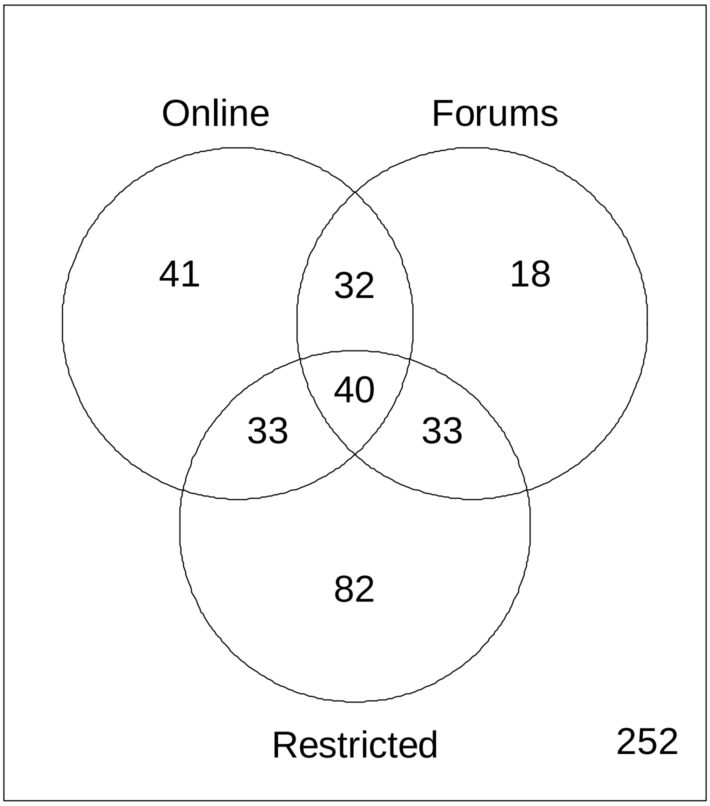
Online appendix to

The Effect of Stakeholder Involvement: Consultation of External Actors & Legislative Duration

Part I. Illustrating the operationalization of consultations

Originally, all cases are coded (non-exclusively) for the presence of online [146], forum [123], gated [116], advice [84], EU agencies [22], management and labor [6], and informal [9] types of consultations, or none at all. The gated, advice, EU agencies, management and labor, and informal types are then combined into a single category to which we give the name 'restricted' [188 cases]. This leaves us with three non-exclusive categories with many overlaps (see the Venn diagram below).

Figure A1. Venn diagram of consultation types



For the final categorization we put:

1) cases that had ‘restricted’ **and either** ‘online’ or ‘forum’ consultation into a ‘combined’ category [106 cases];
2) cases which had **only** ‘restricted’ consultation into a ‘restricted’ category [82 cases];
3) cases which had either ‘online’ **or** ‘forum’ **but no** ‘restricted’ consultation into an ‘open’ category [91 cases];
4) the residual category is ‘no consultation’ [252 cases].

Part II. Non-proportionality and the Cox survival models

The Cox models need to satisfy the assumption of proportionality. The formal tests conducted indicate problems with some of the variables (and specific categories of the main independent variable). However, closer inspection and refitting the models to different subsets of the data reveal that the problem is confined to the observations with very long durations only. Hence, we do not include interactions with time since they will provide misleading inferences for the entire dataset, but we note that the models have a worse fit, and the effect of the variables drops for observations with very long durations.

First, the following table lists the results of the formal Grambsch and Therneau (Grambsch and Therneau, 1994) proportionality tests for the entire model and each of the categories of the main independent variable.

