Application form NWO - Corona: Fast-track data

Section 1 - Application details

Title of proposal	Controlling COVID-19's infection force: A data-driven, real-time, age-varying model
Applicant	dr.dr. 5.1.2e
E-mail	5.1.2e @uu.nl
Telephone	5.1.2e
Organisation	Utrecht University
Duration of the project (in months)	6 months (4 months core)

Section 2 - Public summary

English public summary

Using several readily available real-time big data sources (e.g., traffic), RIVM data, and Corona app data, we will model, validate, and forecast COVID-19's infection force in the Netherlands in real-time. Our infection force model provides the government reliable, real-time, age-varying forecasts on scenarios considered to control the virus outbreak. So, mortality, pressure on health services, and economic disruption can be reliably balanced against restrictions on behavior (e.g., keep schools and business closed).

word count: 80

Dutch public summary

Wij modelleren, valideren en voorspellen de infectiedruk van het Corona virus. Hierbij maken wij gebruik van real-time big data (bijv. verkeersdata), RIVM-data en Corona app data. Ons model adviseert de overheid met betrouwbare, real-time en leeftijdsafhankelijke voorspellingen bij scenario's die worden overwogen om het virus te controleren. Zo kan een betrouwbare afweging worden gemaakt tussen sterfte, de druk op de gezondheidszorg, economische impact enerzijds en beperkingen in bewegingsvrijheid (bijv. het sluiten van scholen en bedrijven) anderzijds.

word count: 79

Section 3 - Budget

3.1 Budget table

Type of costs	Description	Costs (euros)
Personnel	**research assistant 1 (4 months)**	€ 5.1.1c
	research assistant 2 (4 months)	€
	PI (6 months)	€
Materials	**n.a.**	€
		€
Travel	**n.a.**	€
		€
Total		€

3.2 Budget clarification

2 research assistants will be work full-time for a period of 4 months.

* The PI will initiate (2 weeks) and close (6 weeks) the project and will supervise the core project (4 months). He will work 2 days/week on the project. To enable this, other duties will be taken over by colleagues. The additional required funding to realize this is covered by the PI's department, as has been the case so far.

Application form NWO - Corona: Fast-track data

Section 4 – Data management

4.1 Will data be collected or generated that are suitable for reuse?

Yes

4.2 How and when will the data and results from the research be made publicly available?

Directly after the project has started, we will formulate a Data Management Plan (DMP), which will address the relevant aspects of making data FAIR – Findable, Accessible, Interoperable and Re-usable, including what data the project will generate, whether and how it will be made accessible for verification and re-use, and how it will be curated and preserved. With each inclusion of new types of data or data from new sources, the DMP will be updated.

Metadata will be added to the research data, likely in the Dublin Core standard scheme) to document them and make them searchable when public. The data will be stored together with the metadata and other relevant documentation as a package for a period of 10 years for safekeeping and sharing. Files will be stored in a well-documented format on a secure storage location that has the following characteristics: the data package is stored in multiple copies on at least two physically distinct locations, is regularly checked for integrity, and the storage medium is renewed before the expiration date.

RIVM: We have been in contact with the RIVM and they have already shown interest in our ideas. Therefore, we anticipate that the anonymized data and results will be presented and, subsequently, shared with and published on the RIVM website https://www.databronnencovid19.nl/ . Moreover, in case of technical papers (e.g., in medRxiv and arXiv) and journal articles, accompanying anonymized subsets of data will be published as well.

Luscii: We have been in direct contact with Luscii on both this proposal and the data they gather via the OLVG app. We have already been informed on how to request their data.

4.3 Where will the data be stored during the research?

The original data as well as all its derivatives will be encrypted and stored on the university's secured data servers. The data will be primarily accessible for the PI and the researchers working on the data. In case of possible additional assistance, access to subsets of the data will be granted, after a non-disclosure agreement is signed.

4.4 After the project has been completed, how will the data be stored for the long-term?

The original data as well as all its derivatives will be handed over to the RIVM and will remain stored on the university's secured data servers for a period of 10 years.

4.5 Which facilities (ICT, (secure) archive, refrigerators or legal expertise) do you expect will be needed for the storage of data during the research and after the research? Are these available?[1]

Utrecht University has all computing machinery needed for secure storage, secure transport and reliable computing. All data are both anonymized and encrypted before stored and transported.

