# 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 while reflecting on the broader implications of the research. 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[@luijken2021;@power2015]. 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[@hernan2008; @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[@hernan2016; @hernan\_robins\_2016]: 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 the pharmaco-epidemiology literature[@ray2003; @lund2015; @johnson2013] and previous methodologic papers specifically on statins[@danaei2012; @power2015; @emilsson2018], it is surprising that only few studies have considered this design to address this question.

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 (ie. at baseline), we would restrict the study to only those participants who have initiated statin at the time of the measurement (assuming this information is available) 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"[@danaei2013;@garcia\_albeniz2016]. 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 allowed us to go from 6373 eligible participants in the study, to 1578655 potential person-trials.

Up to the time this dissertation was written (October, 2021), no new research paper has been published answering this particular question considering the time-varying nature of statins intake in a different population. Even worse, new studies such as the one by Zhou et al. continued to contrasts current-users vs. non-users at baseline, when the original trial had over 1700 new statin users over follow-up[@zhou2021]. 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 [@danaei2013]. So far, few applications using this tool have been published and no open source package (such as in R or Python) is yet unavailable. Not having open source tools means that not all the research community can access to them. However, less computational intensive techniques that preserve the alignment of treatment and eligibility at time zero have proven to show similar results [@garcia\_albeniz\_2017; @emilsson2018], 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.

\*\*\*

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[@williamson2019]. Most trials where originally performed to answer questions around cardiovascular diseases and as such, they had very refined eligibility criteria (such as participants with history of stroke or with risk of cardiovascular disease) and small length of follow-up[@forette1998; @tzourio2003;@lithell2003; @diener2008; @anderson2011; @williamson2019]. 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[@young2020].

Like in the previous study, our interest was in the sustained effect of these strategies, over 15 years of follow-up – that is, we were interested in the per-protocol effect. To answer this question we used the parametric g-formula, 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[@robins1086; @whatif]. 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[@taubman2009; @young2011]. 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 demands a deep understanding of the data generation mechanisms behind each of the variables from the dataset, since this information will lead how each variable is modelled. Else, it can lead to some challenges as the following. 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. Thus, not acknowledging the different sources of data collection when harmonizing the complete dataset can result in modeling misspecification.

At first, 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. In this plot we observed that the observed mean values of systolic blood pressure by follow-up had the shape of irregular steps going upwards, while the superimposed predicted values looked linear. The irregular steps shape 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 a participant could have two measurements of systolic blood pressure with one year of distance, while another participant may have a gap up to six years between measurements.

To solve this challenge, we had to create variables that indicated the year in which each participant attended each visit and simulate the visit process prior to simulating each covariate. As opposed, the variables that are independent of the visit process did not require this specification. This setup lead to better predictions, since they matched the shaped of the observed values, not just for the systolic blood pressure, but for all covariates and outcome predictions. This case highlights 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. It reflects the need to be well familiarized not just with the data but also with the data collection process specific to a cohort study. Several studies have highlighted the relevance of including the observation plans or data collection design to the causal directed graphs in settings with longitudinal discrete data[@hernan2009; @zhang2011], as well as the need for frequent measurements of the time-varying covariates to prevent bias[@young2019]. Yet, there is room to understand the practical implications and provide solutions to overcome settings when longitudinal data is not ideal in real-case studies. Likewise, this is a call to improve the design and planning of longitudinal studies interested in long-term exposures or interventions.

\*\*\*

Now that I have outlined the technical challenges to implement this study, I will focus on the broader discussion related to specifying a well-defined intervention. In the hypothetical trial of Chapter 3 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/or life-style interventions. This also means that our estimates are based on the consistency assumption that lowering systolic blood pressure through any available means (e.g., dietary changes, medication use, other lifestyle changes) would have the same effect on stroke or dementia risk. Otherwise, they are, at best, interpretable as estimates for an effect of a weighted average of several systolic blood pressure - lowering strategies with weights determined by the frequency that the particular strategies occur in our specific population [@waterkills; @hernan2011]. 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 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 systolic blood pressure 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..

The discussion about well-defined interventions brings me back to a point 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. I consider that 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 measurements of the intervention that targets them, as long as we can be clear about the underlying causal question of interest and transparent about the assumptions, interpretations and limitations of the findings. This process might be straightforward if there is enough scientific evidence on the interventions that would target the exposure of interest: we know the interventions that could lower blood pressure. Nonetheless, this thought process becomes more challenging when there is no clear (nor available) intervention or when we have a measurement in the dataset that is too far from whatever we would have wanted to study. In these cases, it might be less clear how to grasp the underlying causal question.

