

Naar een regelgevingscyclus?

Evaluatie in de Europese Unie

E. Mastenbroek, A.C.M. Meuwese & S. van Voorst

1. Inleiding¹

Evaluaties vervullen in theorie een belangrijke rol in processen van regelgeving. Door kritisch te bezien wat de implementatie, naleving en effectiviteit van (voorgenomen) regelgeving zijn, kunnen fouten in de onderliggende beleidstheorie en in het proces van implementatie worden opgespoord. Dit kan de basis bieden voor aanpassingen van regelgeving, zodat deze daadwerkelijk leidt tot het bereiken van de doeleinden, waarvoor deze ooit in het leven is geroepen.

In de Europese Unie (EU) is deze idee van *evidence-based* regelgeving relatief sterk ontwikkeld. De EU is internationaal gezien een van de voorlopers op het vlak van *Better Regulation*² – een beleidsfilosofie die vanaf de eeuwwisseling haar intrede heeft gedaan in OESO-verband, met als belangrijkste doel het komen tot betere regelgeving. Deze voorloperrol mag geen verbazing wekken. Gezien het democratisch tekort van de EU staat de Europese Commissie onder voortdurende druk om effectief gemeenschappelijke problemen op te lossen, om zo de *output*-legitimiteit van de EU te vergroten.³

Evaluaties vervullen een belangrijke functie in dit kader. De centrale idee is dat het regelgevingproces een cyclus is, waarbij zogenoemde *impact assessments* (IAs) de basis vormen voor latere ex-postwetsevaluatie.⁴ Het sluitstuk van deze gedachtegang wordt gevormd door het recente 'REFIT'-programma.⁵ Centraal hierin staat de gedachte dat regelgeving kritisch moet worden bezien op haar werking en effectiviteit – onvoordelige evaluaties zouden tot herziening of zelfs intrekking van regelgeving moeten kunnen leiden.

In de uitwerking van de *Better Regulation*-filosofie heeft lange tijd de nadruk gelegen op de rol van *impact assessments*. De Europese Commissie heeft een systeem

- 1 Onderhavig artikel vormt het eerste Nederlandstalige verslag van een onderzoeksproject over ex-postevaluatie in de EU, dat de drie auteurs gezamenlijk uitvoeren. Op termijn zullen nadere verklarende analyses van de initiatie en kwaliteit van ex-postevaluaties volgen.
- 2 Strikt genomen is dit gedachtegoed in 2010 van naam veranderd naar *Smart Regulation*. Echter, wij kiezen er bewust voor om in dit artikel te spreken van *Better Regulation*, aangezien dat de oorspronkelijke beleidsfilosofie is.
- 3 G. Majone, *Regulating Europe*, Londen: Routledge 1996; F.W. Scharpf, *Governing in Europe: Effective and Democratic?*, Oxford: Oxford University Press 1999.
- 4 G. Luchetta, 'Impact Assessment and the Policy Cycle in the EU', *The European Journal of Risk Regulation* 2012, p. 561-575.
- 5 Europese Commissie, 'Regulatory Fitness and Performance (REFIT): Results and Next Steps', COM(2013)685 final.

van *impact assessments* opgetuigd dat 'bijna revolutionair'⁶ is genoemd, aangezien *alle* voorstellen tot regelgeving worden beoordeeld op hun verwachte sociale, economische en milieugerelateerde gevolgen.⁷ Hiernaast heeft de Commissie uitgebreide aandacht besteed aan de ex-postevaluatie van programma's door de lidstaten.⁸ In dit aspect van *evidence-based regelgeving*, ofwel regelgeving die gebaseerd is op feiten, lijkt de Europese Commissie relatief weinig te hebben geïnvesteerd.⁹ Hiernaast zijn kritische geluiden te horen geweest over de kwaliteit van zowel *impact assessments*¹⁰ als ex-postevaluaties.¹¹

Het grote belang van evaluaties, in combinatie met empirische vraagtekens over de daadwerkelijke implementatie van het *Better Regulation*-discours, leidt tot een behoefte aan meta-evaluatie: in welke mate krijgt de gedachte van de regelgevingcyclus feitelijk vorm in de EU? In onderhavig stuk beantwoorden wij deze brede vraag alvast door te analyseren hoe het gesteld is met twee noodzakelijke voorwaarden voor een succesvolle regelgevingcyclus: de dekkingsgraad en kwaliteit van deze twee typen evaluaties. Zo hopen wij een bijdrage te bieden aan de wetenschappelijke literatuur over deze thematiek. Dit is met name belangrijk aangezien ex-postwetsevaluaties binnen de EU tot op heden vrijwel geen aandacht hebben gekregen, met uitzondering van enkele exploratieve studies.¹²

De opzet van het artikel is als volgt. De volgende paragraaf geeft kort de beleidsmatige context weer door de ontwikkeling en inhoud van het *Better Regulation*-programma van de Europese Commissie te schetsen. Paragraaf 3 geeft een overzicht van de stand van zaken rond *impact assessments* op basis van secundaire literatuuranalyse. Paragraaf 4 brengt de dekkingsgraad en kwaliteit van ex-post-

- 6 L. Allio, 'Better Regulation and Impact Assessment in the European Commission', in: C. Kirkpatrick & D. Parker (red.), *Regulatory Impact Assessment: Towards Better Regulation*, Cheltenham, UK: Edward Elgar 2007, p. 72-105.
- 7 C. Cecot, R.W. Hahn, A. Renda & L. Schrefler, 'An Evaluation of the Quality of Impact Assessment in the European Union with Lessons for the US and the EU', *Regulation and Governance* 2008, p. 405-424; C.M. Radaelli & A.C.M. Meuwese, 'Hard Questions, Hard Solutions: Proceduralisation through Impact Assessment in the EU', *West European Politics* 2010, p. 136-153; A. Renda, *Impact Assessment in the EU: The State of the Art and the Art of the State*, Brussel: Centre for European Policy Studies 2006.
- 8 S.E. Batterbury, 'Evaluation Concepts and Methods: Principles and Purposes of European Union Cohesion Policy Evaluation', *Regional Studies* 2006, p. 179-188; J. Bachtler & C. Wren, 'The Evaluation of EU Cohesion Policy: Research Questions and Policy Challenges', *Special Issue of Regional Studies* 2006, p. 143-153.
- 9 Europese Rekenkamer, *Effectbeoordelingen in de EU-instellingen: ondersteunen zij de besluitvorming?* (Speciaal verslag nr. 3/2010), Luxemburg: Bureau voor publicaties van de Europese Unie 2010.
- 10 Renda 2006; Cecot e.a. 2008.
- 11 E. Stern, 'Evaluation Policy in the European Union and Its Institutions', in: W.M.K. Trochim, M.M. Mark & L.J. Cooksy (red.), *Evaluation Policy and Evaluation Practice: New Directions for Evaluation*, San Francisco, CA: Jossey-Bass 2009, p. 67-85.
- 12 F.L. Leeuw, J. Toulemonde & A. Brouwers, 'Evaluation Activities in Europe: A Quick Scan of the Market in 1998', *Evaluation* 1999, p. 487-496; H. Summa & J. Toulemonde, 'Evaluation in the European Union: Addressing Complexity and Ambiguity', in: J. Furubo, R.C. Rist & R. Sandahl (red.), *International Atlas of Evaluation*, New Brunswick: Transaction 2002, p. 407-424; T. Fitzpatrick, 'Evaluating Legislation: An Alternative Approach for Evaluating EU Internal Market and Services Law', *Evaluation* 2012, p. 477-499; E.M. Poptcheva, *Library Briefing. Policy and Legislative Evaluation in the EU*, Brussel: European Parliament 2013.

wetsevaluaties in kaart met behulp van twee nieuwe, unieke datasets. Hierbij wordt ook ingegaan op de belangrijkste kenmerken van de twee vormen van evaluaties in de EU, om zo te kunnen bezien in hoeverre zij rijp zijn voor een vruchtbare integratie. Paragraaf 5 neemt de bevindingen samen en komt tot een conclusie.