Application form NWO - Corona: Fast-track data

Section 5 – Project plan

This project will tackle three challenges:

1) Can we determine the outcome of policy measures in advance? Can we calculate the consequences of relieving our intelligent lock down? Yes, we can!

2) Is a thorough validation process available? Yes, there is, using available app-data.

3) Will it work in real-time? Yes, it will use multiple data sources in real-time to generate models.

The basic reproduction number (R0), combined with expert opinions, is the main parameter used to forecast scenario outcomes. The R0 depends on social interaction and personal hygiene and basically indicates whether or a pandemic grows exponentially. However, R0 is defined as the reproduction number in communities where *anyone* is susceptible; hence, it ignores the effect of herd immunity!

We pose that the infection force should be used, instead of the R0, as it considers the interaction between both infected and susceptible individuals. The infection force λ model is also a flexible, time-varying, and age-frailty model, which is a requirement for many aspects of COVID-19, including issues on immunity and age dependency. Moreover, it allows us to make real-time data-driven estimations of infection force, using multiple real-time data sources and validation by available app data.

We estimate a daily infection force λ of a person of age a_j and time-varying frailty $z_j(a_j)$ by using the following, validated Belgian model (Held, 2020):

$$\lambda\{a_j, z_j(a_j)\} = NDL \int_0^\infty \int_0^\infty \beta\{a_j, z_j(a_j); a_k, z_k(a_k)\}\, \lambda\{a_k, z_k(a_k)\}\, S\{a_k, z_k(a_k)\}\, \phi(a_k) da_k dz_k$$

where N is population size, D the duration of the infectious period, L the mean life expectancy, β the transmission between people of age a_j, frailty $z_j(a_j)$ and age a_k, frailty $z_k(a_k)$, λ the infection force of other people, S the number of susceptible people of age a_k and ϕ the probability of being alive at age a_k.

Conveniently, life expectancy and the frailty-age dependency are known (e.g., via the CBS). However, S and β are unknown and often estimated by outdated questionnaire data and time-consuming observations. Instead, we propose a state-of-the-art estimation technique of β and S by publicly available data sources including RIVM data, NDW traffic data, revenue statistics of bars and cafes, and public data on sporting apps like Runkeeper. This allows accurate, real-time estimation of β allowing real-time estimation of infection force λ.

Using Luscii's 33000+ OLVG and similar app data, we will run an additional maximum likelihood-based cross-validation to calculate a daily estimate of β and S.

This cross-validated, real-time calculation of the infection force λ allows hindsight calculation of the individual impact on infection force of all taken measures. In turn, this allows the government to explore future scenarios to control the virus outbreak by calculating and, hence, balance mortality, pressure on health services, and economic disruption, against restrictions on behavior (e.g., cancel events, keep schools and business closed, and keep people at home).

Feasibility: Despite the good contact with RIVM and Luscii, the main project risk is that RIVM will not adopt our model for some reason. A minor project risk is that some data sources are not good or not accessible, which will possibly slightly widen the model's confidence interval.

Word count: 500

References

Adam D (2020) The simulations driving the world's response to COVID-19: How epidemiologists rushed to model the coronavirus pandemic. *Nature*. URL: https://www.nature.com/articles/d41586-020-01003-6

Adolph C et al. (2020). *Pandemic Politics: Timing state-level social distancing responses to COVID-19*. URL: https://www.medrxiv.org/content/medrxiv/early/2020/03/31/2020.03.30.20046326.full.pdf

Brauer F, Castillo-Chavez C, Feng Z (2019). *Mathematical methods in Epidemiology*. New York: Springer.

Chong, YC (2020). *A novel method for the estimation of a dynamic effective reproduction number (Dynamic-R) in the CoViD-19 outbreak*. URL: https://www.medrxiv.org/content/10.1101/2020.02.22.20023267v1

Held L, Hens N, O'Neill P, Wallinga J (2020) *Handbook of Infectious Disease Data Analysis*. Boca Raton, FL, USA: Taylor & Francis Group / CRC Press.

