# Discussion

In this thesis, I aimed to study the effect of several potential targets of intervention related to dementia prevention that have had controversial results in previous observational studies, by applying causal inference theory and corresponding methods. In this section I will outline the principal findings of each project while laying the methodological challenges and the implemented solutions. Next, I will describe the potential future directions and briefly summarize the central points of this dissertation.

## Principal findings and broader implications

The aim in Chapter 2 was to emulate a hypothetical randomized trial - a target trial - for estimating observational analogues to intention-to-treat and per-protocol effects of statins in the risk of dementia. In this study we found that individuals with sustained statin use, but not statin initiation alone had reduced 10-year risks of dementia and dementia or death. Although results should be interpreted with caution, due to the yet small number of statin initiators and number of dementia events, plus potential residual confounding, these findings show how important it is to define and estimate per-protocol effects using observational data.

One of the major and most frequent methodologic flaws in previous observational studies has been the prevalent user bias(cite). This bias refers to the comparison between prevalent users of statins with nonusers, which is subject to selection bias because prevalent users have, by definition, survived under treatment[@danaei2012]. Randomized controlled trials are protected from this bias given that they recruit participants who have not taken statins prior to the study. In contrast, many observational studies do not follow the same eligibility criteria and classify participants by their history or current status of statin use. By emulating the target trial we prevent this bias in two ways: first, by having a clear definition of who would be eligible, which results in excluding prevalent users and second, by having clear definitions of the causal contrast such as “initiating stating treatment” vs. “not initiating statin treatment” or “initiating and sustained statin use” vs “not initiating ever”. Although this has been remarkably emphasized in pharmaco-epidemiology guidelines(cite) and previous methodologic papers (cite) specifically on statins, it is surprising that only few studies have considered this design (cite).

One of the frequent arguments against considering “new users” design is that it may lead to a small sample problems. For example, if we only include participants based on information that was only collected once and we have measured at baseline statin use (such as current indication, length of duration with medication), we may restrict to only those participants who have initiated statin at the time of the measurement and those who have not used it ever (or in a long period before). However, when longitudinal information on statin use is available, we can overcome this situation. In our work, the eligibility criteria was clearly defined to include participants with no statin prescription in the previous two years, and no previous diagnosis of dementia. Since only few participants would be included in the treatment arm at the study baseline, we conceptualized a "sequence of trials". This means that, rather than defining one point in time as the time zero, eligibility criteria was assessed every month between February 1993 and December 2007. This represents 180 trials, each of them with a 1-month enrollment period. As described in Chapter 2, baseline variables are updated at the start of each trial. Data is pooled from all 180 trials into a single model. This design allows us to go from 6373 eligible participants in the study, to 1578655 potential person-trials.

In respect to the per-protocol effect we faced a few limitations. In an RCT participants would take statins over follow-up, and there would be a strict control over adverse-effects over time or contraindications that arise over follow-up. All these would be recorded as standard practice related to safety and quality control. Thus we would be interested in the effect of the sustained use of statins until the outcome of interest or any contradiction over follow-up. In contrast, in this study we faced some limitations: 1) we used data on prescriptions rather than on information on actual intake, 2) we had no information on lack of

Up to the time this dissertation was written (October, 2021), no other research paper has been published answering this or an analogous question considering the time-varying nature of statins intake in a different population. Even worse, new studies such [@Zhou2021], continued to contrasts current-users vs. non-users at baseline, when this trial had over 1700 new statin users over follow-up. However, I do acknowledge that the “sequence of trial” design has limitations in respect to computational challenges and reproducibility that may slow down the process of adopting this method. To performed this analysis we used a SAS macro, developed and accessible at the Causal Inference Lab at Harvard [@]. So far, few applications using this tool have been published and no open source package (such as in R or Python) has been published. Not having open source tools means that not all the research community can access to them. However simpler techniques have proven to show similar results (goodarz), though with less precision. I hope that more network and more educational resources, developed by collaborative work between applied researchers, methodologists and biostatisticians, helps narrow the gap between methods development and applications in a shorter spam of time.

We highlight the importance of phrasing a per-protocol effect question in Chapter 3.

I found a similar gap between methods development and applied research when it comes to studying the effect of hypertension, or better yet, the effect of reducing blood pressure under clinical thresholds and the risk of dementia. In one hand, few randomized controlled trials have looked at the effect of specific antihypertensives, or alternatively, the effect of keeping blood pressure under clinical thresholds, such as the iconic SPRINT-Mind trial. Most trials where originally performed to answer questions around cardiovascular diseases and as such, they had very specific criteria (such as participants with history of stroke or with risk of cardiovascular disease [@sprint]) and small length of follow-up.