2. Context: *Better Regulation*

Het *Better Regulation*-programma van de Europese Commissie, tegenwoordig vaak *Smart Regulation* genoemd, heeft zijn wortels in het Mandelkern-rapport uit 2001, dat namens de lidstaten opriep tot een stevigere feitenbasis onder Europese besluitvorming. Het witboek over *governance* ontwikkelde deze gedachte verder en verbond kernbeginselen van *good governance*, waaronder effectiviteit, verantwoording en openheid, met de noodzaak de kwaliteit van Europese regelgeving te verhogen.¹³ De resulterende *Better Regulation*-filosofie kent verschillende bouwstenen, zoals eisen aan consultatie, programma's voor versimpeling van regelgeving en *impact assessments*.¹⁴ Het laatste aspect heeft veruit de meeste aandacht gekregen van zowel beleidsmakers als wetenschappers.

De Europese Commissie heeft veel aandacht besteed aan het operationaliseren van de belofte van *Better Regulation*. Hierin is vanaf 2002 wel een duidelijke keuze te zien: de Europese Commissie zette in op systematische *impact assessments* als het middel voor het verbeteren van de effectiviteit van regelgeving.¹⁵ Deze keuze werd bestendigd in het Interinstitutioneel Akkoord van 2003, dat de centrale rol van *impact assessments* in het nieuwe programma herhaalde.¹⁶ De nadruk kwam al met al te liggen op ex-antebeoordeling van voorstellen tot regelgeving, in plaats van op het onderzoeken van de daadwerkelijke effectiviteit hiervan.

In 2007 leek het tij te keren: de Commissie herhaalde het belang van ex-post-evaluaties, met het oog op het verbeteren van de effectiviteit en het vergroten van vertrouwen, legitimiteit en verantwoording.¹⁷ Ze beloofde een actieplan om het belang van ex-postevaluaties te bevorderen.¹⁸ Deze ambitie zou meermaals worden herhaald. Commissievoorzitter Barroso benadrukte in zijn politieke richtlijnen voor 2010-2014¹⁹ dat de Commissie de grote investeringen in ex-ante-evaluatie zou evenaren op het vlak van ex-postevaluatie. In haar communicatie over 'Smart Regulation' presenteerde de Europese Commissie²⁰ ex-postevaluatie

13 Europese Commissie, *Europese governance – een witboek*, COM(2001)428, PbEG 2001, C 287/1.

14 C.M. Radaelli & A.C.M. Meuwese, 'Better Regulation in Europe: Between Public Management and Regulatory Reform', *Public Administration* 2009, p. 639-654.

15 Europese Commissie, 'European Governance: Better Lawmaking', COM(2002)275 final.

16 Interinstitutioneel Akkoord over 'beter wetgeven', PbEG 2003, C 321/1.

17 Europese Commissie, 'Communication from the Commission from Ms Grybauskaitė in Agreement with the President. Responding to Strategic Needs: Reinforcing the Use of Evaluation', SEC(2007)213.

18 Fitzpatrick 2012.

19 J.M. Barroso, 'Political Guidelines for the Next Commission', 2009, vinddatum 6 december 2011 via <http://ec.europa.eu/commission_2010-2014/president/pdf/press_20090903_en.pdf>.

20 Europese Commissie, 'Smart Regulation in the European Union', COM(2010)543.

als een centraal instrument voor het regelgevingproces. Vorig jaar presenteerde de Commissie het regelgevingproces als een cyclus: van ontstaan, via ontwerp en implementatie, terug naar de tekentafel – op basis van een herzieningsprocedure geworteld in evaluaties.²¹ Hierbij is een rol weggelegd voor zowel evaluaties van individuele regelgeving als evaluaties van hele beleidsterreinen, zogenaamde *fitness checks*.²²

Met het oog op deze herziening is de Commissie momenteel bezig haar evaluatiebeleid te herzien. Onze meta-evaluatie komt dus op een tijdig moment. De tijd is rijp om te bezien in hoeverre een aantal noodzakelijke voorwaarden voor een goed functionerende regelgevingscyclus reeds aanwezig is. Hierbij gaat het allereerst om de dekkinggraad van *impact assessments* en ex-postwetsevaluaties; het is belangrijk dat er systematisch wordt geëvalueerd, om zo de schijn van subjectiviteit weg te nemen.²³ Daarnaast is de kwaliteit van de evaluaties van belang. Evaluaties kunnen pas bijdragen aan verbetering van effectiviteit als zij voldoen aan methodologische randvoorwaarden en eisen inzake bruikbaarheid. Tot slot staat of valt het rondmaken van de regelgevingscyclus met ‘vergelijkbaarheid’: het stellen van vergelijkbare evaluatievragen en het gebruik van vergelijkbare indicatoren. In de volgende paragrafen analyseren wij hoe de twee hoofdelementen van de Europese regelgevingscyclus er wat dit betreft voor staan.

3. *Impact assessments*: de balans

Toen de contouren van *Better Regulation* geschetst waren en ‘*impact assessment*’ was komen bovendrijven als vlaggenschip, laste de Europese Commissie een pilotperiode in.²⁴ Twee jaar lang werden *impact assessments* alleen uitgevoerd op een select aantal wetgevingsdossiers om ervaring met het nieuwe instrument op te doen. In de *Guidelines* van 2005 werd de opgedane ervaring gebundeld en werden belangrijke methodologische knopen doorgehakt.²⁵ Men besloot in te zetten op zowel kwantificering, ‘waar aangewezen’, als ‘multicriteria-analyse’, waarbij verschillende beslisriteria, zoals kosteneffectiviteit, efficiëntie en beleidscoherentie, naast elkaar gezet worden zodat de ‘*trade-offs*’ van de verschillende opties inzichtelijk worden. Deze analyse heeft als doel besluitvormers te voorzien van een vergelijking van de opties op basis van één of meer criteria. Doelmatigheid en doeltreffendheid zijn belangrijke criteria, maar ook meer kwalitatieve criteria, zoals coherentie en ‘mate van bescherming van fundamentele rechten’, komen voor en worden aangemoedigd. In de beginfase gold dat vrijwel alle systematische ex-

21 Europese Commissie 2013.

22 COM(2012)746. Er is bijvoorbeeld een ‘fitness check’ gaande van de raffinage-industrie, waarvan de resultaten naar verwachting in september 2014 gepubliceerd zullen worden, zie <http://ec.europa.eu/energy/observatory/oil/doc/refining/20130505_refining_fitness_check.pdf>.

23 Poptcheva 2013.

24 Europese Commissie, ‘Actieplan “Vereenvoudiging en verbetering van de regelgeving”’, 5 juni 2002, COM(2002)278; Europese Commissie, ‘Mededeling over effectbeoordeling’, COM(2002)276.