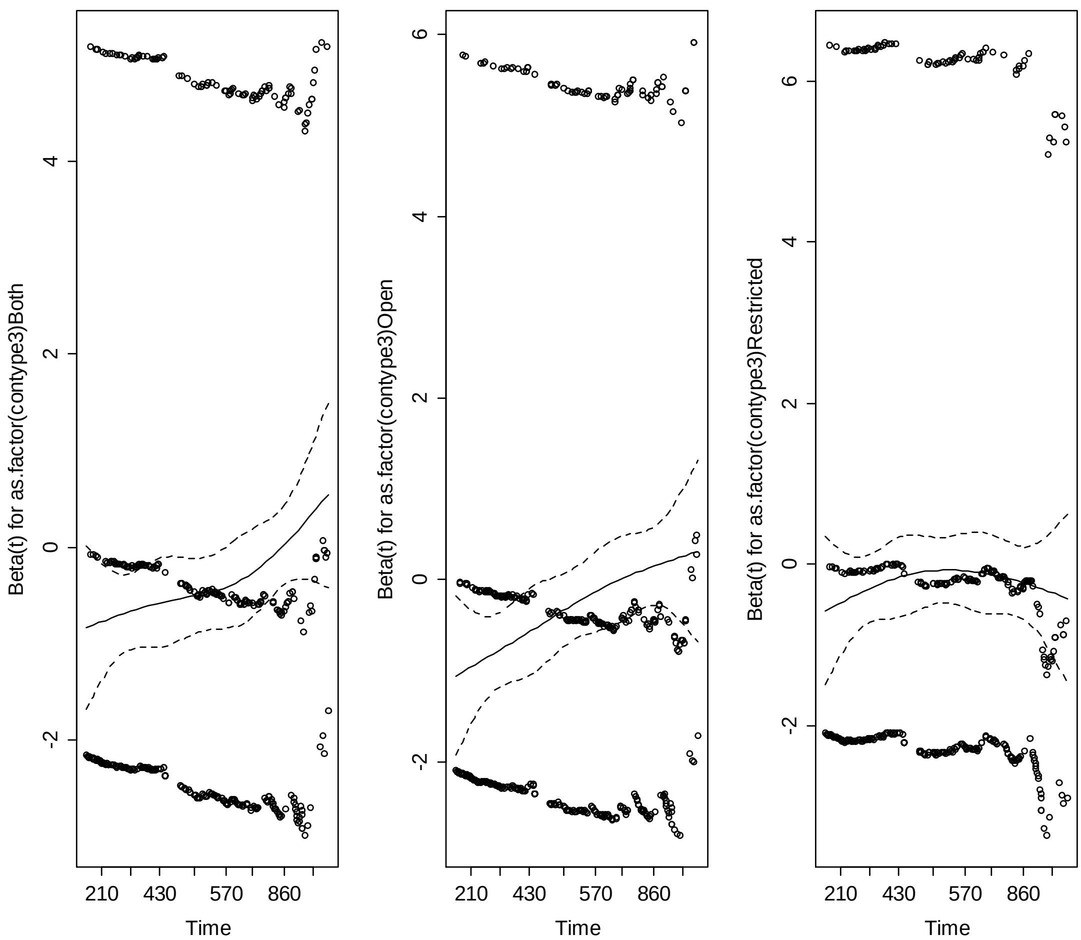
Table A1. Proportionality tests for the models reported in the main text

		Model 1	***Model 2***	***Model 3***
		rho (p-value)	rho (p-value)	rho (p-value)
Type of consultation:				
	Open	**0.15 (0.002)**	**0.14 (0.002)**	0.08 (0.09)
	Restricted	0.02 (0.72)	0.06 (0.22)	0.04 (0.34)
	Combined	**0.12 (0.01)**	**0.12 (0.007)**	**0.09 (0.04)**
Global test		NA (0.003)	NA (0.003)	NA (<0.001)

The tests indicate potential problems with non-proportionality for the Open and Combined categories. However, in the full Model 3 the problems are at the borderline of statistical significance (the chi-square distribution is used in the tests). More importantly, however, a visual inspection of the plots of the scaled Schoenfeld residuals vs. duration reveals that the pattern of the effect over time is complex: while there is negative tendency in the beginning, the pattern reverses for observations with very long duration (to the right of the x-axis). Using the most common method of correcting for violations of the non-proportionality assumption, i.e. interacting the offending variables with a linear, log or quadratic function of time - produces misleading results. Hence, it suggests that in the beginning of the observation period, the effect of consultation on legislative duration is negative (decreases duration) and only later turns into a positive one (increases duration). This inference is misleading because it is not consistent with the data (see below).