In the other hand, we have observational studies that have either looked at the effect of antihypertensives with the design issues we described above (such as the prevalent user design and defined at only one time-point) [@ding2020], described the longitudinal systolic blood pressure patterns across blood pressure categories and outcome level [@rajan2018] or they categorize participants under different cutt-offs of systolic blood pressure, either at baseline or collapsing time-varying information over follow-up as unique categories [@walker2019].

We attempt to bridge these two sides of research in Chapter 3. The aim of this study was to emulate a target trial to estimate the sustained effect of several hypothetical interventions on systolic blood pressure control, including in combination with an intervention on smoking over follow-up, on the risk of first-ever stroke and dementia using data from 15 years of follow-up in the Rotterdam Study. All interventions that involved reducing SBP were associated with a stroke risk reduction of about 10%, and joint interventions on SBP and smoking status further decreased the risk of stroke in over 15% (e.g. reducing SBP by 20% if above 140mmHg and quit smoking risk ratio: 0.83; 95% CI 0.71 - 0.94). In contrast, we did not observe a change in the risk of dementia. As opposed, all point estimates were above one. These results need to be interpreted in the context of death as a competing event. Given that we have targeted at a total effect, part of the effect in the risk of dementia is mediated by how interventions affect the risk of death.

Like in the previous study, our interest was in the sustained effect of these strategies, over 20 years of follow-up – that is, we were interested in the per-protocol effect. Rather than answering this question with inverse-probability weighting (or IPW), we used the parametric g-formula. The parametric g-formula is a method that allows us to fit regression models to estimate the complete joint distribution of the outcome given the time-varying exposures and time-varying confounding. Under the assumption of no unmeasured confounding and no model-misspecification, we can use high-dimensional data to simulate the risk of an outcome as if everybody would receive a certain intervention. Each simulation represents a different “treatment arm” and randomization is mimicked because every simulation recreates the same pseudopopulation and only the value of systolic blood pressure is modified according to the defined strategy. This method allows us to define as many useful interventions we want, so it can be a powerful tool.

But, as they say, with great power comes great responsibility. The parametric g-formula is very sensitive to model misspecifications. Since all variables are modeled (the intervention, outcome, competing events, and all included confounders), this method pushes the researcher to understand the nature of each variable, how it was collected and coded or treated in the dataset. This introduced some challenges during the analysis and fitting of the g-formula. In this study, systolic blood pressure was measured at every visit in the Ergo-Center, thus, each participant had up to 5 measurements of this variable, and the date of the measurement. We also collected other time-varying covariates that were measured during these visits, such as smoking, body mass index, alcohol intake, cholesterol and hypertension treatment. We also included time-varying covariates that represented an incident diagnosis of diabetes, heart disease, cancer. These variables, as well as the outcome of stroke, dementia and death, were collected from several sources such as integration of the electronical medical records, and we had specific dates for each variable that were unrelated to the dates of the visit process.

Having transformed the data to a person-year (or long) format, and after fitting the parametric g-formula, predictions for several variables including systolic blood pressure were off. Fortunately, the reason of this problem was visually represented when plotting the predicted values of the systolic blood pressure vs. the observed values over time, represented in Figure 2A. In this plot we observed, in red line the observed mean values of SPB at each year, vs. the predicted values of SPB in dotted blue lines. The shape of observed trajectory looks like irregular steps, while the predicted values look linear. These irregular steps represent the years in which values changed for each individual, and those values are specific to the visit process. However, they are not entirely regular because intervals between visits were not symmetric, and intervals between two consecutive visits were not symmetric across individuals. That means that participant X could have their SBP measured with one year of distance, while participant Y may have a gap up to six years between SBP measurements.

To solve this challenge, we had to create variables that indicated the year in which each participant attended each visit. Then, we set the g-formula parameters to simulate the visit process and to simulate each covariate that was measured at the visits only after each visit was first simulated. As opposed, the variables that are independent of the visit process did not require this specification. This setup lead to better predictions, as we observe in Figure 2B, not just for the systolic blood pressure, but for all covariates and outcome predictions. This experience increased my awareness of the additional challenges that remain underexplored in longitudinal studies, and that is not restricted to the use of the parametric g-formula, but in general every time we use longitudinal data that comes from different sources. This highlights how important it is to be familiarized with the data collection process and the data in general. And that the application of the g-formula will be molded to the cohort characteristics and unique features. (add references, or briefly discuss the visit process papers from before, Miguel Hernan visit process, jessica in time between measurements, acknowledge limitations of our data!)

Now that I have outlined the technical challenges to implement this study, I will focus on the broader discussion related to having a well-defined intervention. In this hypothetical trial we did not specify how we would reduce the blood pressure. I am aware that a non-pre-specified blood pressure intervention is not the ideal research question when in fact, the most appropriate well-defined question would explore interventions that specifically target lower-pressure medication (indicated for hypertension) and life-style factors in a population-based cohort with sufficient length of follow-up. However, we consider that our approach gets closer to understanding the effect of lowering blood pressure, by simulating the effect of different strategies targeting systolic blood pressure over time, considering the potential effect of time-varying covariates and the development of comorbidities that affect systolic blood pressure over time. Previously, these considerations have not been specifically addressed in observational studies.