25 Europese Commissie, *Impact Assessment Guidelines*, SEC(2009)92; zie ook Europese Commissie, ‘Mededeling “Better Regulation for Growth and Jobs in the European Union”’, COM(2005)97.

anteanalyse een verbetering was.²⁶ Mede om die reden werd de permanente spaagaat waarin *impact assessments* ‘Europese stijl’ verkeerden, tussen puur informatie-instrument en besluitvormingsinstrument, als weinig problematisch ervaren.²⁷ Terwijl de kwaliteit omstreden was, legde de Commissie de nadruk op het ‘leereffect’ van *impact assessments*. De Commissie heeft ditzelfde leereffect steeds aangevoerd ter rechtvaardiging van de keuze om *impact assessments* intern uit te voeren en hooguit onderdelen te laten uitbesteden – een belangrijk verschil met ex-postevaluaties, die vaak door externen uitgevoerd worden.

Hoewel meerdere onderzoekers oordelen hebben geveld over de *kwaliteit* van Europese *impact assessments*,²⁸ bestaat er geen ‘standaardwerk’ over dit onderwerp. Hooguit valt de consensus te ontwaren dat het methodologisch niet goed mogelijk is om de totale nettomeerwaarde van *impact assessment* in de EU te onderzoeken, in tegenstelling tot de situatie in de Verenigde Staten (VS),²⁹ omdat in de EU het type bevindingen sterk varieert per *impact assessment*. Dit heeft alles te maken met de specifieke insteek van het instrument zoals dit in de EU-context is geïmplementeerd: het hoofddoel is niet om doelmatigheid of doeltreffendheid in kaart te brengen, maar om politieke besluitvormers in staat te stellen beter geïnformeerde keuzes te maken.³⁰ Het motto van de Europese Commissie luidt dan ook nog steeds: ‘IA is an aid to and not a substitute for political decision-making.’³¹ Een aantal consequenties van dit uitgangspunt blijft echter wringen.

Zo werd vanaf 2005de roep om onafhankelijke toetsing van *impact assessments*, zowel in het Europees Parlement als bij *business stakeholders*, steeds sterker. In 2006 heeft de Commissie daarom de Impact Assessment Board (IAB) opgericht: een interne toetsingsinstantie, die bestaat uit op persoonlijke titel benoemde senior-ambtenaren. Inmiddels stuurt de IAB ongeveer een derde van de *impact assessments* ter verbetering terug naar het directoraat-generaal (DG), dat eerste

26 A.C.M. Meuwese, *Impact Assessment in EU Lawmaking*, Alphen aan den Rijn: Kluwer Law International 2008.

27 C.M. Radaelli, ‘Whither Better Regulation for the Lisbon Agenda?’, *Journal of European Public Policy* 2007, p. 190-207.

28 Renda 2006; Cecot e.a. 2008; P. Ladegaard, ‘Measuring RIA Quality and Performance’, in: C. Kirkpatrick & D. Parker (red.), *Regulatory Impact Assessment: Towards Better Regulation?*, Cheltenham: Edward Elgar 2007, p. 56-71; N. Lee & C. Kirkpatrick, *A Pilot Study on the Quality of European Commission Extended Impact Assessment*, Manchester: Impact Assessment Research Centre, Institute for Development Policy and Management, University of Manchester 2004; C.M. Radaelli & F. De Francesco, *Regulatory Quality in Europe: Concepts, Measures, and Policy Processes*, Manchester: University Press 2007; C.M. Radaelli & F. De Francesco, ‘Indicators of Regulatory Quality’, in: C. Kirkpatrick & D. Parker (red.), *Regulatory Impact Assessment: Towards Better Regulation?*, Cheltenham: Edward Elgar 2007, p. 36-55.

29 R.W. Hahn, J.K. Burnett, Y.I. Chan, E.A. Mader & P.R. Moyle, *Assessing the Quality of Regulatory Impact Analyses*, Washington, DC: AEI-Brooking Joint Center for Regulatory Studies 2000; R.W. Hahn & C. Cecot, *The Economic Significance of ‘Insignificant’ Rules* (WP no. RA-06), Washington, DC: AEI Brooking Center 2006; R.W. Hahn & P.C. Tetlock, ‘Has Economic Analysis Improved Regulatory Decisions?’, *Journal of Economic Perspectives* 2008, p. 67-84.

30 Meuwese 2008.

31 Europese Commissie 2002.

auteur was. Ondanks het feit dat het werk van de IAB positief werd beoordeeld,³² is de Europese Commissie momenteel op zoek naar een wijze om de Board verder te professionaliseren, bijvoorbeeld door met duidelijke beoordelingsmaatstaven te gaan werken.³³ In 2014 lijkt het Europese systeem van *impact assessment* op een kruispunt te zijn aangekomen. Het instrument *impact assessment* is geaccepteerd en geïnternaliseerd in alle gelederen, ook die van de externe *stakeholders*, maar critici blijven vraagtekens zetten bij de meerwaarde van het instrument en de kwaliteit van de uitvoering. Terwijl de eerste jaren vooral kritiek kwam op de mate en diepgang van kwantificering,³⁴ betreft de kritiek nu met name het gebrek aan transparantie, het feit dat heel wat analyse ‘verloren gaat in de vertaling’ richting politieke besluitvormers,³⁵ en de gebrekkigheid van het wetenschappelijke bewijs.³⁶

Sinds de al genoemde pilotperiode is de reikwijdte van *impact assessment* wel geleidelijk uitgebreid. Waar eerst nog alle initiatieven uit het Werkprogramma (Commission Legislative and Work Programme (CLWP)) een *impact assessment* moesten ondergaan, werden dat later ‘alle wetgevende initiatieven’ om te voorkomen dat DG’s initiatieven buiten het Werkprogramma zouden proberen te houden, om zo de IA-verplichting te ontduiken. Toen de druk toenam om ook gedelegeerde regelgeving, gedelegeerde maatregelen (art. 290 van het Verdrag betreffende de werking van de Europese Unie (VWEU)) en implementatiemaatregelen (art. 291 VWEU) binnen het IA-kader te brengen, werd de totale investering in ex-ante-evaluatie wel heel groot. Sindsdien is men op zoek naar manieren om de ‘echt belangrijke’ regelgevende initiatieven te selecteren. Een kwantitatieve ondergrens, zoals de \$ 100 miljoen die geldt in de VS, is lastig bruikbaar in de Europese context, omdat *impact assessments* daar juist bedoeld zijn om meerdere opties in kaart te brengen. De laatste versie van de *Guidelines*, uit 2009,³⁷ spreekt van wetgevingsinitiatieven (inclusief gedelegeerde regelgeving) met duidelijk identificeerbare economische, maatschappelijke en milieugerelateerde effecten en ‘niet-wetgevingsinitiatieven’ (bijvoorbeeld mededelingen en programma’s) die toekomstig beleid definiëren.³⁸ In ieder geval moeten DG’s hun voornemens om al dan niet een *impact assessment* te starten, uiteenzetten in een *Roadmap*, een voorloper van een *impact assessment*. Deze praktijk wordt gemonitord door het

32 J.B. Wiener & A. Alemanno, ‘Comparing Regulatory Oversight Bodies across the Atlantic: The Office of Information and Regulatory Affairs in the US and the Impact Assessment Board in the EU’, in: P. Lindseth & S. Rose Ackerman, *Comparative Administrative Law*, Cheltenham, UK: Edward Elgar 2010.

33 Informeel interview met Commissie-ambtenaar, 19 mei 2014, Brussel.

34 Renda 2006; Cecot e.a. 2008.

35 Informeel interview met Commissie-ambtenaar, 19 mei 2014, Brussel.

36 ‘EU twisting facts to fit political agenda, chief scientist says’, interview met Anne Glover. Zie <www.euractiv.com/sections/eu-priorities-2020/eu-twisting-facts-fit-political-agenda-chief-scientist-says-302399>, geraadpleegd op 28 mei 2014.