Figure A2. Plots of the scaled Schoenfeld residuals vs. duration

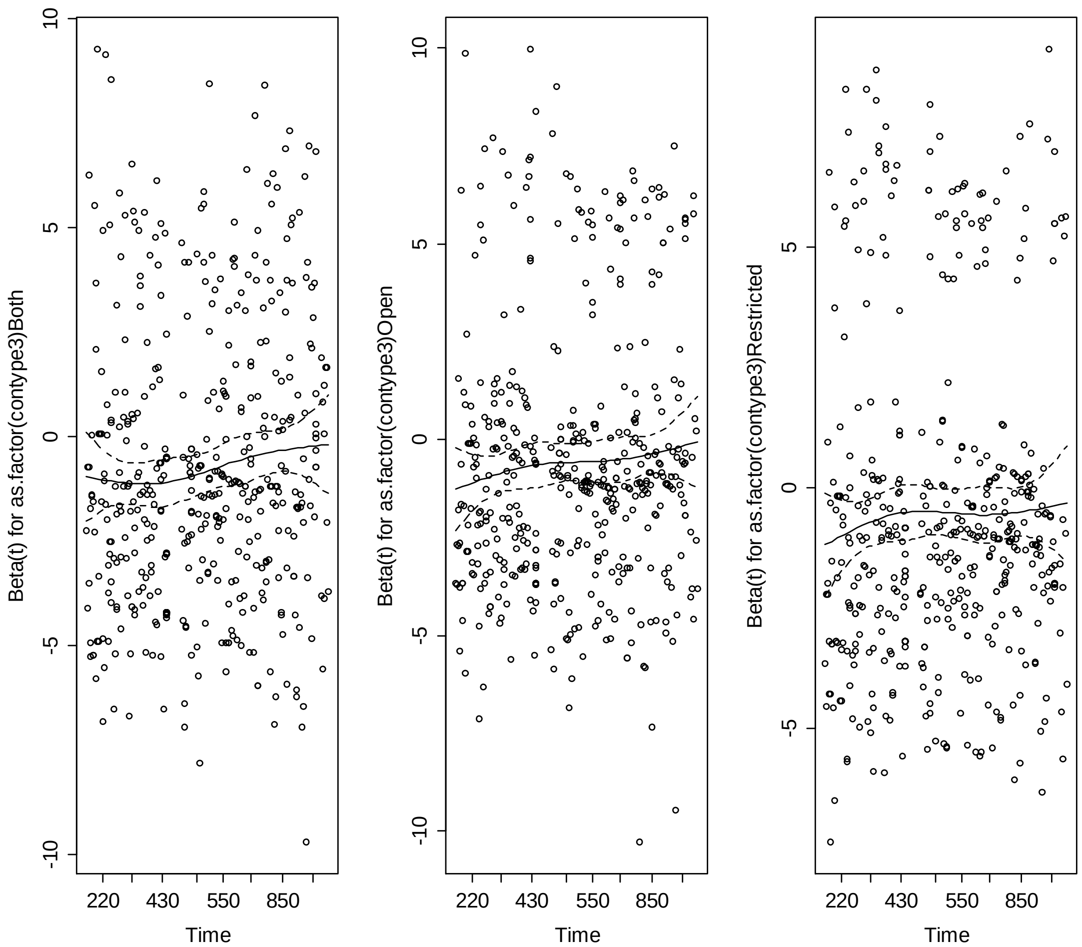
Model 1



The problem arises because we impose on the data a smooth (linear, log, quadratic) function of time while the real non-proportionality follows a rather complex pattern. What is true is that the effect of conducting consultations gets smaller over time in the full model (see the plot for Model 3). This implies that the model fits cases with average and small duration better than the extreme cases. However, in the cases with extreme duration, errors in record keeping and data coding, censoring, and other 'special' reasons are likely to play a bigger role, so this is not a problem for our inferences as such.

