31(3), 357–393. https://doi.org/10.2307/3235246

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.17986

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. Educational Researcher, 5(10), 3–8. https://doi.org/10.3102/0013189X005010003

Goertz, G., & Mahoney, J. (2012). A tale of two cultures: Qualitative and quantitative research in the social sciences. Princeton University Press.

Graciano, M. C., et al. (2018). Efetividades do Programa Bolsa Verde no Assentamento Canudos em Goiás: uma análise da segurança alimentar e da preservação ambiental. Interações (Campo Grande).

Harrer, M., Cuijpers, P., Furukawa, T. A., & Ebert, D. D. (2021). Doing meta-analysis with R: A hands-on guide. Chapman and Hall/CRC.

Hedström, P., & Ylikoski, P. (2010). Causal mechanisms in the social sciences. Annual Review of Sociology, 36, 49–67. https://doi.org/10.1146/annurev.soc.012809.102632

Hunter, W., & Sugiyama, N. B. (2014). Transforming subjects into citizens: Insights from Brazil's Bolsa Família. Perspectives on Politics, 12(4), 829–845. https://doi.org/10.1017/S1537592714002151

Imai, K., Keele, L., Tingley, D., & Yamamoto, T. (2011). Unpacking the black box of causality: Learning about causal mechanisms from experimental and observational studies. American Political Science Review, 105(4), 765–789. https://doi.org/10.1017/S0003055411000414

Kano, H., & Hayashi, T. I. (2021). A framework for implementing evidence in policymaking: Perspectives and phases of evidence evaluation in the science-policy interaction. Environmental Science & Policy, 116, 86–95. https://doi.org/10.1016/j.envsci.2020.11.003

Kasanoski, D. S. (2018). Bolsa Verde: benefício social e incentivo à conservação ambiental? [Dissertação de mestrado, Universidade de Brasília].

Keele, L. (2015). The statistics of causal inference: A view from political methodology. Political Analysis, 23(3), 313–335. https://doi.org/10.1093/pan/mpv007

Koga, N. M., et al. (2020). Análise de políticas públicas e uso de evidências pelas burocracias: proposta analítica e evidências do caso brasileiro (Texto para Discussão No. 2560). Ipea.

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.17986

Koga, N. M., et al. (2024). Apresentação — Boletim de Análise Político-Institucional: governança e cultura de evidências. Ipea.

Kopittke, A. L. W., & Ramos, M. P. (2021). O que funciona e o que não funciona para reduzir homicídios no Brasil: uma revisão sistemática. Revista Brasileira de Segurança Pública, 15(1), 90–113.

Lassance, A. (2026). Manual de uso de ferramentas de inteligência artificial em apoio à análise ex ante de políticas públicas. Ipea.

Lotta, G. (2018). Burocracia, redes sociais e interação: uma análise da implementação de políticas públicas. Revista de Sociologia e Política, 26(66), 145–173.

Ludwig, J., Kling, J. R., & Mullainathan, S. (2011). Mechanism experiments and policy evaluations. Journal of Economic Perspectives, 25(3), 17–38. https://doi.org/10.1257/jep.25.3.17

MacDonald, H., et al. (2024). Searching for studies: A guide to information retrieval for Campbell systematic reviews. Campbell Systematic Reviews, 20(1), e1384.

MacNeill, T., & Drummond, C. (2025). Green basic income: Evaluating the Bolsa Verde project in the Brazilian Amazon. Basic Income Studies, 20(1), 125–149.

Matland, R. E. (1995). Synthesizing the implementation literature: The ambiguity-conflict model of policy implementation. Journal of Public Administration Research and Theory, 5(2), 145–174.

Melo, R. da M. S., & Duarte, G. B. (2010). Impacto do Programa Bolsa Família sobre a frequência escolar: o caso da agricultura familiar no Nordeste do Brasil. Revista de Economia e Sociologia Rural, 48(3), 635–656.

Moens, M., et al. (2025). Artificial intelligence as team member versus manual screening to conduct systematic reviews. Journal of Clinical Epidemiology, 179, 111646.

Moreira, C. C. S. (2017). Entre a conservação ambiental e a transferência de renda: o Programa Bolsa Verde em uma reserva extrativista marinha [Dissertação de mestrado, Universidade Federal do Pará].

Nascimento, R. M. do. (2021). Efeitos do Programa Bolsa Verde sobre o desmatamento: uma análise sobre os assentamentos rurais [Tese de doutorado]. Universidade Federal de Viçosa.

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.17986

Novak, M. A. L. (2018). Pagamento por serviços ambientais: a política pública do Programa Bolsa Verde [Dissertação de mestrado, Universidade Estadual do Centro-Oeste].