37 Impact Assessment Guidelines, Commissiemededeeling, 15 januari 2009, SEC(2009)92. Men is momenteel bezig met nieuwe *Guidelines*; deze zouden veel korter moeten worden dan de huidige, maar wel voorzien worden van twee onderliggende ‘lagen’ van meer praktische tools en *best practices* op het intranet van de Commissie.

38 Zie ook <http://ec.europa.eu/smart-regulation/impact/index_en.htm>.

Secretariaat-Generaal. Vandaag de dag worden ook *impact assessments* uitgevoerd voor gedelegeerde regelgeving, zij het niet stelselmatig.³⁹

Hoe ziet het IA-kader er vandaag de dag uit? Zoals gebruikelijk is voor IA-systemen wereldwijd⁴⁰ hanteert de Commissie een stappenplan waarin vanuit een grondige probleemdefinitie en scherpe formulering van de doeleinden, op zowel overkoepelend als operationeel niveau (SMART-doelen), analyse van meerdere opties plaatsvindt. De overkoepelende vraag die in een *impact assessment* aan de orde komt, is: welke *trade-offs* zijn verbonden aan welke beleidsopties in het licht van de probleemdefinitie en de doeleinden? Consultatieprocessen binnen de Europese Commissie zijn nauw verbonden met het IA-proces, of zouden dat moeten zijn. Een voor de regelgevingcyclus belangrijk aspect van het IA-stappenplan is dat elk IA-rapport ook een paragraaf over 'evaluatie en monitoring' moet bevatten. Er moet dus in het stadium van het ontwerpen van regelgeving al worden nagedacht over de wijze waarom een paar jaar later evaluatie zal plaatsvinden en welke data eventueel in de tussentijd verzameld zullen worden. De *Guidelines* maken – op dit punt en op andere – onderscheid tussen 'niet-wetgevingsinitiatieven' (dan moeten *monitoring* en evaluatie vooral breed worden gehouden), *cross-cutting* wetgevingsinitiatieven (waarbij men denkt aan algemene indicatoren), smalle wetgevingsinitiatieven (indicatoren die zich richten op algemene en specifieke doeleinden) en comitologie, oftewel wat we nu eerder gedelegeerde wetgeving zouden noemen (waarvoor concrete indicatoren die een afspiegeling zijn van specifieke en operationele doeleinden worden aanbevolen).⁴¹ Toch lijkt, afgaande op de door de Commissie geselecteerde *best practices*,⁴² het devies vooral te zijn 'hoe concreter, hoe beter'.⁴³ Zo wordt een *impact assessment* waarin voor elk operationeel doeleinde al precies staat aangegeven welke indicatoren gebruikt zullen worden om over enkele jaren de doeltreffendheid te meten, aangemerkt als *best practice*. In theorie voorzien de *Guidelines* ook in het faciliteren van de regelgevingcyclus: '[e]nsure that evaluations are designed and timed in a way that the results can be used as input for future impact assessments'.⁴⁴ In de praktijk zijn

39 A. Alemanno & A.C.M. Meuwese, 'What Role for Impact Assessment in Non-Legislative Rule-Making?', *European Law Journal* 2013, p. 76-92. Zie ook achtereenvolgende rapporten van de IAB: Impact Assessment Board, 'Report for the Year 2009', SEC(2009)1728; Impact Assessment Board, 'Report for the year 2010', SEC(2011)126; Impact Assessment Board, 'Report for the year 2011', SEC(2012)101.

40 OESO, *Regulatory Impact Analysis: A Tool for Policy Coherence*, Parijs: OESO 2009.

41 Impact Assessment Guidelines, 2009, p. 16-17.

42 Europese Commissie, 'Impact Assessment Freight Transport Logistics Action Plan', Commissie-mededeling van 18 oktober 2007, SEC(2007)1320. Zie voor meerdere 'best practices': <http://ec.europa.eu/smart-regulation/impact/commission_guidelines/best_pract_lib_en.htm#_60>.

43 Impact Assessment Guidelines, 2009, p. 28: 'Link objectives with your future monitoring and evaluation activities: without clear objectives you cannot monitor and evaluate whether your policy is on track. SMART objectives are needed to define good indicators. This in turn will allow you to monitor progress and evaluate the extent to which you have achieved your objectives.' Zie ook p. 49 en 50, <http://ec.europa.eu/smart-regulation/impact/commission_guidelines/docs/iag_2009_en.pdf>.

44 Impact Assessment Guidelines, 2009, p. 49.

ons echter (nog) geen voorbeelden bekend van *impact assessments* die expliciet aansluiten bij ex-postevaluaties of andersom.

4. Ex-postevaluatie: de feiten op een rij

De Europese Commissie kent een gedecentraliseerd evaluatiesysteem. Dit houdt in dat de verantwoordelijkheid voor ontwerp, programmering en uitvoering van evaluaties bij de afzonderlijke DG's ligt.⁴⁵ Zij worden ondersteund door een netwerk van evaluatoren, die zorgen voor kennisuitwisseling en professionalisering.⁴⁶ Coördinatie van dit systeem berustte oorspronkelijk bij DG Budget, maar is met ingang van 2009 overgenomen door het Secretariaat-Generaal, dat al langer de verantwoordelijkheid droeg voor *impact assessments*. Veel evaluaties worden extern uitbesteed op basis van aanbestedingsprocedures. Daarnaast voert de Commissie regelmatig zelf evaluaties uit.⁴⁷

Oorspronkelijk betrof ex-postevaluatie met name financiële programma's, zoals de structuurfondsen en het cohesiebeleid, ontwikkelingssamenwerking en wetenschapsbeleid.^{48,49} Centraal in dergelijke programma-evaluaties staat de vraag welke resultaten zijn bereikt met Europees geld. Recentelijk is het besef bij de Europese Commissie ontwikkeld dat systematische evaluatie ook relevant is voor andere typen beleidsinstrumenten, waaronder regelgeving en *soft law*.⁵⁰ In 2000 is echter bepaald dat alle belangrijke activiteiten van de Europese Commissie geëvalueerd moeten worden.⁵¹

De Europese Commissie definieert evaluaties als 'reliable and objective evidence-based assessments of the results and impacts of its interventions in the light of the needs they aim to satisfy'.⁵² Met resultaten wordt hierbij niet alleen bedoeld op *outcomes* van interventies, maar ook op *outputs*, in de zin van stappen in het implementatieproces dat een interventie met resultaten moet verbinden⁵³ – met name de juridische implementatie in lidstaten. De Commissie probeert op deze manier verschillende doelen te bereiken. Evaluaties van programma's dienen allereerst om een verantwoorde basis voor budgettaire besluitvorming te vormen.⁵⁴ Daarnaast noemt de Commissie als doeleinden: democratische verantwoording en verbetering van het ontwerp en de kwaliteit van interventies.⁵⁵ Als een afgeleide hiervan dienen evaluaties ertoe om de legittimiteit van besluiten te

45 Summa & Toulemonde 2002, p. 414.

46 Summa & Toulemonde 2002, p. 414.

47 'Evaluation Standards and Good Practice', Commissiemededeling, 23 december 2002, COM(2002)5267, p. 3.