Figure A3. Plots of the scaled Schoenfeld residuals vs. duration

Model 3



In order to explore deeper the potential problem with non-proportionality we re-fit the models to subset of the original data (we only show the results for the consultation variable, but otherwise the models are specified as above). Table A2(1) refits the models to four non-overlapping subsets of the data depending on the duration outcome (0 to 1st quantile, 1st quantile to median, median to 3rd quantile, 3rd quantile to maximum). Table A2(2) refits the models to all observations up to the median duration, and to all observations up to the 3rd quantile of duration (the details of full dataset models are included for comparison).

Table A2. The Cox models fitted to subsets of the data (1)

	0 to 1st quantile duration Coeff. (st. error)	*1st quantile to median duration* Coeff. (st. error)	*Median to 3rd quantile duration* Coeff. (st. error)	*3rd quantile to maximum duration* Coeff. (st. error)
Model 1				
Open	-0.38 (0.30) *p-value=0.13*	-0.36 (0.26) *p-value=0.16*	**0.29 (0.25)** ***p-value=0.25***	**0.18 (0.28)** ***p-value=0.52***
Restricted	-0.34 (0.23) *p-value=0.20*	-0.57 (0.28) *p-value=0.04*	**0. 85 (0.28)** ***p-value=0.003***	**-0.20 (0.31)** ***p-value=0.52***
Combined	-0.21 (0.25) *p-value=0.39*	-0.51 (0.25) *p-value=0.04*	-0.12 (0.24) *p-value=0.61*	***0.05 (0.29)*** ***p-value=0.84***
Pseudo R-square	0.03	0.06	0.08	0.01
N	134	132	132	133
Model 2				
Open	-0.51 (0.40) *p-value=0.20*	-0.54 (0.32) *p-value=0.09*	-0.28 (0.43) *p-value=0.52*	No convergence
Restricted	-0.70 (0.31) *p-value=0.03*	-0.59 (0.34) *p-value=0.08*	**0.07 (0.42)** ***p-value=0.87***	No convergence
Combined	-0.35 (0.40) *p-value=0.37*	-1.20 (0.36) *p-value<0.001*	-0.85 (0.41) *p-value=0.04*	No convergence
Pseudo R-square	0.18	0.38	0.25	No convergence
N	131	128	125	128
Model 3				
Open	-0.21 (0.41) *p-value=0.48*	-0.28 (0.36) *p-value=0.45*	**0.10 (0.47)** ***p-value=0.82***	No convergence
Restricted	-0.79 (0.33) *p-value=0.02*	-0.45 (0.34) *p-value=0.18*	**0.54 (0.45)** ***p-value=0.23***	No convergence
Combined	-0.28 (0.40) *p-value=0.48*	-0.73 (0.43) *p-value=0.09*	-0.65 (0.42) *p-value=0.12*	No convergence
Pseudo R-square	0.24	0.41	0.29	No convergence
N	131	128	119	128

Table A2. The Cox models fitted to subsets of the data (2)

