Topic	Guiding protocol	Application
Administrative information		
Title	1,2,4	An evidence synthesis of the academic regulatory failure literature: Protocol
Date		7 August 2020
Key words		Regulatory failure, public regulation, regulatory governance, regulatory delivery
Update/version	1, 4	This is the first version (v1) of the protocol that was approved by JH and NV on 10 August 2020.
Registration	1, 4	The protocol will be submitted to the Open Science Foundation (OSF) for preregistration. The original version of the protocol (v1) was published on 13 August 2020, prior to undertaking the evidence synthesis, on https://jeroenvanderheijden.net/?p=410
Corresponding author(s)	1, 4	Corresponding author: Jeroen van der Heijden, Victoria University of Wellington Mobile +64 (0)22 563 5082 Room 821a, Rutherford House, Pipitea Campus PO Box 600, Wellington 6140, New Zealand jeroenvanderheijden@vuw.ac.nz
Contributors, contributions	1, 4	JH is author of this protocol. The protocol takes inspiration from a series of handbooks on meta-analysis, systematic reviews, and evidence synthesis (Card, 2015; Cooper, 2017; Cooper, Hedges, & Valentine, 2019; Eklund Karlson & Takahashi, 2017; Gough, Oliver, & Thomas, 2012; Heyvaert, Hannes, & Onghena, 2017; Howell Major & Savin-Baden, 2010; Ringquist, 2013; Saini & Shlonsky, 2012). The second reviewer (NV) read, provided feedback and approved this protocol.
Amendments	1, 4	If this protocol needs amendments, these will be explained in an updated version of the protocol. Each amendment, description of changes made, and the rational for making changes will be described.
Support	1, 4	This evidence synthesis does not receive targeted funding or financial support. JH is the Chair in Regulatory Practice at VUW. This role is partially funded by the New Zealand Government Regulatory Practice (G-REG) initiative. The evidence review will be used for educational purposes, including the G-REG ongoing professional education program. NV was contracted by the Chair for support with the review.
Sponsor	1, 4	None.
Role of sponsor	1, 4	N/A.
Introduction		
Rationale	1, 2, 3, 4	Regulatory failure is much talked about but little understood. Discussions about regulatory failure are often about different understandings of what can be expected of regulatory governance and public regulation. The rhetoric of regulatory failure (typically a blame game) easily (and often) overshadows the analytical explanation of it (Baldwin, Cave, & Lodge, 2012, Chapter 5; Breyer, 1979; Wilson, 1984). As a result, it remains unknown: - What types of failure can be distinguished from the evidence base of the academic literature on regulatory failure; - Under what circumstances regulation is likely to fail and how it fails; - What strategies are available and have been proven successful in the prevention of regulatory failure.
Objectives	1, 2, 3, 4	The aim of this evidence synthesis is to synthesize the evidence base of the academic regulatory failure literature and evaluate the effectiveness (and lack thereof) of strategies to prevent regulatory failure in real-world situations. To this end, this evidence synthesis will answer the following questions: 1. Under what circumstance, when, and how is regulation likely to fail?

		2. What evidence-based typology can be distilled from the regulatory failure literature? 3. What strategies are applied by governments and other regulators to prevent regulatory failure and with what level of success? 4. For questions 1 and 3, if heterogeneity is found in studies on regulatory failure: What is the role of context on the occurrence and prevention of regulatory failure?
Methods		
Eligibility criteria	1, 2, 3, 4	Studies will be selected according to the PICO criteria (participants, interventions, comparators, outcomes) outlined below: *Study designs* We will include any type of empirical study that assesses the causes of regulatory failure, the performance of regulatory failure prevention strategies, or both, as well as single n, medium n, and large n studies of observed situations of regulatory failure. We will only assess unique empirical studies once. If an individual study is reported in multiple publications, we will collate reported research and performance information for that study. *Participants* There will be no restrictions by background, type or number of participants (including, but not limited to, people, firms, regulators, and jurisdictions) of studies. *Interventions* Of interest are studies that empirically observe failure of a regulatory intervention, or the empirically study the performance of an intervention that seeks to prevent regulatory failure. *Comparators* Given the broad perspective of regulatory failure (failure to maintain regulatory independence, over-regulation, under-regulation, failing enforcement, and so on) and the unlikeliness of (a significant number of) studies that compare a pre-failure, post-failure, or non-failure situation with an observed instance of regulatory failure, several types of studies will be relevant to include to allow for cross-study comparisons: - Direct non-comparative observations: studies that provide deep insight in an observed instance of regulatory failure, following, for example, process tracing methodology, but that do not compare the failure situation with pre-failure, post-failure, or non-failure situation - Direct comparative observations: as per above, but a direct comparison is made between the failed situation and a pre-failure, post-failure, or non-failure situation - Counterfactual observations: studies that compare an observed instance of regulatory failure with a counterfactual non-regulatory failure situation *Outcomes* Endpoints important for typology building are of primary interest. Endpoints important for decision making are of secondary interest (needless to say, there is overlap between the two). - Endpoints important for typology building: o The stage of the regulatory process in which regulatory failure is observed o The causal process of regulatory failure observed