48 Fitzpatrick 2012, p. 479; Stern 2009, p. 69.

49 Summa & Toulemonde 2002, p. 409.

50 Fitzpatrick 2012, p. 479.

51 Stern 2009, p. 71.

52 Europese Commissie 2000.

53 Fitzpatrick 2012, p. 480.

54 Fitzpatrick 2012, p. 479.

55 Stern 2009, p. 72; Europese Commissie 2004, p. 9.

vergroten.⁵⁶ Hierbij geldt dat de nadruk door de tijd heen is verschoven van summatieve evaluatie, gericht op het afleggen van verantwoording over bereikte beleidsprestaties, naar een meer *formatieve* rol van evaluaties: evaluaties met als doel het verbeteren of bijstellen van regelgeving.⁵⁷

Het beleidskader voor het uitvoeren van evaluaties bestaat uit drie pijlers. Voor programma's wordt de basis gevormd door de *Financial Regulation*.⁵⁸ Daarnaast bestaan er sectorspecifieke evaluatiebepalingen, bijvoorbeeld op het vlak van de structuurfondsen en evaluatiebepalingen in specifieke regelgeving, vaak op aandringen van het Europees Parlement en de Raad van de Europese Unie. Ten derde hanteert de Commissie eigen richtlijnen voor de uitvoering van evaluaties, zowel procesmatig als inhoudelijk. De meest recente richtlijnen zijn vervat in een Communicatie uit 2007, die een herzien evaluatiekader en kwaliteitscriteria biedt voor alle typen ex-postevaluatie.^{59,60} Deze richtlijnen betreffen de organisatie, de planning, het ontwerp, de uitvoering en de disseminatie van evaluaties. Hoewel een centraal uitgangspunt van de richtlijnen de bruikbaarheid voor doeleinden van de besluitvormers – dat wil zeggen de Commissie zelf – betreft, heeft de Commissie in haar evaluatiekader ook oog voor de kennisbehoeften van *stakeholders*. Om deze reden wordt in de evaluatierichtlijnen veel aandacht besteed aan randvoorwaarden om tot betrouwbare, robuuste en complete resultaten te komen. Zo moeten evaluaties gebaseerd zijn op *evidence* en goede analyse, staat onafhankelijkheid van evaluatoren centraal en moeten de resultaten gebaseerd zijn op geschikte methoden, betrouwbare data en deugdelijke analyse.⁶¹

Ondanks de duidelijke boodschap vervat in de interne richtlijnen is er veel kritiek op de kwaliteit van ex-postevaluaties. Een belangrijk zwak punt van de evaluaties lijkt, net als bij *impact assessments*, de objectiviteit van de rapporten te zijn. Volgens Stern⁶² bestaat in de Europese evaluatiegemeenschap de perceptie dat de Commissie onafhankelijkheid van externe evaluatoren niet bepaald hoog in het vaandel heeft staan.⁶³ Deze observatie⁶⁴ is in lijn met bredere zorgen die geuit zijn door Anne Glover, de belangrijkste wetenschappelijke adviseur van de Commissie. Zij stelde onlangs dat de Commissie wetenschappelijke *evidence* vaak dienstbaar maakt aan politieke doeleinden, hetgeen onder meer komt doordat externe adviesbureaus weinig prikkels hebben om daadwerkelijk onafhankelijk te

56 Europese Commissie 2007.

57 Fitzpatrick 2012, p. 479.

58 Stern 2009, p. 70.

59 Europese Commissie 2007.

60 Fitzpatrick (2012) bekritiseert de gedachte dat wetgevingsevaluaties op één lijn gesteld moeten worden met andere typen evaluatie. Zij stelt dat er eigenlijk een speciale methodologie zou moeten worden ontwikkeld, geënt op het evalueren van de werking en effecten van wetgeving.

61 Europese Commissie 2007, p. 23-24.

62 Stern 2009, p. 72.

63 Zie ook R.A.J. van Gestel & H.-W. Micklitz, 'Comparative Law and EU Legislation: Inspiration, Evaluation or Justification', in: M. Adams & D. Heirbaut (red.), *The Method and Culture of Comparative Law* (essays in honour of Mark van Hoecke), Oxford: Hart Publishing 2014, p. 301-317.

64 Stern (2009) staft deze observatie niet empirisch; deze lijkt te zijn gestoeld op informele gesprekken in de Brusselse evaluatiewereld. Letterlijk stelt hij: 'However, there is a widespread perception in the evaluation community that independence is not always highly valued.'

opereren.⁶⁵ Een andere vraag die zich voordoet, betreft het systematische aspect – in hoeverre voldoet de Commissie aan haar eigen, sinds 2000, gestelde eis dat alle belangrijke instrumenten aan ex-postevaluatie moeten worden onderworpen? Tot dusverre was er een gebrek aan systematisch inzicht in deze twee aspecten, terwijl dit cruciale voorwaarden zijn voor het optimaal functioneren van de regelgevingcyclus. Dit heeft deels te maken met het ontbreken van een systematisch evaluatiesysteem of database.⁶⁶ In het navolgende trachten wij de kennislacune op deze aspecten deels te dichten door gebruik van twee eigen, nieuwe datasets met evaluatierapporten en Europese regelgeving.

De eerste dataset bevat 160 ex-postwettelijke evaluaties uitgevoerd door of in opdracht van de Europese Commissie.⁶⁷ De bekeken periode loopt van begin 2000 tot oktober 2012. *Impact assessments* en evaluaties van beleidsprogramma's zijn bij voorbaat buitengesloten, evenals studies die alleen kijken naar de interne aangelegenheden van de Commissie of naar buitenlands beleid. Wij hebben alleen evaluaties opgenomen die het effect van regelgeving op Europese burgers, bedrijven of andere overheden betreffen.⁶⁸ Studies die alleen in het Frans verschenen, zijn om praktische redenen ook weggelaten.⁶⁹ Rapportages van de Commissie die alleen resultaten van andere evaluaties presenteren, zijn ook verwijderd om te voorkomen dat studies dubbel werden opgenomen. De onderliggende studie is in deze gevallen wel opgenomen in de dataset. Interim-evaluaties zijn wel opgenomen, tenzij er een zeer vergelijkbare ex-poststudie beschikbaar was, aangezien er bij evaluaties van regelgeving geen duidelijk verschil tussen interim en ex post bestaat.

De evaluaties zijn verzameld uit verschillende bronnen: de *multi-annual overview of evaluation* van de Europese Commissie,⁷⁰ evaluatieoverzichten van de Commissie over specifieke jaren,⁷¹ werkprogramma's van de Commissie⁷² en lijsten van evaluaties gepubliceerd op de websites van DG's. Dit is aangevuld met een reeds

65 'EU twisting facts to fit political agenda, chief scientist says', interview met Anne Glover. Zie <www.euractiv.com/sections/eu-priorities-2020/eu-twisting-facts-fit-political-agenda-chief-scientist-says-302399>, geraadpleegd op 28 mei 2014.

66 Summa & Toulemonde 2002, p. 410.

67 Omdat onze nadruk ligt op de evaluatiefunctie van de Europese Commissie, hebben wij geen evaluaties door individuele overheden of andere *stakeholders* opgenomen in de dataset.

68 De achterliggende gedachte hierbij is dat ex-postevaluatie met name belangrijk is gezien de gebrekkige legitimiteit van de EU. Dit maakt dat *output*-legitimiteit – legitimiteit van het resulterende beleid – belangrijk wordt. Deze legitimiteit kan worden vergroot door middel van evaluaties, aangezien zij een basis kunnen vormen voor de verbetering van slecht werkende regelgeving.

69 Door een verschil in taalbeheersing zouden de Franstalige evaluaties niet op dezelfde manier kunnen worden beoordeeld, hetgeen tot een probleem met betrouwbaarheid zou kunnen leiden. Dientengevolge zijn zes evaluatierapporten verwijderd uit de subset.