	Full dataset	***0 to Median (523 days) duration***	***0 to 3rd quantile (795 days) duration***
Model 1	Coeff. (st. error)	Coeff. (st. error)	Coeff. (st. error)
Open	- 0.33 (0.13) *p-value=0.01*	-0.39 (0.20) *p-value=0.05*	-0.30 (0.15) *p-value=0.04*
Restricted	- 0.22 (0.13) *p-value=0.10*	-0.19 (0.18) *p-value=0.29*	0.08 (0.15) *p-value=0.56*
Combined	- 0.38 (0.12) *p-value=0.002*	-0.34 (0.18) *p-value=0.05*	-0.37 (0.14) *p-value=0.009*
Pseudo R-square	0.025	0.025	0.029
N	531	266	398
Model 2			
Open	- 0.74 (0.15) *p-value<0.001*	- 0.52 (0.22) *p-value=0.02*	- 0.65 (0.18) *p-value<0.001*
Restricted	- 0.52 (0.15) *p-value<0.001*	- 0.34 (0.22) *p-value=0.12*	- 0.20 (0.19) *p-value=0.29*
Combined	- 0.78 (0.15) *p-value<0.001*	- 0.56 (0.23) *p-value=0.02*	- 0.78 (0.18) *p-value<0.001*
Pseudo R-square	0.249	0.281	0.216
N	512	259	384
Model 3			
Open	- 0.62 (0.15) *p-value<0.001*	- 0.34 (0.25) *p-value=0.16*	- 0.36 (0.18) *p-value=0.04*
Restricted	- 0.60 (0.15) *p-value<0.001*	- 0.30 (0.22) *p-value=0.19*	- 0.21 (0.19) *p-value=0.30*
Combined	- 0.76 (0.15) *p-value<0.001*	- 0.38 (0.16) *p-value=0.15*	- 0.66 (0.18) *p-value<0.001*
Pseudo R-square	0.339	0.333	0.331
N	501	259	378

Table A3. Testing the proportional hazard assumption for the models fitted to the subsets of the data (see Table A2[2])

a. Subset: 0 to Median (523 days) duration

	Model 1	***Model 2***	***Model 3***
	rho (p-value)	rho (p-value)	rho (p-value)
Type of consultation:			
Open	0.02 (0.75)	0.01 (0.85)	-0.02 (0.80)
Restricted	-0.02 (0.72)	0.04 (0.34)	0.06 (0.36)
Combined	-0.02 (0.72)	-0.05 (0.39)	-0.02 (0.76)
Global test	NA (0.939)	NA (0.328)	NA (0.17)

b. Subset: 0 to 3rd quantile (795 days) duration

	Model 1	***Model 2***	***Model 3***
	rho (p-value)	rho (p-value)	rho (p-value)
Type of consultation:			
Open	**0.16 (0.001)**	0.08 (0.08)	0.04 (0.48)
Restricted	**0.13 (0.01)**	0.04 (0.38)	0.06 (0.17)
Combined	0.08 (0.10)	0.03 (0.56)	0.03 (0.52)
Global test	NA (0.003)	NA (0.02)	NA (0.003)

c. Subset: 0 to 1436 days duration (excluding outliers)

	Model 1	***Model 2***	***Model 3***
	rho (p-value)	rho (p-value)	rho (p-value)
Type of consultation:			
Open	**0.13 (0.006)**	**0.12 (0.007)**	0.06 (0.18)
Restricted	0.01 (0.84)	0.05 (0.29)	0.02 (0.63)
Combined	**0.10 (0.02)**	**0.12 (0.01)**	0.08 (0.08)
Global test	NA (0.01)	NA (0.02)	NA (<0.001)

The results in the tables indicate that the effects we find in the whole dataset are similar when we replicate the models to the subsets of observations up to the median duration, and up to the 3rd quantile duration. When possible to estimate, the effects of consultation appear different for the subset of observations with longer than the 3rd quantile durations. In the full model (Model 3) the effects of all types of consultation diminish in size when the model is fit in the subsets, however, the significance for open and combined consultations remains while the one for restricted diminishes.