\*\*\*

This 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”[@yamada1999]. The studied population included women and men aged 60 and older who resided in Hiroshima in early 1990’s. 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). Since then, numerous studies have studied the association between cancer and dementia, most of them retrieving similar findings [@ospina2020, @vanderwillik2018] and quoting this study as reference. These studies also 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 an inductive reasoning perspective, meaning that the analysis and interpretation of data patterns and results preceded the hypothesis development to answer that particular question[@banegas2000]. In fact, Yamada’s et al work did not aim to study the relationship between cancer and dementia as a primary question, and is susceptible of selection bias given the highly specific population of atomic bomb survivors recruited in the initial study. Furthermore, the analysis falls into the example of the Table 2 fallacy [@westreich2012]. 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 acknowledged and 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 studying this question 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 several molecular pathways that could explain this inverse association. One of these is related to the protein Pin-1 that is involved in different processes during the cell cycle, such as in cell proliferation 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 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 diagnosis 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. We could have stated a question about another mechanism of interest instead, such as the effect of chemotherapy vs. no treatment, and consider cancer diagnosis as the proxy for chemotherapy. 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, which is the key point of discussion.

In this work we depict a causal directed acyclic graph that represents this question and 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 discussed, 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 and weighting, 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 an intervention on Pin-1, we would ideally align the time of eligibility criteria with the time of measurement of Pin-1 (that would correspond with taking an action on it). For this particular question, considering that the time between the first biological changes and cancer manifestations can range between five and forty years[@], would we want to include participants at risk of cognitive impairment and free of cancer through screening?

The challenges to define the eligibility criteria for preventive treatment of Alzheimer’s disease and related dementias go beyond the study presented in Chapter 4, and are one of the main concerns of the field[@cite]. However, this should not mean that we should obviate the discussion since it is already known, we should raise it every time we study a biomarker, in the context of causal and prediction research. Only by being explicit of the intentions of our research, even if they are out of the scope of the data available to answer the question in mind, we can lead a more transparent agenda of research in this fascinating and challenging field.

\*\*\*

In Chapter 4 I also discussed the challenge of having death as a competing event, and 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 (CDE), which represents a hypothetical scenario were death could have been eliminated through-out the follow-up[@young2020]. The relevance of this question is debatable, since it does not represent a real-world scenario, and although there are novel estimands with alternative interpretations to answer this question[@stenruds2020], 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 how they intend to satisfy it.

Prior studies had raised concerns related to the competing event of death as a potential source of bias[@ospina2020; @hayes\_larson2020]. However, only few studies have elaborated on how to overcome the problem[@hanson2016]. As I mentioned previously, most studies treated death as a censoring event[@roe2010; @driver2012; @nudelmann2014; @freedman2016; @frain2017; @bowles2017; @prinelli2018], but without clear understanding of the underlying causal contrast or question of interest and without outlining the assumptions required to get a valid estimate. While most papers only mention death as part of the time of calculation for length of follow-up, such as “we followed participants until dementia diagnosis, death or last date of follow-up”, this is not enough to understand which estimand is being targeted. Frain et al. give some intuition about the underlying misconception that censoring for death is equivalent to ignoring death in the following quote: “*When the goal is to measure the association between two diseases for the purpose of determining a causal relationship, then it is appropriate to ignore the competing risk, as is routinely done when using Cox models in an elderly cohort.*”[@frain2017]. Although Hanson et al. had elaborated on the problem of considering death as an uninformative event and independent of dementia in 2016[@hanson2016], they concluded that more careful consideration of model specifications is needed, which is true but rather than focusing on the estimator, we need more attention when choosing the estimand first.

I hope that our study and results prove that considering death as a censoring event has direct repercussions on the interpretation of results, and that further action is needed to satisfy the independent censoring assumption. In other words, we need to actively block the confounding paths between death and dementia. In Chapter 4 we showed that results change substantially when we include weights based of time-varying covariates as a way to satisfy the independent censoring assumption. In both setting (time-independent vs. time-varying cancer definition), adding weights for death shift point estimates towards and above the null. The use of inverse probability weighting to block the shared common causes between death and dementia has been described almost ten years ago specifically in the setting of cognitive decline with truncation for death[@weave2012] but few studies implement this method or any other alternatives[@vangeloven2014] in this research field.