70 Europese Commissie, 'Multiannual Overview (2002-2009) of Evaluations and Impact Assessments', Secretariaat-Generaal, mei 2010, vinddatum 1 december 2011 via <http://ec.europa.eu/dgs/secretariat_general/evaluation/docs/multiannual_overview_en.pdf>.

71 Europese Commissie, lijst van evaluaties 2010-2011, vinddatum 1 december 2011 via <http://ec.europa.eu/smart-regulation/evaluation/index_en.htm>.

72 Europese Commissie, Commissie werkprogramma's, vinddatum 24 april 2012 via <http://ec.europa.eu/atwork/programmes/index_en.htm>.

E. Mastenbroek, A.C.M. Meuwese & S. van Voorst

bestaande lijst van onderzoeksbureau Eureval, die beschikbaar werd gesteld op ons verzoek. In totaal werden 96 evaluaties uit deze bronnen verzameld. Vervolgens is er via de zoekmachine Google systematisch gezocht naar evaluaties van alle verordeningen en richtlijnen tussen 1996 en 2004, wat nog eens 50 resultaten opleverde. Ten slotte is de zoekmachine van evaluaties van de Commissie geraadpleegd die tijdens 2012 beschikbaar werd, waarmee nog 14 evaluaties werden gevonden.⁷³

De tweede dataset is specifiek bedoeld om vast te stellen hoeveel Europese wetgeving wordt geëvalueerd. Gezien het uitgangspunt van de Commissie dat de uitgaven aan evaluatie proportioneel moeten zijn aan het belang van de wetgeving in kwestie,⁷⁴ hebben wij in deze dataset alleen belangrijke wetgeving opgenomen. Amendementen, codificaties, rectificaties en herroepingen werden om die reden bij voorbaat buitengesloten. Hetzelfde geldt voor gedelegeerde wetgeving⁷⁵ en wetgeving van de Raad op uitvoeringsniveau. Wetgeving rond interne aangelegenheden en buitenlands beleid werd ook buitengesloten, vanwege onze gerichtheid op de legitimiteit van regelgeving. Vanwege de arbeidsintensieve dataverzameling is de dataset beperkt tot de jaren 2000 tot en met 2002,⁷⁶ hetgeen in totaal 156 belangrijke verordeningen en richtlijnen opleverde.

Over de inhoud van de evaluaties valt het volgende op te merken. 89 van de 160 gevonden evaluaties (56%) zijn te kenmerken als productevaluaties. Dit wil zeggen dat hun doel is om een oordeel te vellen over de resultaten die met de regelgeving bereikt is. 2 van deze 89 studies beogen uitsluitend te kijken of de doelen van de wetgeving gehaald zijn. Door 39 evaluaties wordt ook een effectiviteitsvraag gesteld: heeft de wetgeving een bijdrage geleverd aan het bereiken van de gestelde doelen? 31 studies kijken naar efficiëntie: tegen welke kosten heeft de wetgeving wel of niet haar doelen bereikt? 13 studies kijken in brede zin naar de effecten van hun object. In deze evaluaties wordt niet alleen aandacht gegeven aan de efficiëntie van de wetgeving in kwestie, maar ook aan neveneffecten op bijvoorbeeld het milieu of de economie.

De overige 71 rapporten zijn te kenmerken als procesevaluaties. Zij betreffen de transitie (24 gevallen), invoering (2 gevallen), uitvoering (25 gevallen), nale-

73 De zoekmachine van evaluaties van de Commissie is te vinden via <<http://ec.europa.eu/smart-regulation/evaluation/search/search.do>>.

74 Stern 2009, p. 71.

75 Wij zijn ons ervan bewust dat gedelegeerde wetgeving ook belangrijke gevolgen kan hebben, evenals wijzigingen van richtlijnen of verordeningen. Door deze selectie bevat de dataset dan ook waarschijnlijk niet *alle* belangrijke wetgeving uit de jaren 2000-2002. Echter, wij hebben ons laten leiden door het doel te komen tot een dataset met *alleen* belangrijke wetgeving, in plaats van *alle* belangrijke wetgeving.

76 De keuze voor deze periode is een uitvloeisel van de keuze voor de periode waarin evaluaties zijn verzameld (2000-2002). Om de initiatie van evaluatie valide te operationaliseren is het belangrijk om geen evaluaties over het hoofd te zien. Aangezien het gemiddeld zo'n vijf jaar duurt voor een ex-postevaluatie wordt uitgevoerd, hebben wij een ruime marge genomen. De kans dat regelgeving uit de jaren 2000-2002 alsnog wordt geëvalueerd – als dat binnen tien jaar niet gebeurd is – kan als klein worden bestempeld. Het verdient wel aanbeveling om deze periode, naarmate de tijd vordert, op te rekken, omdat de evaluatiepraktijk aan verandering onderhevig kan zijn.

Tabel 1. *Type evaluatie in de dataset*

Type evaluaties	Aantal studies	Percentage
Geen duidelijke onderzoeksvraag	4	3%
Transpositie	24	15%
Invoering	2	1%
Uitvoering	25	16%
Naleving	16	10%
Handhaving	4	3%
Doelbereiking	2	1%
Effectiviteit	39	24%
Efficiëntie	31	19%
Effecten	13	8%

ving (16 gevallen) en handhaving (4 gevallen) van Europese verordeningen en richtlijnen, zonder een oordeel te geven of met de wetgeving iets is bereikt. Tabel 1 biedt een overzicht van de verschillende soorten aangetroffen evaluaties.

De Europese Rekenkamer heeft reeds vastgesteld dat een vrij beperkt deel van de ex-postevaluaties gaat over regelgeving. Van alle ex-postevaluaties van de Commissie uit 2007 had bijvoorbeeld slechts 24% betrekking op wetgeving.⁷⁷ Met onze data kunnen we een systematischer beeld geven van de dekkingsgraad van ex-postwetsevaluatie.⁷⁸ De dataset laat zien dat van de 156 belangrijke verordeningen en richtlijnen aangenomen tussen 2000 en 2002 er 44 gevallen één keer zijn geëvalueerd en 4 gevallen twee keer. Deze twee groepen vormen samen 31% van de dataset, wat betekent dat bijna 7 uit de 10 belangrijke verordeningen en richtlijnen achteraf nooit zijn onderzocht. Dit resultaat stemt redelijk overeen met het percentage verordeningen met significante impact dat volgens de Commissie zelf is geëvalueerd (29%).⁷⁹ Tussen de 108 verordeningen en richtlijnen die niet zijn geëvalueerd, zitten verschillende politiek significante onderwerpen, zoals de richtlijn over hulp aan illegalen (2002/90), de richtlijn over privacy en elektronische communicatie (2002/58) en de richtlijn met betrekking tot Ozon in de lucht (2002/3).

De kwaliteit van de evaluaties is gemeten aan de hand van acht criteria.⁸⁰ In de eerste plaats moet de studie goed zijn afgebakend. Hiervoor is gekeken naar de

77 Europese Rekenkamer 2010, p. 42.

78 De data die hier worden beschreven, zijn reeds gepresenteerd in een conferentiepapier: E. Mastenbroek & S. van Voorst, 'Towards Better Problem-Solving in the European Union? Analyzing Ex-Post Evaluation of EU Legislation', NIG Annual Conference, 28-29 november 2013, Enschede.