		○ The impact of regulatory failure observed on the regulator (monetary costs, extra time required to achieve compliance, etc.) ○ The impact of regulatory failure observed on the targets of regulation (monetary costs, property damaged, lives lost, extra time required to achieve compliance, etc.) ○ The political implications of regulatory failure observed (loss of confidence in policymakers, shift in policy rhetoric, etc.) ○ The impact of regulatory failure on society at large (monetary costs, property damaged, lives lost, etc.) ○ As per the above for regulatory failure prevention strategies observed - Additional endpoints important for decision making: ○ Changes in relationships with regulatees (improved, deteriorated, etc.) ○ Spill-over effects (reduced levels of compliance because regulatees become aware of the regulatory failure) ○ Changed working conditions for staff (improved, deteriorated, etc.) ○ Changes in the relationship between regulators and policymakers/politicians (directly responsible ministers, members of parliament, etc.) *Timing* There will be no restrictions by time, length, or repetitions (including no repetitions) of studies. *Setting* There will be no restrictions by the setting(s) of studies. *Language* We will include articles reported in English. *Academic literature* We will only include published peer-reviewed articles, including 'online first' and 'early access' publications. We will acknowledge the limitations of excluding non-published academic work and academic publications other than peer-reviewed articles when reporting findings from the evidence synthesis (Vevea, Coburn, & Sutton, 2019). *Non-academic literature* There will be no selection of non-academic literature. Our interest is primarily in empirical findings on regulatory failure reported by the academic community. We will acknowledge the limitations of excluding non-academic literature when reporting findings from the evidence synthesis (Mahood, Van Eerd, & Irvin, 2014). This is, to the best of our knowledge, the first systematic evaluation of the regulatory failure literature following formal meta-research methodology. As such, the synthesis may become a starting point for future studies to synthesise other areas of the knowledge on regulatory failure—including, but not limited to, grey literature and academic literature beyond peer-reviewed journal articles.
Information sources	1, 4	Both qualitative and quantitative studies will be sought. Documents will be sourced from the following databases: WorldCat, Scopus and Web of Science.
Search strategy	1, 2, 3, 4	Search strategies for each database:

		- WorldCat search:
		  o all articles with the words "regulatory failure" in any searchable field, published in English, since 1900, in the following subject areas: business and economics, law, sociology, political science.
		- Scopus search:
		  o all articles with the words "regulatory failure" in their titles, abstracts, or keywords, published in English, since 1900 in the following subject areas: social sciences; business, management and accounting; economic, econometrics and finance.
		- Web of Science search:
		  o all articles with the words "regulatory failure" in any searchable field, published in English, since 1900, in the following disciplines: business economics, government law, public administration, social issues, social sciences other topics.

Study records and data management	1, 4	We will use MS Excel to for data management (document selection and document coding) and a shared Dropbox folder to share files.

Selection process	1, 2, 4	Two reviewers (JH and NV) will independently screen articles against the following inclusion criteria. This will be done in three rounds:
		- Round 1
		  o Article titles, abstract and keywords will be screened to exclude articles that are unlikely to deal with regulatory failure or its prevention, that are explicitly not empirical, or both. The reviewers will use the following scores: yes (include), no (exclude), unsure (include). Intercoder reliability scores will be reported (agreement percentage and Cohen's kappa). In this round, we take a liberal approach to inclusion and will only exclude the combinations of 'no' and 'no' (i.e., if at least one of the coders uses the score 'yes' or 'unsure', the article is included for screening in the next step).
		- Round 2
		  o Article research design sections (or similar) will be screened to exclude articles that are not dealing with an observed instance (or instances) of regulatory failure or a regulatory failure prevention strategy, and article method sections (or similar) will be screened to exclude articles that are explicitly not empirical. The reviewers will use the following scores: yes (include), no (exclude), unsure (include). Intercoder reliability scores will be reported (agreement percentage and Cohen's kappa). In this round, we resolve all intercoder disagreements for the combinations 'yes' and 'no' (i.e., all articles coded with the combination 'yes' and 'yes', 'yes' and 'unsure', and 'unsure' and 'unsure' are included for screening in the next step; all other combinations are resolved).
		- Round 3
		  o Articles are screened in full to exclude articles that are not dealing with an observed instance (or instances) of regulatory failure or a regulatory failure prevention strategy, that are explicitly not empirical, or both. The reviewers will use the following scores: yes (include), no (exclude), unsure (include). Intercoder reliability scores will