This topic is of major importance for the cancer-dementia debate because participants with cancer have a higher risk of death compared to individuals free of cancer. In our study 63% of participants with cancer diagnosis died prior to having a dementia diagnosis, while only 15% of participants free of cancer died prior to a cancer diagnosis. Having descriptive information about the incidence of death across cancer arms can be very insightful, but not a common practice, which brings Chapter 5 to discussion.

\*\*\*

In this chapter I present a systematic review of longitudinal studies focused (implicitly or explicitly) on causal effects, in order to summarize how death during follow-up is handled in the design, analysis, reporting, and interpretation of results. Out of 57 papers that were included, the number or proportion of individuals who died over time was reported in 56% of papers; 18% presented these numbers by exposure level. Only 11% had a clear and complete description of how death was treated in the main analysis, while 47% did not include any description on how death was handled in the methods section. The vast majority (93%) presented estimates of a hazard ratio, mostly under a Cox proportional hazards model though none reported the correct interpretation given the presence of a competing event nor discussed the assumptions related to death as a competing event. Furthermore, 86% interpreted hazard ratios as inferring something about a risk (e.g. “the exposure increased the risk of dementia, HR:X, 95%CI”) and only one study gave an explicit interpretation that matched the target causal parameter of interest.

I would like to emphasize two main concerns that raise from these results. First and foremost, there is an evident limitation or resistance to phrase clear causal questions. To retain research articles with a causal aim we had to outline a several points on the eligibility criteria because most of them phrased their aim as interested in looking at the “association”. A large movement to embrace the “causal” word is already held in epidemiology[@hernan\_cword2018; @hernan2019; @goetghebeur; @olarte2021] but it should spread into dementia research too. This issue gives an idea on what is the starting point when we develop educational resources to teach about different estimand when competing events are present.

Subsequently, and as second point, phrasing questions or estimands including a definition of the competing event of death is rare. Thereafter, almost all studies reported hazard ratios as primary results and were frequently interpreted as implying something about a risk. Ospina et al. give insights on why hazard ratios may be preferred on the following text: “*studies that report cumulative incidence proportions of AD are subject to competing risks bias because the cumulative incidence proportion does not account for death…We considered longitudinal studies that used rate-based estimators such as HRs or IRRs or case-control studies that used incidence density sampling as having no competing risks bias..*”. This misconception can be clarified if we return to the classical literature in competing events that gave notions on different estimands. In the 70's, Tsiatsis and others outlined two types of risks: the "net risk" and the "crude risk"[@tsiatsis1975; @peterson1976]. The net risk was defined as the risk of the main outcome in settings where the competing event could have been eliminated. As opposed, the crude risk represented the risk of the main outcome when the competing event is also present. Both estimands represent a cumulative incidence that takes death into account in different ways, and under different assumptions. Decades later, Young et al. prove how these risks translate into two different causal contrasts: the total effect and the controlled direct effect[@young2020]. This work has specifically focused in proposing both directed acyclic graphs and single intervention world graphs for settings with competing events, and has formalized how, under explicit assumptions, they allow for identification of different estimands.

I believe this work has been fundamental to change the narrative on how competing events are usually taught: heavily based on the estimators and with few emphasis on the questions. While several authors dichotomize recommendations into: use cause-specific hazard models for etiologic questions and Fine-Gray model for prediction modeling[@lau2009; @austin2016], Young et. al propose these estimands as alternatives when the interest is in answer a causal question[@young2020]. This work has also motivated into considering analogous estimands to answer prediction questions[@geloven2020]. To illustrate concepts in a language tailored for an applied audience, in Chapter 6 we present a hypothetical randomized trial on smoking cessation in late-midlife, and emulate such a trial using observational data from the Rotterdam Study. We estimated a total effect of smoking cessation (compared to continued smoking) on 20-year dementia risk of 2.1 (95%CI: -0.1, 4.2) percentage points and a controlled direct effect of smoking cessation on 20-year dementia risk had death been fully prevented of -1.9 (-5.1, 1.4) percentage points. Our study highlights how different causal contrasts can result in different estimates, here going in opposite directions. Perhaps the biggest take-away of our findings, recent methodologic innovations, and our guidelines is a simple one: we cannot begin to describe "bias" due to a competing event, let alone do something about that supposed bias, without stating clearly what question we were seeking an answer for.