The proportionality test presented in Table A3 support the conclusions that in the full specification (Model 3) no problems with non-proportionality are evident for the levels of the consultation variable. When the models are fitted to the cases with durations from zero to the median one, all three specifications of the model pass the proportionality test. When the models are fitted to the subset of observations from 0 to the 3rd quantile, both models which include control variables pass the tests (Model 2 and 3). As mentioned above, the removal of extreme cases (longer than 1436 days – 11 cases) is enough to clear possible problems with non-proportionality for the model which includes all controls. Overall, on the basis of the formal tests and the visual inspection of the residuals, our conclusion is that the effect of consultation on duration is rather stable, with the exception of the cases having extreme duration.

However, in Model 3 problems with proportionality remain for some of the control variables, and as a result, for the model as a whole. In response, we turn to another method of correcting for violations of the non-proportionality assumption, i.e. we fit the model with stratification over the offending variables (see the right-most column of Table 2 of the main text). The model is stratified by Commission DG, number of EP amendments (dichotomized at the median), and number of EP committees involved (dichotomized at 1). The stratification removes all evidence of non-proportionality as evident from the formal test:

Table A4. Therneau and Grambsch Proportionality tests for Model 4 reported in the main text

	rho	chisq	p
contype3Both	0.0466	1.0242	0.312
contype3Open	0.0452	0.9903	0.320
contype3Restricted	0.0251	0.2697	0.604
typedirective	0.0283	0.3683	0.544
typeregulation	0.0351	0.5870	0.444
amendmentYes	-0.0110	0.0582	0.809
GLOBAL	NA	2.0463	0.915

As a final precaution against non-proportionality affecting our results, we fit a parametric log-logistic survival model which does not require the assumption of proportional effects (Keele 2010). The results support our conclusions with all types of consultation significantly increasing legislative duration and no significant differences between the types of consultation.

Table A5. Model 5 Log-logistic parametric survival model. Scale-location parameterization. Positive coefficients imply longer duration.

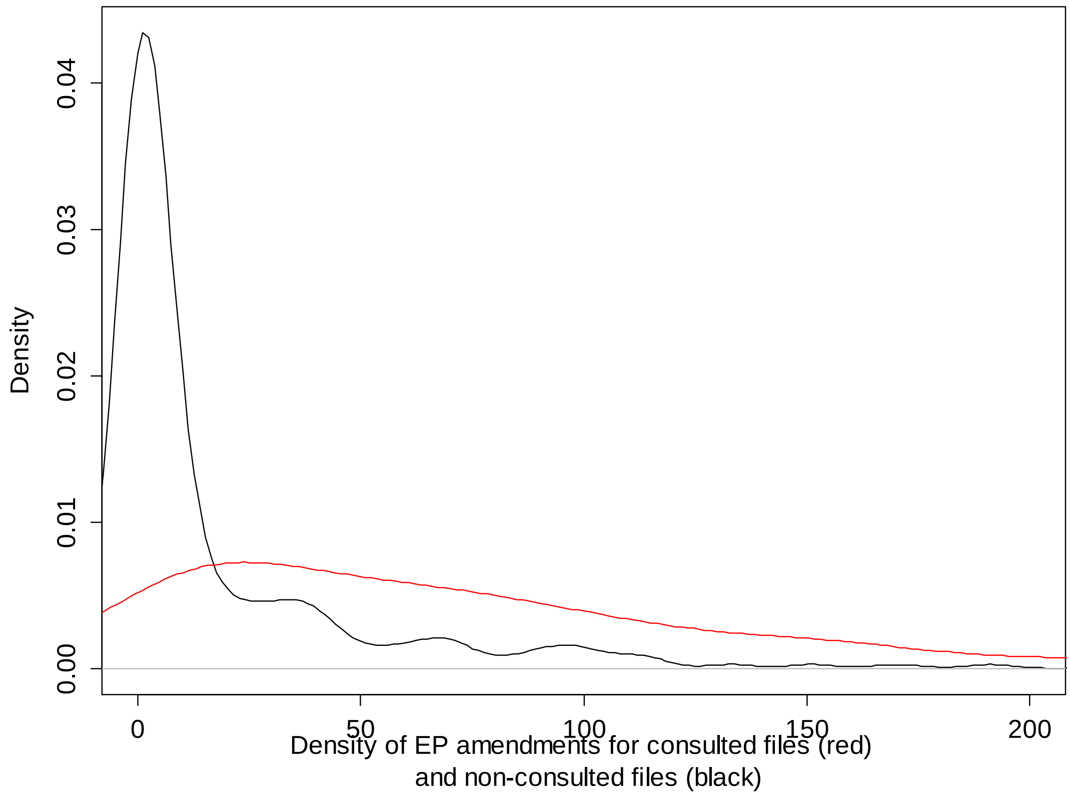
	Value	Std. Error	z	p
(Intercept)	6.89486	0.32094	21.483	***
contype3Both	0.32512	0.06471	5.024	***
contype3Open	0.27327	0.06519	4.192	***
contype3Restricted	0.23594	0.06924	3.408	***
typedirective	0.22294	0.08305	2.684	***
typeregulation	0.11003	0.07617	1.445	
amendmentYes	-0.18305	0.05041	-3.631	***
epamend	0.00101	0.00023	4.376	***
ncom	-0.31318	0.04404	-7.111	***