Imperial College COVID-19 Response Team (2020). *Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand*. URL: https://spiral.imperial.ac.uk:8443/bitstream/10044/1/77482/5/Imperial%20College%20COVID19%20NPI%20modelling%2016-03-2020.pdf

Kupferschmidt K, Cohen J (2020). Can China's COVID-19 strategy work elsewhere? *Science, 367(6482)*, 1061-1062.

Lai A et al. (2020). Early phylogenetic estimate of the effective reproduction number of SARS-CoV-2. *Journal of Medical Virology*. URL: https://doi-org /10.1002/jmv.25723

Lipsitch M et al. (2003). Transmission Dynamics and Control of Severe Acute Respiratory Syndrome. *Science, 300*, 1966-1970.

Reesing H et al. (2020). *Large scale, real-time and longitudinal tele-monitoring ofCOVID-19 in the general population*. URL: https://luscii.com/wp-content/uploads/2018/04/DOC-COV-24032020.pdf

Tuite AR, Fisman DN (2020). Reporting, epidemic growth, and reproduction numbers for the 2019 novel coronavirus (2019-nCoV) epidemic. *Annals of Internal Medicine* DOI: https://doi.org/10.7326/M20-0358

Wu JT, Leung K, Leung GM (2020). Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: A modelling study, *Lancet, 395*, 10225, DOI: https://doi.org/10.1016/S0140-6736(20)30260-9

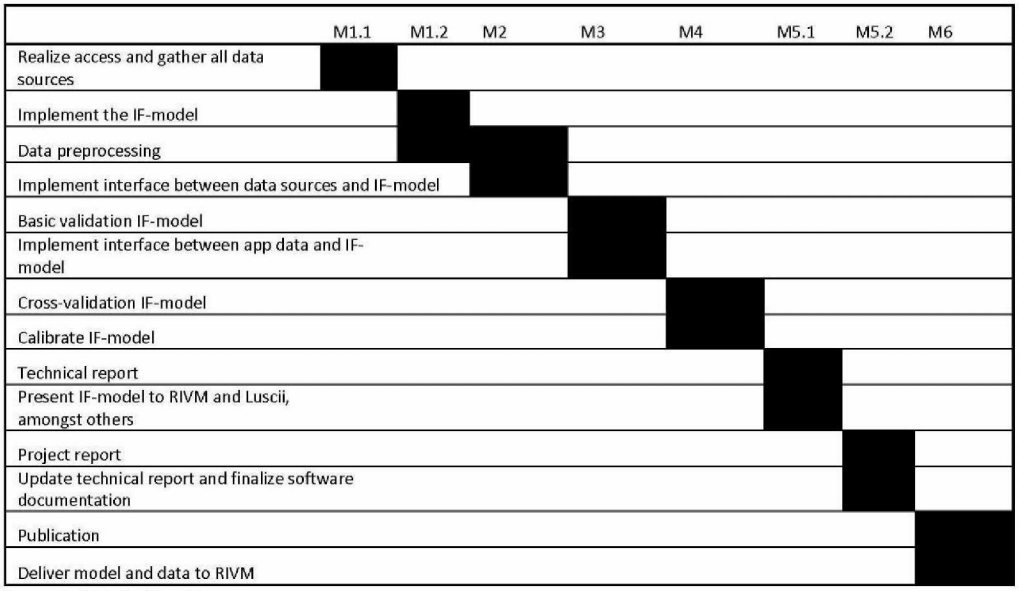
Yu D, Lin Q, Chiu APY, He D (2017). Effects of reactive social distancing on the 1918 influenza pandemic. *PLoS ONE, 12(7)*, e0180545.

Application form NWO - Corona: Fast-track data

Section 6 – Timeline

Literature, model, and initial data sets are identified. **Start:** March 26, 2020.

The PI will start/continue **immediately** after its acceptance.



Legend: IF: Infection Force

Word count: 100

Section 7 – Ethics

We will use 3 types of data:

i) data we will anonymize ourselves;
ii) already anonymized data; and
iii) aggregated data, which is anonymized per definition.

In none of these cases, ethical issues are expected to occur. Moreover, and most importantly, the data will not be analyzed on individual level but at cohort level.

The theoretical framework, research methods, and analysis will be specified in detail after which the proposal will be submitted to the faculty's ethical review board, which will secure a fast-track review of the proposal and will facilitate iterative updates regarding permission for additional data collection.