79 Europese Commissie 2013.

80 19 van de 160 evaluaties zijn door twee van de auteurs beoordeeld om de intercodeurbetrouwbaarheid te bepalen. De waarde van Kappa lag voor geen van de acht criteria lager dan 0,650 en was in alle gevallen significant. Dit duidt op een redelijk tot hoge intercodeurbetrouwbaarheid.

aanwezigheid van een *heldere probleemstelling*.⁸¹ maakt de probleemstelling van het rapport duidelijk op basis van welk criterium er wordt geëvalueerd (zoals doelbereiking, effectiviteit of efficiëntie)? Daarnaast moet de interne validiteit van de evaluatie in orde zijn, waarvoor is gekeken naar de aanwezigheid van een *heldere operationalisatie*.⁸² wordt er vóór de presentatie van de onderzoeksresultaten in het rapport duidelijk gemaakt welke indicatoren zijn gebruikt? Om de externe validiteit te meten is per rapport vastgesteld of er een *representatieve selectie van lidstaten* is en een *representatieve keuze van cases*. Dit betekent concreet dat een eventuele selectie van lidstaten en cases moet zijn toegelicht in het rapport. Tevens is gekeken of de *response rate* voor eventuele surveys of interviews meer was dan 50%.⁸³ Voor betrouwbaarheid is gekeken naar de *herhaalbaarheid van het onderzoek*: zijn alle lijsten van vragen, respondenten en gebruikte documenten toegevoegd in de tekst of als bijlagen, zodat het empirisch onderzoek desgewenst opnieuw kan worden uitgevoerd?⁸⁴ Ook is het van belang dat de studie een robuuste methodologie heeft.⁸⁵ Daarvoor is bekeken of de evaluatie een *beargumenteerde keuze voor methoden* maakt. Ten slotte is de objectiviteit van een studie relevant.⁸⁶ Om dit in kaart te brengen is gekeken of er ten minste twee van de volgende vier typen onderzoeksmethoden zijn toegepast: inhoudsanalyse, observatie, enquêtes of interviews (*triangulatie*). *Stakeholder-consultatie* is hier beschouwd als een vorm van interviews indien deze gebruikt is om empirisch materiaal voor de evaluatie te verzamelen. Opgeteld vormen de criteria een schaal van 1 tot 8.

Tabel 2 toont de scores op de verschillende kwaliteitscriteria. Zoals te zien is, scoren relatief veel cases op een representatieve selectie van landen en triangulatie. Daarbij moet wel worden opgemerkt dat met name Noord- en West-Europese landen goed zijn vertegenwoordigd. Portugal is bijvoorbeeld maar in zeven studies meegenomen als specifieke casus en België in dertien, terwijl deze landen qua bevolkingsomvang even groot zijn. Vooral als het gaat om herhaalbaarheid en methodologische verantwoording scoren de meeste evaluaties slecht. Zoals te zien in de tabel zou slechts iets meer dan een kwart van de studies te herhalen zijn, terwijl slechts 17% van de rapporten de keuze voor de gebruikte methoden verantwoordt.

81 R. Schwartz & J. Mayne, 'Assuring the Quality of Evaluative Information: Theory and Practice', *Evaluation and Programme Planning* 2005, p. 6.

82 R. Adcock & D. Collier, 'Measurement Validity: A Shared Standard for Qualitative and Quantitative Research', *American Political Science Review* 2001, p. 529.

83 Dit is een gemiddelde van de waarden die te vinden zijn in de literatuur, die varieerden van 20 tot 80%. J.C. Valentine, 'Judging the Quality of Primary Research', in: H. Cooper & L.V. Hedges (red.), *The Handbook of Research Synthesis*, 2de druk, New York: Russel Sage Foundation 2009, p. 135; F.J. Fowler, *Survey Research Methods*, 4de druk, New York: Sage Publications 2009, elektronische versie vindbaar op <<http://srmo.sagepub.com/view/survey-research-methods/n1.xml?row=6&rskey=GLWnNo>>.

84 N. Golofshani, 'Understanding Reliability and Validity in Qualitative Research', *The Qualitative Report Volume* 2003, p. 599.

85 Schwartz & Mayne 2005, p. 6.

86 Schwartz & Mayne 2005, p. 6.

Tabel 2. Scores op de kwaliteitsaspecten

Kwaliteitsaspect	Aantal scorende rapporten	%
Representatieve landselectie	104	65
Triangulatie	97	61
Operationalisatie	92	58
Probleemdefinitie	90	56
Response	76	48
Representatieve casusselectie	74	46
Herhaalbaarheid	45	28
Methodologische verantwoording	27	17

Naast de inhoudelijke kwaliteit van de studies is ook gekeken naar bruikbaarheid voor beleidsmakers. Dit is gebeurd met behulp van twee criteria. Het eerste criterium is de aanwezigheid van heldere aanbevelingen:⁸⁷ noemt het rapport in zijn conclusie interventies die kunnen worden gepleegd om de wetgeving te verbeteren die in de macht liggen van de Europese Commissie? Het tweede criterium is de aanwezigheid van een heldere samenvatting.⁸⁸ Deze samenvatting mag maximaal tien pagina's zijn.⁸⁹ Rapporten die zelf minder dan tien pagina's lang zijn, scoorden automatisch op dit criterium, omdat deze hoe dan ook beknopt genoeg zijn om te lezen. De data lieten zien dat 113 evaluaties (71%) duidelijke aanbevelingen bevatten voor de Europese Commissie. Er zijn 103 evaluaties (64%) met een heldere samenvatting. Gemiddeld genomen scoort dus een hoger aantal evaluaties op deze bruikbaarheidscriteria dan op de criteria van inhoudelijke kwaliteit.

5. Bevindingen en conclusie

Hoge kwaliteit en een hoge dekkingsgraad in beide fasen van de regelgevingcyclus zijn noodzakelijke voorwaarden om deze goed te laten functioneren. Het *impact assessment*, dat als instrument voor de voorbereiding van regelgeving binnen de Europese Commissie al zo'n tien jaar heeft kunnen rijpen, heeft vandaag de dag eerder een kwaliteitsprobleem dan een kwantiteitsprobleem. In hoeverre dat kwaliteitsprobleem ook een imagoprobleem is, valt bij gebrek aan een systematische studie moeilijk vast te stellen. Feit is dat er allerlei concrete gevallen aan te wijzen zijn waarin de *assessments* niet solide waren.⁹⁰ Het sterker concentreren van middelen ten behoeve van 'belangrijke regelgevingsinitiatieven' zou kwaliteitsverho-

87 K. Forss & J. Carlsson, 'The Quest for Quality – Or Can Evaluation Findings Be Trusted?', *Evaluation* (3) 1997, afl. 4, p. 495.

88 Poptcheva 2013, p. 4.

89 University of South California, *Executive Summary*, zonder datum, vinddatum 10 mei 2014 op <<http://libguides.usc.edu/content.php?pid=83009&sid=1481087>>.

90 Meuwese 2008, p. 185 e.v.; F. Vibert, *The Itch to Regulate: Confirmation Bias and the European Commission's New System of Impact Assessment*, Londen: European Policy Forum 2005.

gend kunnen werken, maar heeft als risico dat het omzeilen van elke effectbeoordeling veel eenvoudiger wordt.

De dekkingsgraad van ex-postwetsevaluaties loopt sterk achter bij die van *impact assessments*. Wij vonden dat 69% van de belangrijke verordeningen en richtlijnen tot op heden niet ex post is geëvalueerd, waar de dekking van *impact assessments* wat wetgevingsinitiatieven betreft vrijwel volledig is. Deze *impact assessments* zijn echter geen afdoende alternatief, aangezien zij een beperktere feitenbasis hebben dan evaluatie van de daadwerkelijke impact van regelgeving.