This also means that our estimates are based on the consistency assumption that lowering SBP through any available means (e.g., dietary changes, medication use, other lifestyle changes) would have the same effect on stroke or dementia risk, or otherwise are at best interpretable as estimates for an effect of a weighted average of several SBP-lowering strategies with weights determined by the frequency that the particular strategies occur in our specific population. Furthermore, given how each strategy was defined, interpretation of the effect of these strategies should be limited to the individuals with systolic blood pressure that is above the clinical threshold (regardless of whether they are currently receiving or not any treatment at the moment), those who have a systolic blood pressure (SBP) below the clinical threshold are not affected by any of the specified interventions. Thus, while there are certainly limitations in terms of ambiguity to the interventions studied in the current paper, they represent an improvement (in terms of clarity and for informing decision-making) over etiologic studies that address SBP’s effects with a simplified version of the complexity of real data, and a step toward the types of interventions we may consider in as public health interventions. This exercise is only intended to complement the larger agenda which yet would include studying the effect of different antihypertensives and life-style interventions, as well as the study of the interventions to improve adherence achieve a sustained control of hypertension in different sub-populations.

As I mentioned in the introduction of this dissertation, during my first years of the PhD, the debate in respect to conceptualizing causal questions when the measured exposures are not measurements of an intervention, filled me with insecurities about how to study exposures related to dementia etiology. But Chapter 3 represents a step towards stepping out of the dilemma. I believe we can study the effect of biomarkers or other exposures of interest, even if we don’t have the intervention that targets them, as long as we can be clear about the underlying causal question of interest and transparent both during the interpretation and limitations of the findings. This process might be straightforward if we already know the interventions that would target the exposure of interest: we know the interventions that could lower blood pressure just as the interventions to reduce cholesterol. This thought process becomes more challenging when there is clear (or not yet real) intervention or when we have a measurement in the dataset that is too far from whatever we would have wanted to measure. In these cases, it might be less clear how to grasp the underlying causal question.

The latter scenario brings Chapter 4 to discussion, but before discussing the aim of the study that represents this chapter, I will briefly outline how this line of research gain interest through-out time. In 1999, Yamada et al. published a paper called “Prevalence and Risks of Dementia in the Japanese Population: RERF’s Adult Health Study Hiroshima Subjects”. The population that was studied was women and men aged 60 and older who resided in Hiroshima at the moment a larger study was taken place. In this study they measured the prevalence of dementia and fitted a logistic regression model including socio-demographic variables, radiation exposure and history of several comorbidities including cancer. In this study, they concluded that the prevalence of Alzheimer’s Disease decreased with a history of cancer (OR: 0.3, 95%CI: 0.05 – 0.98). These findings were controversial, and since then more than twenty studies have studied the association between cancer and dementia, most of them retrieving similar findings [@ospina2020, @vanderwillik2018]. These studies propose multiple hypothesis to explain this association, including several biological, behavioral and environmental factors. Thus, this area of research seems to have heavily grown from inductive reasoning, meaning that the analysis and interpretation of data patterns and results came prior to the hypothesis development to answer that particular question. Although there is recent lab-based studies that give more clarity to this interesting association, the evidence from observational studies is too unclear.

However, researchers have categorized the potential sources of bias that could explain this association in three terms: confounding, measurement error and selection bias [@ganguli2015; @driverbiogeront2014; @frain2017, @ospina2020]. Yet, no study has questioned why “cancer diagnosis” or “history of cancer” is an interesting exposure in the first place. That is, we would never randomize participants to having or not cancer. It is somehow clear the research community is interested in understanding this association to unveil the potential mechanisms of action that could lead to a better design of therapy treatments to prevent or stop the progression of dementia [@snyder2016]. But if we continue asking questions in population-based observational studies, how much can we learn from these underlying mechanisms of interest if we keep focusing on “cancer” as the exposure of interest? Likewise, we cannot truly understand the potential sources of bias if we are not clear about what is our truly question of interest.

Recent lab-based studies have discovered molecular pathways that could explain this inverse association. One of these is related to the protein Pin-1. This protein is involved in different processes during the cell cycle, such as in cell proliferation, motility and apoptosis. It works as a molecular timer that activates or inactivates different pathways, like a switch. In cancer, Pin-1 is overstimulated and increases cell proliferation, angiogenesis, migration and invasion, and inhibits apoptosis of tumor cells in several ways. In opposite, Pin-1 is inhibited in Alzheimer’s disease, and previous studies have shown that Pin1 knockout mice developed a syndrome similar to AD characterized by hyper-phosphorylated tau and neurodegeneration[@lin2020; @drivepin2015; @driverbiogeront2014; @liou2003].