		be reported (agreement percentage and Cohen's kappa). In this round, coders resolve all conflicts. - Round 4: o Articles are screened in full to cluster articles that report on the same study (this to prevent 'double counting' of individual studies). To ensure consistency across the reviewers, we will conduct calibration exercises. We will resolve disagreements by discussion.
Data items (variables and outcomes)		
Data abstraction	1,2, 3, 4	Data will be abstracted from the articles following the PICO criteria discussed above. Data abstracted will be recorded in text (to be coded later). A standardized Excel form will be used to ensure that both reviewers (JH and NV) abstract similar data from the articles. To ensure consistency across the reviewers, we will conduct calibration exercises. We will resolve disagreements by discussion.
Primary variables	1, 2, 3, 4	JH and NV will abstract: - Geographical location lead author - Geographical location(s) of regulatory failure (or prevention strategy) studied - Calendar year(s) regulatory failure (or prevention strategy) occurred - Calendar year(s) regulatory failure (or prevention strategy) was studied - Policy area(s) regulatory failure (or prevention strategy) studied - Type of observation (direct, counterfactual, indirect – see above) - Regulatory stage at which failure is observed (or prevention strategy was introduced)
Secondary variables	1, 2, 3	JH will abstract: - Causal narrative of regulatory failure (or prevention strategy) - Other relevant observations about the 'input', 'throughput' or 'output' of the instance of regulatory failure (or prevention strategy) studied
Primary outcomes	1, 2, 3, 4	JH will abstract (if applicable/reported): - Impact on regulator (as per above) - Impact on target of regulation (as per above) - Political impact of regulatory failure (or prevention strategy) - Societal impact of regulatory failure (or prevention strategy)
Secondary outcomes	1, 2, 3	JH to abstract.: - Decreased/improved relationships with regulatees - Spill-over effects - Decreased/improved working conditions for staff - Decreased/improved relationship between regulator and policymaker(s)/politician(s)
Note		The exact set of data items to be abstracted will be affected by the selection process. Engagement with the literature in this process will give a better understanding of the variables and outcomes that can be abstracted from the selected articles.
Risk of bias individual studies		
Risk assessment	1, 4	We expect to predominantly find single-n in-depth, or small-n comparative qualitative studies. This limits the extent to which we can use accepted protocols for assessing potential bias of these studies. To the extent possible, the reviewers (JH and NV) will assess selection bias, performance bias, detection bias, attrition bias and reporting bias (Mayo-Wilson & Grant, 2019).

		This will be done at the time of round 4 of the selection process (discussed above). The reviewers will use the following scores: low risk of bias, unclear risk of bias, high risk of bias.
Data synthesis		
Quantitative synthesis, methods	1, 2, 4	We do not expect to find a large enough number of studies that report an effect-size to allow for a quantitative synthesis (e.g., fixed effects or random effects models).
Additional quantitative analyses	1, 2, 4	If possible, we will provide descriptive statistics when presenting the findings from the evidence synthesis – while keeping in mind the risks of such 'vote counting' (Borenstein, Hedges, Higgins, & Rothstein, 2009).
Qualitative synthesis, methods	3	If the abstracted data allow, we will follow a realist synthesis approach, "a theory-driven synthesis aimed at unpacking mechanisms of how complex programs or interventions work (or why they fail) in particular contexts and settings" (Heyvaert et al., 2017, 237).
Additional qualitative analyses	3	If the abstracted data do not allow for a realist synthesis approach, we will fall back on a more traditional (but systematic) synthesis approach and present findings in a transparent, consistent and comprehensive manner. We will aggregate and integrate data where possible and aim to generate a new inductive understanding of regulatory failure (i.e., an interpretive approach) (Howell Major & Savin-Baden, 2010; Saini & Shlonsky, 2012).
Meta-bias(es)		
Assessment	1, 3, 4	The main publication bias that our evidence synthesis is subject to is the sole focus on articles in peer-reviewed academic journals. There is a risk that we will find a relatively high number of positive experiences with regulatory failure (and prevention strategies) that may not be representative of all the experiences with regulatory failure (and prevention strategies) (i.e., a selective reporting bias). We will be explicit about this possible bias when presenting the findings of the evidence synthesis. Unfortunately, we will not be able to run tests for sample biases (Hardwicke et al., 2020) as we expect to predominantly find single-n in-depth, or small-n comparative qualitative studies.
Confidence or quality assessment		
Method	1, 2, 3, 4	We expect to predominantly find single-n in-depth, or small-n comparative qualitative studies. This limits the extent to which we can use accepted protocols for assessing the quality of these studies. To the extent possible, the reviewers (JH and NV) will assess the quality dimension of the studies using the CASP Checklist for qualitative research (CASP, 2018). This will be done at the time of round 4 of the selection process (discussed above). The coders will use the following scores: risk of low quality, unclear risk of low quality, little risk of low quality. An average risk estimation will be calculated from the coders' scores. Please note, this quality dimension assessment is not meant to judge the quality of the individual studies, but to assess how much weight we can reasonably assign to findings presented in the evidence synthesis (Heyvaert et al., 2017).