In Chapter 6 I developed in depth the concepts related to the controlled direct effect and the total effect while there are other estimands to be considered, such as: the survivor average causal effect (SACE)[@frangakis2002], the separable effects [@stensrud2019] and the total effect on the composite outcome. And yet, the reader may feel unsatisfied with these estimand, since their interpretation might be unrealistic, effects may be unwanted (for example, the total effect on the Pin-1/cancer and dementia risk may be undesired for obvious reasons), or unreasonable assumptions. At least I know I share these thoughts and feelings around competing events, but throughout this dissertation I’ve realized that there is no right or wrong answer. There is no one estimand that is better than the other, it all comes down to which one is more suitable to the question of interest, how strong is the association between the intervention/exposure of interest and death, how frequent is death over follow-up, how much information we have over follow-up to satisfy the required assumptions, etc. In some settings, we may not be able to answer the question of interest, but rather only test the sharp null hypothesis (as in Chapter 4).

To finalize, I would like to briefly give a special mention to the controlled direct effect for two reasons: many researchers may feel uncomfortable with answering a question that refers to a hypothetical scenario where death is prevented and second, as discussed in Chapter 5, many studies are in fact trying to answer this question (implicitly) by censoring for death. Since many researchers have outlined their discomfort around this estimand over the last few decades, I would like to share some of the most poignant arguments I’ve read on the topic. Therry Therneau, author of the R "survival" package, provides his opinion on this estimand in the documentation of the corresponding package as follows: "*in this hypothetical world it is indeed true that many more subjects would progress to X, but it is also not a world that any of us will ever inhabit. This author views the result in much the same light as discussions of survival after the zombie apocalypse*"[@therneau2021]. Furthermore, one of the three principles stated by Andersen et al. for biostatistical and epidemiological applications is: "*stick to this world*”[@2012]. Basile Chaix et al. comment on the article on IPW for settings with truncation for death by Weave at al.[@weuve2012] as follows: “*In replacing dead participants by cloning the living, IPW generates a sample in which participants are not allowed to die. Moreover, IPW attributes particularly high weights to the participants most likely to die, ie, to people with poor health characteristics associated with death in the attrition model. In doing so, IPW not only prevents people from dying but also artificially maintains the lives of people in very poor health—arguably a form of statistical cruelty*”[@chaix2012].

These recommendations may lead the reader to ask, why did this estimand was ever considered in the first place?. To understand the relevance of this estimand it is essential to return to the history of epidemiology and methods developments from the 18th century. During the smallpox epidemic, several attempts to develop preventive treatments were studied, including smallpox inoculation. Given that inoculation could also lead to smallpox and death, and randomized trials were not popular at the time, there was a lot of controversy around the topic[@karn1932]. To this matter, Bernoulli used available data and compared the observed period life expectancy to a counterfactual scenario where inoculation was mandatory to each individual at birth (thus, an scenario where smallpox was eradicated), concluding that early inoculation would result in an increase of years to life expectancy[@karn1932; @colombo2015]. In this setting, eliminating smallpox as a cause of death may sound reasonable (if we stand from the perspective of someone living in the 21st century), even if the intervention was not available at the moment. Although Bernoulli faced serious criticism about the work [@seth2014, @colombo2015], this counterfactual scenario became popular rapidly, as researchers were interested in assessing life expectancy had cancer, pulmonary tuberculosis or heart disease could have been eliminated [@karn1933]. And while criticism to the assumptions tied to this question were discussed since the beginning of this story, Tsiatsis provides a point that should not be dismissed: “relying on this assumption requires deep knowledge on the biological process and expertise knowledge on the topic”[@tsiatsis1975].

That being said, imagining scenarios where death is prevented may be reasonable, depending on the research question of interest, and if we can first conceptualize the intervention that would prevent death over follow-up (and have sufficient data). Although dementia happens in late-life, indicators of how much life-expectancy has improved worldwide over the last century, and that high income countries are reducing mortality from preventable diseases, makes me more optimistic about this approach and far from considering a zombie apocalypse any time soon. And to finalize, this story also resonates with the notion of conceptualizing clear questions even in settings where we don’t have the treatment yet, measured, tested or discovered. As Carol Buck wrote in 1975: “*To search for all the refutable consequences of a hypothesis demands highly imaginative thinking. Imagination is needed to arrive at the hypothesis in the first place, let alone to suggest rigorous tests for it.*” I believe that causal reasoning forces us to be creative and imaginative, otherwise, how can we even grasp counterfactual thinking?

### Directions for future research