We will ask the ethical review board to allow us to start immediately with data collection, as all data is readily available and anonymized and we will not gather new data.

Word count: 126

Application form NWO - Corona: Fast-track data

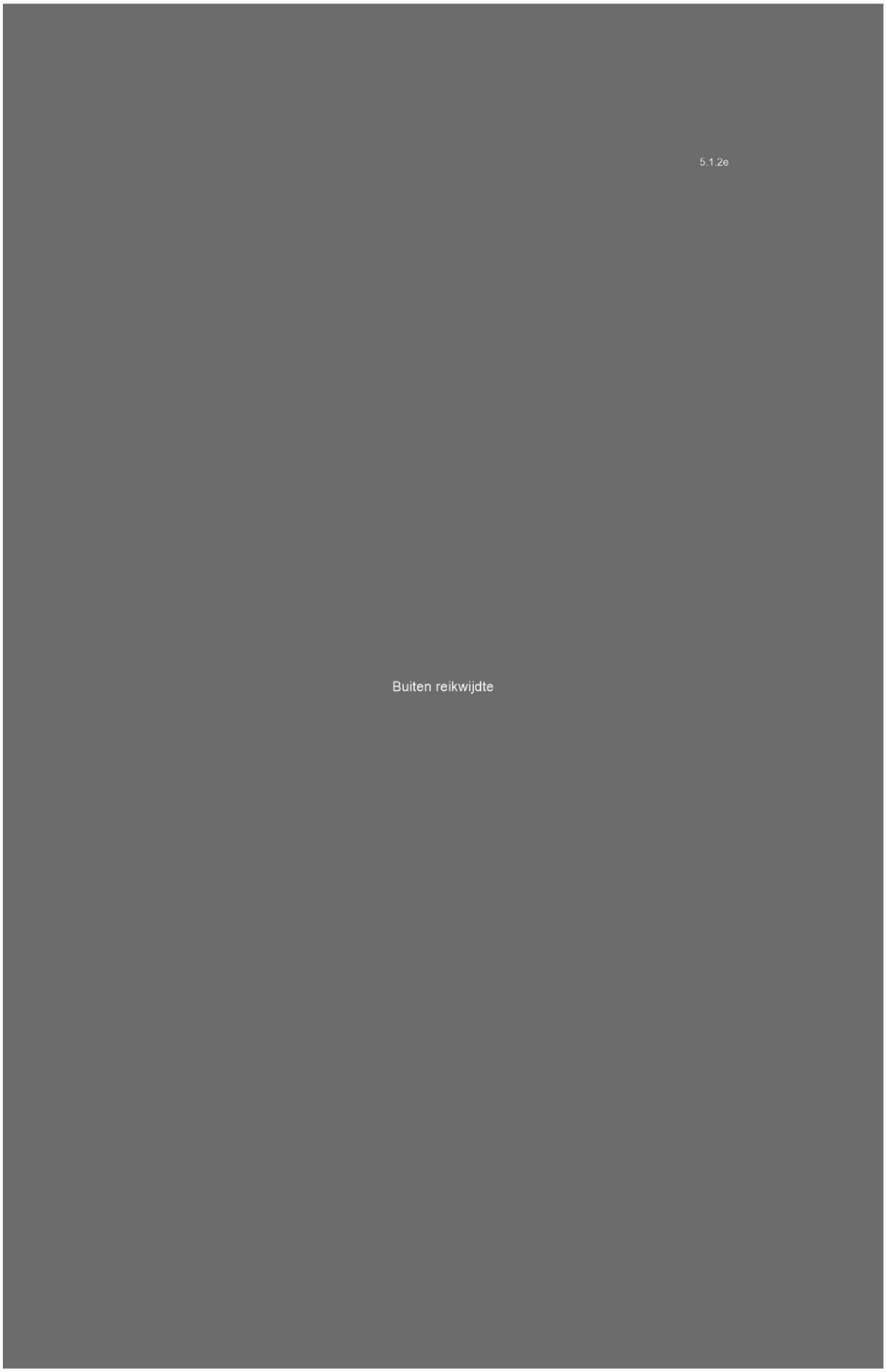
Signature

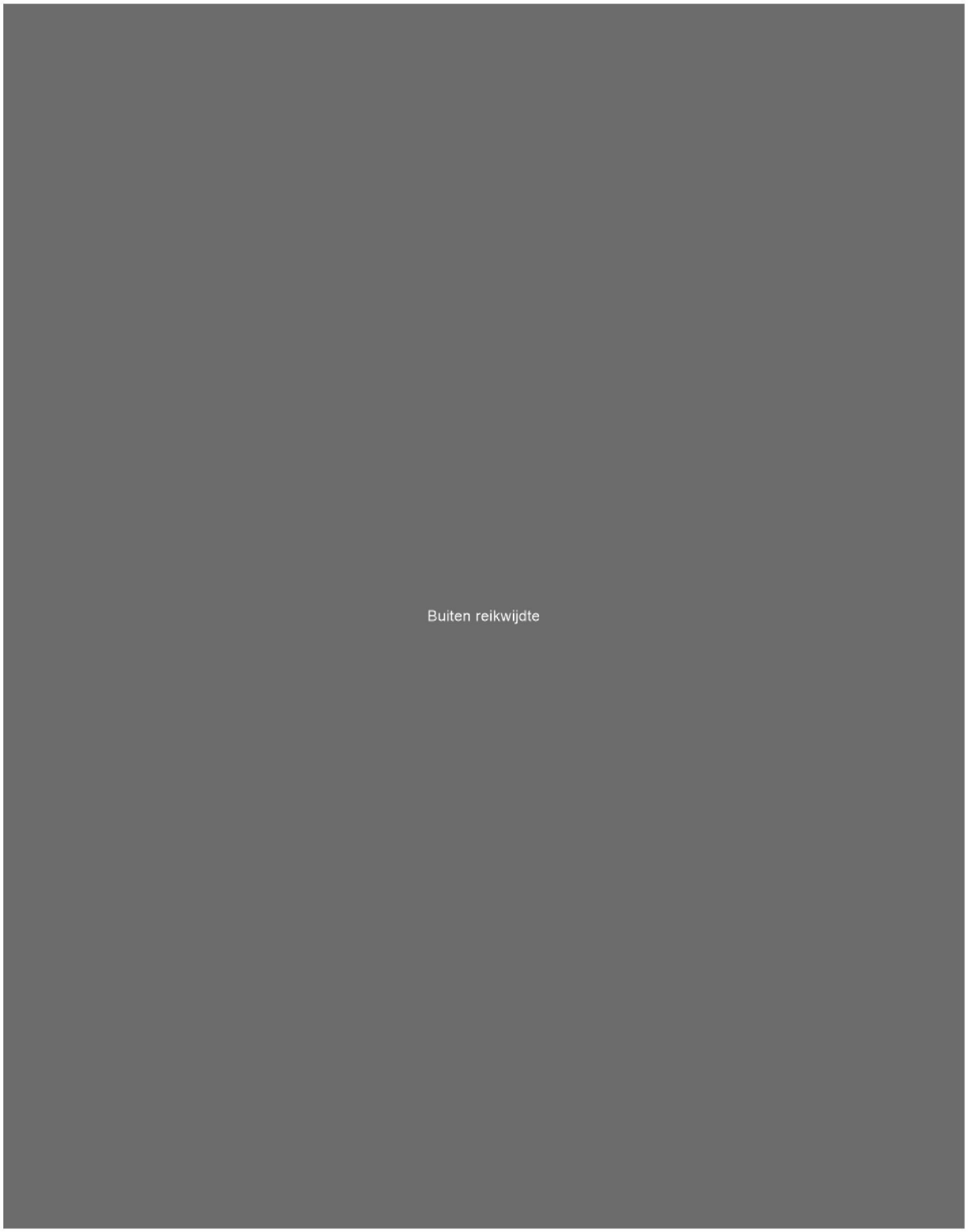
- By submitting this document I declare that I satisfy the nationally and internationally accepted standards for scientific conduct as stated in the Netherlands Code of Conduct for Scientific Practice (Association of Universities in the Netherlands).
- I endorse and follow the Code Openness Animal Experiments (if applicable)
- I endorse and follow the Code Biosecurity (if applicable)
- I have completed this form truthfully.

Name: 5.1.2e
Place: The Hague
Date: 8 April 2020

Appendix: CV

See the next two pages.


5.1.2e
Buiten reikwijdte


Buiten reikwijdte