Guides used: (1) AMSTAR 2 = MeaSurement Tool to Assess Systematic Reviews version 2 (Shea et al., 2017); (2) MARS = American Psychological Association (APA) Meta-Analysis Reporting Standards (essentially the MARS protocol is modified from, Cooper, 2017); (3) MMRS = Mixes Methods Research Synthesis protocol (Heyvaert et al., 2017); (4) PRISMA-P = Preferred Reporting Items for Systematic Meta-Analyses (Shamseer et al., 2015).

References

Baldwin, R., Cave, M., & Lodge, M. (2012). *Understanding Regulation: Theory, Strategy and Practice - second edition*. Oxford: Oxford University Press.

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2009). *Introduction to meta-analysis*. Chichester: Wiley.

Breyer, S. (1979). Analyzing Regulatory Failure: Mismatches, Less Restrictive Alternatives, and Reform. *Harvard Law Review, 92*(3), 547-609.

Card, N. (2015). *Applied Meta-Analysis for Social Science Research*. New York: The Guilford Press.

Cooper, H. (2017). *Research synthesis and meta-analysis: A step-by-step approach* (5th ed.). Los Angeles: SAGE.

Cooper, H., Hedges, L., & Valentine, J. (Eds.). (2019). *The handbook of research synthesis and meta-analysis* (3rd ed.). New York: Russel Sage Foundatation.

Eklund Karlson, L., & Takahashi, R. (2017). *A Resource for Developing an Evidence Synthesis Report for Policy-Making* Copenhagen: WHO Regional Office for Europe.

Gough, D., Oliver, S., & Thomas, J. (2012). *An introduction to systematic reviews*. London: Sage.

Hardwicke, T. E., Serghiou, S., Janiaud, P., Danchev, V., Crüwell, S., Goodman, S., & Ioannidis, J. (2020). Calibrating the Scientific Ecosystem Through Meta-Research. *Annual Review of Statistics and Its Application, 7*(1), 1-27 (early view).

Heyvaert, M., Hannes, K., & Onghena, P. (2017). *Using mixed methods research synthesis for literature reviews*. Los Angeles: SAGE.

Howell Major, C., & Savin-Baden, M. (2010). *An Introduction to Qualitative Research Synthesis : Managing the Information Explosion in Social Science Research*. Abingdon: Routledge.

Mahood, Q., Van Eerd, D., & Irvin, E. (2014). Searching for grey literature for systematic reviews: challenges and benefits. *Research Synthesis Methods, 5*(3), 221-234.

Mayo-Wilson, E., & Grant, S. (2019). Transparent reporting: Registrations, protocols, and final reports. In H. Cooper, L. Hedges, & J. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (3rd ed., pp. 471-488). New York: Russel Sage Foundation.

Ringquist, E. (2013). *Meta-Analysis for Public Management and Policy*. San Francisco: John Wiley.

Saini, M., & Shlonsky, A. (2012). *Systematic Synthesis of Qualitative Research*. Oxford: Oxford University Press.

Shamseer, L., Moher, D., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., . . . Stewart, L. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. *BMJ, 350*(g7647), 1-25.

Shea, B., Reeves, B., Wells, G., Thuku, M., Hamel, C., Moran, J., . . . Henry, D. (2017). AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ, 358*(j4008), 1-.

Vevea, J., Coburn, K., & Sutton, A. (2019). Publication bias. In H. Cooper, L. Hedges, & J. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (3rd ed., pp. 383-429). New York: Russel Sage Foundation.

Wilson, G. (1984). Social Regulation and Explanations of Regulatory Failure. *Policy Studies, 32*(2), 203-225.