Novak, M. A. L., & Gonzaga, C. A. M. (2024). Consequências socioambientais após descontinuação do Programa Bolsa Verde em assentamentos rurais. Revista NERA.

Nunes, W. (2020). Políticas públicas e construção de capacidades estatais: comparando políticas industriais e sociais no Brasil. Dados, 63(4), e20190133.

Nutley, S. M., Powell, A. E., & Davies, H. T. O. (2013). What counts as good evidence? Alliance for Useful Evidence.

Pawson, R. (2002). Evidence-based policy: The promise of realist synthesis. Evaluation, 8(3), 340–358. https://doi.org/10.1177/135638902401462448

Pawson, R., Greenhalgh, T., Harvey, G., & Walshe, K. (2005). Realist review — a new method of systematic review designed for complex policy interventions. Journal of Health Services Research & Policy, 10(Suppl. 1), 21–34.

Petticrew, M., & Roberts, H. (2003). Evidence, hierarchies, and typologies: Horses for courses. Journal of Epidemiology & Community Health, 57(7), 527–529.

Petticrew, M., & Roberts, H. (2006). Systematic reviews in the social sciences: A practical guide. Blackwell.

Porta, M. (Ed.). (2016). A dictionary of epidemiology (6th ed.). Oxford University Press.

Priem, J., Piwowar, H., & Orr, R. (2022). OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts [Pré-print]. arXiv.

Queiroz e Silva, S. (2016). Políticas socioambientais na Amazônia: o Programa Bolsa Verde e a inclusão produtiva das famílias [Tese de doutorado, Universidade Federal do Amazonas].

Ramos, M. P., & Schabbach, L. M. (2012). O estado da arte da avaliação de políticas públicas: conceituação e exemplos de avaliação no Brasil. Revista de Administração Pública, 46(5), 1271–1294.

Sartori, G. (1970). Concept misformation in comparative politics. The American Political Science Review, 64(4), 1033–1053. https://doi.org/10.2307/1958356

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.17986

Silva, G. P. da. (2015). Desafios ontológicos e epistemológicos para os métodos mistos na ciência política. Revista Política Hoje, 24(1), 11–30.

Slough, T., & Tyson, S. A. (2023). External validity and meta-analysis. American Journal of Political Science, 67(2), 440–455. https://doi.org/10.1111/ajps.12664

Small, M. L. (2011). How to conduct a mixed methods study: Recent trends in a rapidly growing literature. Annual Review of Sociology, 37, 57–86. https://doi.org/10.1146/annurev.soc.012809.102657

Snilstveit, B., Oliver, S., & Vojtkova, M. (2012). Narrative approaches to systematic review and synthesis of evidence for international development policy and practice. Journal of Development Effectiveness, 4(3), 409–429.

Souza, M. T. de, Silva, M. D. da, & Carvalho, R. de. (2010). Revisão integrativa: o que é e como fazer. Einstein, 8(1), 102–106.

Spohn, M., Lasky-Fink, J. & Linos, Elizabeth. (2026). What policymakers want: Documenting a mismatch in policymaker demand and research supply. Policy Brief. The People Lab.

Stern, C., et al. (2020). Methodological guidance for the conduct of mixed methods systematic reviews. JBI Evidence Synthesis, 18(10), 2108–2118.

Sutton, A., Clowes, M., Preston, L., & Booth, A. (2019). Meeting the review family: Exploring review types and associated information retrieval requirements. Health Information & Libraries Journal, 36(3), 202–222.

Vetter, T. R. (2019). Systematic review and meta-analysis: Sometimes bigger is indeed better. Anesthesia & Analgesia, 128(3), 575–583.

Viana, J. P. (2013). Avaliação de duas ações governamentais recentes em apoio a extrativistas: garantia de preços mínimos. Ipea.

Viana, J. P. (2014). Dois anos de Bolsa Verde: seria a meta alcançável? (Texto para Discussão). Ipea.

Wagner, G., Lukyanenko, R., & Paré, G. (2026). Generative artificial intelligence for literature reviews. Journal of Information Technology.

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.17986

Wong, P. Y., et al. (2026). Individual pay for collective performance: Evidence from deforestation in Brazil. Journal of Environmental Economics and Management, 133, 103308.

Xiao, Y., & Watson, M. (2019). Guidance on conducting a systematic literature review. Journal of Planning Education and Research, 39(1), 93–112. https://doi.org/10.1177/0739456X17723971

DECLARAÇÕES

CONTRIBUIÇÃO DE AUTORIA (CREDIT).