Het kwaliteitsprobleem dat bij *impact assessments* aanwezig lijkt te zijn, lijkt ook te bestaan bij de ex-postwetsevaluaties die tot op heden zijn uitgevoerd. De gevonden evaluatierapporten scoren met name laag op de criteria die het belangrijkste zijn voor objectiviteit – representatieve casusselectie, herhaalbaarheid en methodologische verantwoording. Voor een daadwerkelijk goed functioneren van de regelgevingscyclus is dit een probleem, aangezien het juist de rol van ex-postwetsevaluatie zou kunnen zijn om te achterhalen in hoeverre de verwachte effecten zich daadwerkelijk voordoen.

Als we de kaders en het soort vragen dat in beide evaluatie-exercities gehanteerd worden,⁹¹ vergelijken, valt allereerst op dat het vrij strakke kader voor *impact assessments*, in de vorm van de *Guidelines* en een stevige institutionele inbedding, bij de ex-postwetsevaluaties minder ontwikkeld is. De Europese Commissie is al enige tijd bezig met de herziening van dit beleidskader; een exercitie die in 2014 haar beslag zou moeten krijgen. Verder zijn de functies die de twee typen ‘evaluaties’ bekleden in het proces duidelijk verschillend: voor goed *impact assessment* is het van belang dat meerdere opties serieus worden onderzocht; bij ex-postwetsevaluatie is specificiteit juist aangewezen: wat werkt wel of niet en waarom? Dit soort informatie kan vervolgens worden gebruikt voor een scherpe probleemdefinitie in een *impact assessment*. Omgekeerd kunnen ex-postwetsevaluaties ertoe dienen om assumpties omtrent impacts, daadwerkelijk te toetsen.

Op basis van de eerste stappen die wij gezet hebben in ons onderzoek naar de Europese regelgevingscyclus, komen alvast de volgende verbeterpunten naar voren. Er waren al zorgen over de kwaliteit van *impact assessments*. Deze zorgen blijken ook voor ex-postwetsevaluaties op hun plaats. Dit creëert het beeld dat de Europese Commissie voor de gehele regelgevingscyclus moeite heeft om duidelijk te maken wat ‘analyse’ is en wat ‘politieke besluitvorming’. Het voorstel van de *chief scientific advisor* om *assessments* los te koppelen van politieke dynamiek heeft institutioneel gezien nogal wat voeten in de aarde, maar is een belangrijke stap naar voren.⁹² In ieder geval zou moeten worden bekeken of tot dusver op de achtergrond gebleven ‘verantwoordingsfora’, zoals de Europese Rekenkamer, geen rol

91 R.A.J. van Gestel & J.B.M. Vranken, ‘Assessing the Accuracy of Ex Ante Evaluation through Feedback Research: A Case Study’, in: J. Verschuuren (red.), *The Impact of Legislation*, Leiden: Nijhoff 2009, p. 199-228.

92 ‘EU twisting facts to fit political agenda, chief scientist says’, interview met Anne Glover. Zie <www.euractiv.com/sections/eu-priorities-2020/eu-twisting-facts-fit-political-agenda-chief-scientist-says-302399>, geraadpleegd op 28 mei 2014.

zouden kunnen krijgen bij concrete ex-postwetsevaluaties en *impact assessments*.⁹³ Tevens is te overwegen om de rol van *stakeholders* in evaluatieprocessen te vergroten, om een eenzijdige insteek van de evaluatie te voorkomen. Tot slot lijkt, om de geloofwaardigheid van het Europese ex-postevaluatiesysteem te garanderen, een verzekering van basale methodologische kwaliteitscriteria voor evaluaties op haar plaats.

Wat de dekkingsgraad betreft scoren *impact assessments* beter dan ex-postwets-evaluaties. Het is niet duidelijk op welke gronden de Commissie al dan niet overgaat op een ex-postwetsevaluatie. Een aanbeveling is dan ook het instellen van een verantwoordingsplicht voor de Commissie jegens het Europees Parlement over de evaluatieagenda: op welke gronden komt de selectie van te evalueren regelgeving tot stand en hoe worden *impact assessments* hierbij ingezet? Het is tevens aan te bevelen dat de Europese Commissie een systeem van ex-postevaluatie ontwikkelt, waarmee in ieder geval regelgeving met significante effecten – zoals voorspeld in het *impact assessment* – systematisch wordt geëvalueerd. Voor wetgeving die zo buiten de regelgevingscyclus valt, omdat aan het begin geen *impact assessment* is gemaakt, zou een vorm van petitie tot evaluatie mogelijk moeten zijn.

Tot slot leidt dit artikel tot enkele vragen voor nader onderzoek. Het onderzoek naar ex-postwetsevaluatie in de EU staat in de kinderschoenen. De gesignaleerde variatie in de kwaliteit en initiatie, allereerst, vraagt om nadere uitleg. Onder welke omstandigheden wordt geëvalueerd? En hoe kunnen verschillen in evaluatiekwaliteit worden verklaard? Om deze vragen te beantwoorden zal moeten worden ingegaan op de methodologische problemen met ex-postwetsevaluatie in een EU van 28 lidstaten, maar ook op de politieke context waarin evaluatie plaatsvindt. Daarnaast zouden de verzamelde rapporten op inhoud kunnen worden geëvalueerd, zoals dat ook is gebeurd in de Nederlandse context.⁹⁴ Een andere belangrijke vraag betreft het gevolg van evaluaties: in welke mate werken zij door in verantwoordingsprocessen, in hoeverre leiden zij daadwerkelijk tot herziening van wetgeving, en wat is de relatie met infractieprocedures door de Europese Commissie?

Nadere aandacht vergt ook de wisselwerking tussen *impact assessments* en ex-postwetsevaluaties. Ten eerste zou het interessant zijn om te bezien in hoeverre er *impact assessments* en ex-postevaluaties al inhoudelijk aan elkaar gerelateerd zijn, aangezien het eerste instrument reeds enkele jaren tot het standaardrepertoire van de Europese wetgever behoort. Zijn er al ex-postevaluaties die een onderliggend *impact assessment* als uitgangspunt hebben genomen, of *impact assessments* die de resultaten van een ex-postevaluatie gebruiken? Ten tweede zou het aardig zijn om internationaal vergelijkend onderzoek uit te voeren naar de

93 De algemene ‘audit’ van het IA-systeem die de Europese Rekenkamer eerder heeft verricht, verliep enigszins problematisch, maar toetsing van concrete *assessments* zou wel eens meer in het straatje van deze instelling kunnen passen. Zie A.C.M. Meuwese, ‘The European Court of Auditors Steps out of Its Comfort Zone with an “Impact Assessment Audit”’, *European Journal of Risk Regulation* 2011, p. 104-107.

94 E. Niemeijer & C.M. Klein Haarhuis, ‘Synthesising Legislative Evaluations – Putting the Pieces Together’, *Evaluation* 2009, p. 403-425.

E. Mastenbroek, A.C.M. Meuwese & S. van Voorst

theoretische en feitelijke verbinding van de twee instrumenten. Ten derde zou systematisch onderzoek naar de kwaliteit van *impact assessments*, vergelijkbaar met het onderzoek naar ex-postwetsevaluatie dat wij in dit artikel gepresenteerd hebben, een interessante kijk op de Europese praktijk van *Better Regulation* kunnen bieden. Want de regelgevingcyclus kan slechts werken wanneer deze gebaseerd is op een overtuigende feitenbasis.