For this reason, in Chapter 4, I begin by taking a more deductive reasoning approach to disentangle potential sources of bias that could explain this inverse association. To this matter, I bring the Pin-1 hypothesis to stage from the very beginning and phrase the question of interest as: \_What is the effect of this Pin1-targeting drug on the risk of dementia over time compared to standard treatments?\_. With this question we aimed to explore how we might learn about this effect using real-world data on cancer and dementia. To connect this particular causal question to the observable data we progressively build a causal directed acyclic graph (DAG), outlining the assumptions needed to study the effect. We highlight the challenges that arise which may introduce bias, and describe how these can be prevented (up to certain extent) through different analytic decisions.

Before I discuss the results, I want to acknowledge that considering cancer diagnosis as a proxy for Pin-1 over-expression may sound wrong, or it may make the reader feel uncomfortable. I agree it is not an easy step to take, and up to some extend it forces us to be creative and imaginative. The reason to endorse with this question is the evidence in Pin-1, and how much is used as one of the mechanisms to explain this association in observational studies. Alternatively we could have phrased another mechanism of interest, such as the effect of chemotherapy vs. no treatment, and consider cancer diagnosis as the proxy for chemotherapy. However, this question would have led to a very different design, even if using the same data, and might bring to attention different sources of bias.

After illustrating a causal directed acyclic graph that represents this question, though yet in a simplified fashion, we show how different analytic decisions may result in very different results. For example, when we consider cancer diagnosis as a “time-fixed” measurement, defined as “ever” vs. “never”, and prior to adjusting for confounding and for censoring death, we observe a protective association with a risk ratio (RR) of 0.70 (95%CI: 0.49, 0.93) and a hazard ratio (HR) of 0.52 (95%CI: 0.39, 0.69). Though adjusting for measured confounding only minimally changed the observed association, the association is closer to the null after including censoring weights for death [RR: 0.91 (95%CI: 0.65, 1.19); HR: 0.72 (95%CI: 0.54, 0.98)]. In contrast, when we considered cancer diagnosis as a time-varying proxy for Pin-1, the fully adjusted model results in a RR of: 1.05 (95%CI: 0.79, 1.29) and a HR of 1.09 (95%CI: 0.8X, 1.5X), though confidence intervals cross the null.

As we observed, defining cancer as “ever vs. never” can introduce immortal-time bias. Although we may attempt to prevent this bias by considering instead as time-varying measurement, and adjusting for time-varying confounders, as well as eliminating “death” through censoring, we can only truly prevent it by clearly defining the time-zero. That means, if we could have designed a prospective study to specifically study the effect of Pin-1 expression, what would be the eligibility criteria to recruit participants? Considering that the time between the first biological changes and cancer manifestations can range between five and forty years, would we include participants at a certain age? Would we want to include participants at risk of cognitive impairment such as that a possible intervention on Pin-1 over-expression and free of cancer manifestations? These questions go beyond the specific study presented in Chapter 4, but it is a call to a broader discussion that is not present in the current literature, specifically talking about the cancer - dementia context. However this echoes one of the biggest challenges in the research of Alzheimer’s disease and potential prevention or treatment therapies at large, even in settings where the targets of intervention have been clearly identified and studied[@ackley2021].

In this study we also discussed the challenge of having death as a competing event, and we showed how much results change when we include weights based of time-varying covariates to satisfy the independent censoring assumption. This challenge is very important, because in our setting, 63% of participants with cancer diagnosis over follow-up died prior to having a dementia diagnosis, in contrast, only 15% of participants free of cancer died prior to a cancer diagnosis. These descriptive numbers already put in evidence the problem. Out of all the papers discussed by the latest systematic review in the topic, by Ospina et al[], only x reported the number of participants who died in each arm.

As I will discuss in more depth in the following paragraphs, when death is a competing event, the causal contrast (or estimand) of interests needs to have death as part of it’s definition. To this matter, I chose to estimate the controlled direct effect, which represents a hypothetical scenario were death could have been eliminated through-out the follow-up. The relevance of this question is debatable, since it does not represent a real-world scenario, and although there are novel estimands to answer this question[@stenruds2020] to be discussed in the Future Directions section, I chose this causal contrast for two reasons: (1) it is an estimand that isolates the direct effect of cancer (or Pin-1) in dementia and (2) most other studies treat death as a censoring event (implicitly, and maybe unintendedly aiming to address a CDE) but do not evoke the independent censoring assumption or who they intend to satisfy it. With that

Even worse, some authors have misleadingly refer to censoring as ignoring, such as Frain et al. [@frain]