Bruno Marques Schaefer: Conceituação (Liderança); Curadoria de Dados (Igual); Metodologia (Liderança); Administração de projeto (Liderança); Supervisão (Liderança); Validação (Liderança); Escrita - rascunho original (Igual); Escrita - revisão e edição (Igual).

Tomás Paixão Borges: Conceituação (Igual); Curadoria de Dados (Igual); Análise formal (Igual); Metodologia (Igual); Validação (Liderança); Escrita - rascunho original (Igual); Escrita - revisão e edição (Igual).

Carlos Freitas: Conceituação (Igual); Curadoria de Dados (Igual); Análise formal (Igual); Metodologia (Igual); Validação (Suporte); Escrita - rascunho original (Igual); Escrita - revisão e edição (Igual).

Frederico Augusto Auad de Gomes Filho: Conceituação (Igual); Curadoria de Dados (Igual); Análise formal (Igual); Metodologia (Igual); Validação (Suporte); Escrita - rascunho original (Igual); Escrita - revisão e edição (Igual).

Ana Paula Lima dos Santos: Conceituação (Igual); Curadoria de Dados (Igual); Análise formal (Igual); Metodologia (Igual); Validação (Suporte); Escrita - rascunho original (Igual); Escrita - revisão e edição (Igual).

CONFLITO DE INTERESSE.

Os autores declaram não haver conflito de interesse.

FINANCIAMENTO.

Bruno Marques Schaefer declara financiamento do PROCIÊNCIA UERJ (2023–2026). O trabalho também recebeu recursos por meio de três emendas parlamentares da Assembleia Legislativa do Estado do Rio de Janeiro (ALERJ), destinadas pelas deputadas

SciELO Preprints - Este documento é um preprint e sua situação atual está disponível em: https://doi.org/10.1590/SciELOPreprints.17986

Dani Balbi (emendas nº 2711 e 2714) e Dani Monteiro (emenda nº 976). As emendas não se vincularam ao conteúdo específico deste artigo, mas fomentaram o trabalho mais amplo de políticas informadas por evidências e a estruturação do projeto O que Funciona (OQF), iniciativa do MAPE (Laboratório de Monitoramento e Avaliação de Políticas e Eleições) - https://mape.org.br/oqf/.

DECLARAÇÃO DE DISPONIBILIDADE DE DADOS DE PESQUISA.

Os scripts em R, as bases intermediárias e o material suplementar estão depositados no Open Science Framework e podem ser acessados publicamente (https://osf.io/pb9xz/overview). Além disso, o protocolo está disponível em: https://osf.io/b792u/overview

USO DE INTELIGÊNCIA ARTIFICIAL.

Modelos de linguagem foram utilizados como avaliadores adicionais na etapa de triagem substantiva, conforme descrito na seção 4.3, com registro do prompt, das decisões e das medidas de concordância. A redação do manuscrito é de responsabilidade dos autores.

Este preprint foi submetido sob as seguintes condições:

- Os autores declaram que os necessários Termos de Consentimento Livre e Esclarecido de participantes ou pacientes na pesquisa foram obtidos e estão descritos no manuscrito, quando aplicável.
- Os autores declaram que a elaboração do manuscrito seguiu as normas éticas de comunicação científica.
- Os autores declaram que estão cientes que são os únicos responsáveis pelo conteúdo do preprint e que o depósito no SciELO Preprints não significa nenhum compromisso de parte do SciELO, exceto sua preservação e disseminação.
- Os autores declaram que os dados, aplicativos e outros conteúdos subjacentes ao manuscrito estão referenciados.
- O manuscrito depositado está no formato PDF.
- Os autores declaram que a pesquisa que deu origem ao manuscrito seguiu as boas práticas éticas e que as necessárias aprovações de comitês de ética de pesquisa, quando aplicável, estão descritas no manuscrito.
- Os autores declaram que uma vez que um manuscrito é postado no servidor SciELO Preprints, o mesmo só poderá ser retirado mediante pedido à Secretaria Editorial do SciELO Preprints, que afixará um aviso de retratação no seu lugar.
- Os autores concordam que o manuscrito aprovado será disponibilizado sob licença Creative Commons CC-BY.
- O autor submissor declara que as contribuições de todos os autores e declaração de conflito de interesses estão incluídas de maneira explícita e em seções específicas do manuscrito.
- Os autores declaram que o manuscrito não foi depositado e/ou disponibilizado previamente em outro servidor de preprints ou publicado em um periódico.
- Caso o manuscrito esteja em processo de avaliação ou sendo preparado para publicação mas ainda não publicado por um periódico, os autores declaram que receberam autorização do periódico para realizar este depósito.
- O autor submissor declara que todos os autores do manuscrito concordam com a submissão ao SciELO Preprints.