Part III. Matching analysis

Matching refers to a "variety of procedures that restrict and reorganize the original sample in preparation for a statistical analysis". The goal of matching is to adjust the data prior to the parametric analysis so that the link between the treatment variable and the relevant pre-treatment variables is eliminated or reduced (Ho et al., 2007, 211). Matching can alleviate problems in estimating causal effects due to imbalances and incomplete overlaps. Imbalance refers to situations in which the distributions of the control variables differ for the treatment and non-treatment groups (Gelman and Hill, 2007, 199) while lack of complete overlap implies that some treated units do not have corresponding no-treatment units with regard to the values or categories of some relevant control variable. In short, matching and related techniques address the problem where we have no counterfactuals for the treated observations.

In order to prepare the data for the matching analysis, we dichotomize our main explanatory variable – consultations. Since we saw that there are no significant differences between the different types of consultation, the dichotomization is not problematic. From the many possibilities for preprocessing we choose the full matching procedure, as implemented by the MatchIt library for R (Ho et al., 2007). Exact matching discards too many observations because one of our control variables – the number of EP amendments is continuous. Nearest neighbor matching did not improve the balance between the consultation and no-consultation groups. Full matching achieved excellent improvements in the balance between the groups. In full matching, a fully matched sample is composed of matched sets, where each matched set contains one treated unit and one or more controls (or one control unit and one or more treated units) (Ho et al., 2007). Here is an indication of the imbalances involved in the original dataset and the improvements after matching: Originally, the mean number of EP amendments for files, which had any type of interest group consultation, was 92 vs. 29 for the files which had none. After matching, the corresponding numbers are 92 vs. 97. In other words, the balance improvement is 91%. Similarly, the mean number of committees involved in consultation files was 1.65 vs. 1.48 for the non-consultation dossiers. After matching, the difference is only 0.6 (1.64 vs. 1.71). Figure A4 shows the imbalances in the number of EP amendments between consultation and no consultation files by plotting the observed density of amendments for both groups.

Figure A4. Possible imbalances in the dataset: EP amendments and consultation



The results from the analyses of the unmatched and the matched data are presented in Table A6. The models still support the finding that consultation increases the duration of legislative decision making. Model B1 shows the effect of the occurrence of consultation on the hazard of legislative adoption. As expected the coefficient is negative implying that files that have been consulted with external actors take longer to conclude. Model B2 estimates the same effect but after matching. The effect is still significant, and is very similar in size. Models B3 and B4 include a list of control variables – type of legislative act, number of EP amendments, and the number of EP committees involved (the same variables were also used in the matching procedure). Again the models provide very similar results and the effect of consultation appears rather robust.

Overall, our conclusions remain the same but now, through matching we have the additional safeguard that the results are not driven by imbalances in the data. As suspected in hypothesis 1, involving civil society in legislative preparation does seem to increase the transaction costs of forming legislative coalitions and prolong the legislative processes. In contrast, we did not find evidence for hypothesis 2 that there should be a systematic difference in how much open and restricted types of consultative exercises prolong matters.

Table A6. Comparison survival analysis models with and without matching

	Model B1	*Model B2 (matching)*	*Model B3*	*Model B4 (matching)*
Variable	Coeff. (st. error)	Coeff. (st. error)	Coeff. (st. error)	Coeff. (st. error)
Consultation (Yes)	- 0.42 (0.09) *p-value<0.001*	- 0.40 (0.10) *p-value<0.001*	- 0.47 (0.10) *p-value<0.001*	- 0.39 (0.11) *p-value<0.001*
Type of act (baseline: decision)				
Directive	-	-	- 0.74 (0.16) *p-value<0.001*	- 0.77 (0.16) *p-value<0.001*
Regulation	-	-	- 0.72 (0.16) *p-value<0.001*	-0.83 (0.17) *p-value<0.001*
Number of EP amendments (10)	-	-	- 0.01 (0.00) *p-value=0.018*	- 0.01 (0.1) *p-value=0.40*
Number of EP committees	-	-	0.44 (0.10) *p-value<0.001*	0.65 (0.14) *p-value<0.001*
Distance	-	-0.14 (0.27) *p-value=0.598*	-	- 1.55 (0.78) *p-value=0.05*
Pseudo R-square	0.04	0.04	0.12	0.13

Notes: N=502; dependent variable – hazard of adoption of a co-decision legislative act. Baseline category for Consultation – ‘No consultation’. Baseline category for Type of act – ‘Decision’. Full matching performed.

Part IV. Multiple comparisons

Formal multiple comparison tests (based on Tukey's all-pair comparisons) for the values of the categorical variable 'consultation' [as implemented in the *multcomp* package, for details see Torsten Hothorn, Frank Bretz and Peter Westfall (2008), Simultaneous Inference in General Parametric Models. *Biometrical Journal*, **50**(3), 346–363].

Model A1

```
                         Estimate Std. Error z value Pr(>|z|)
Open - Both == 0          0.06196    0.15060   0.411  0.97602
Restricted - Both == 0  0.17214    0.15515   1.109  0.67922
Restricted - Open == 0  0.11018    0.15874   0.694  0.89752
```

Model A2

```
                         Estimate Std. Error z value Pr(>|z|)
Open - Both == 0          0.03708    0.16391   0.226  0.99588
Restricted - Both == 0  0.26060    0.16985   1.534  0.41455
Restricted - Open == 0  0.22352    0.17596   1.270  0.57969
```

Model A3

```
                         Estimate Std. Error z value Pr(>|z|)
Open - Both == 0          0.14321    0.16403   0.873    0.818
Restricted - Both == 0  0.16463    0.17283   0.953    0.775
Restricted - Open == 0  0.02142    0.17776   0.121    0.999
```

References:

Gelman A and Hill J (2007) *Data analysis using regressson and multilevel/hierarchical models*. Cambridge: Cambridge University Press.

Grambsch P and Therneau T (1994) Proportional hazards tests and diagnostics based on weighted residuals, *Biometrika* 81: 515-26.

Ho D E, Imai K, King G and Stuart E A (2007) Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference, *Political Analysis* 15(3): 199-236.

Keele, Luke. 2010. "Proportionally Difficult: Testing for Nonproportional Hazards in Cox Models." *Political Analysis* 18, no. 2 (2010): 189-205.

Therneau, Terry M., and Patricia M. Grambsch. 2000. Modeling survival data: Extending the Cox model. New York: Springer-Verlag.