* Chapter 2
  + Prevalent user bias and sequence of trials
  + Per protocol effect
  + Say something about the difference between methods development and applications
* Connect this to chapter 3.
* Chapter 3
  + Previous differences between observational and rcts, that means they are not comparable X
  + Aim and results X
  + Again mention that we needed time-varying covariates, we have to be very careful on the modelling specifications for each covariate. X
  + This forced us to conceptualize the data generating mechanism of each variable, and since they all come from different sources we needed to be sure about it. 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. Maybe show a picture of before and after for systolic blood pressure. X
  + Then we move to the results and discuss the competing event of death. We don’t think this is biased, its just the total effect. Which highlights how important it is to also look at death to have a complete picture. TBD
  + But this means that we need to have data on time-varying predictors of death, either if it is to get the total effect (because we need to model in properly) vs. the cde were we need enough information to prevent it from happening. In our case, we did not get accurate predictions of the natural course when estimating the cde and this could be part of the challenge. TBD
  + To note there was one more issue for the smoking variable, in relation to inconsistencies. For example x amount of people said they smoked at visit 1 and at visit 2 they said they never smoked. Inconsistent patterns are (maybe this is not that worthy) TBD
  + Now that we have discussed the technical challenges, I can 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, that is it could be with changing life style habits, medication (any), etc. This could lead to criticism X
  + Say that we don’t always have access to the measurements of the real intervention, but just a proxy of it. But also, it may be that we are interested specifically in the target of intervention, regardless of the way, as it would be conceptualized as a public health policy.
  + But this does not mean we should also put effort into understanding what are the most efficacious interventions, but also how well is adherence and acceptance of these in different sub-populations.
* The problem of measuring a proxy rather than the intervention we would like to measure is the key challenge in Chapter 4.
  + There is an extensive controversy related to the inverse association of cancer and dementia in previous observational studies. Which has lead to ….
  + Clearly, the research community is interested in understanding this association to unveil potential mechanisms of action that could lead to a better design of therapy treatments to prevent or stop progression of dementia.
  + But truly no one ever questions why we use “cancer diagnosis” as the intended variable of interest.
  + In chapter 4. The aim of our study was …..
  + This project in some way was an exercise to trace back and seek the intended intervention, which is very challenging because there are many “factors” that potentially explain the inverse association between these two diseases. But to understand the potential sources of bias, each of these would have a very different graphical representation in a dag.
  + Therefore we had to commit to one target of interest and we chose Pin1 given the state of the literature.
  + By outlining as part of the research question all the steps that go from what we would really want to measure to what we have measured as proxy, is a creative and challenging process to represent this is a dag, but this puts in evidence all the potential sources of bias, clearly, and thus it unveils the ways (into some extent) in which we could prevent them.
  + In this study we proved that using time-varying covariates to satisfy the independent censoring assumption would flip the direction of the association
  + Mention that it is very hard to define time zero in settings of biomarkers, worse when a disease diagnosis is considered the exposure. Sometimes yes, a measurement of a biomarker may lead to an action, but in settings where the biomarker is measured in a different setting, this becomes more challenging.
  + By representing each step that goes from the imaginary hypothesis to the data-generation mechanisns, through causal diagrams, we were able to dissect several sources of bias, some of which we can prevent by design or within the analysis, but some leave room to bigger discussions in respect to how to continue
* Too many ways to introduce my fascination for competing events

\*\*About competing events.\*\* (yet I need to organize my thoughts)

* Competing events is probably the most exciting methodologic problem I faced in my career. Specially because it is often taught as an advanced method in survival analysis and not frequently used or well understood. But, as opposed to the believe that research questions with competing events are the exception, I believe that they happen very frequently. The field of dementia and other aging-related diseases is especially susceptible to competing events, given that participants may die from other conditions before developing dementia.
* In Chapter 5 I reviewed over fifty papers published in top journals. Our aim was to understand what are the current practices and.
* Although about 80% of papers had follow-up longer than 5 years, almost half of the papers did not even mention how many participants had died over follow-up, and only few had described the number of deaths on each exposure group. In some cases, death was not even mention when the lenght of follow-up for each participant was described.
* The idea that when competing events are present, there is more than one "risk" estimand, is very old. In fact, prior to the development of Cox-regression models, in the 70's Tsiatsis had already outline at least two types of risks: the "net risk" and the "crude risk". 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. This lead to many discussions in that time already, with a big emphasis on the problems of relying on the independent censoring assumption when this is not verifiable in the observed data. But they raised one key point that is often forgotten, relying on this assumption requires deep knowledge on the biological process and expertise knowledge on the topic.
* Many researchers question what is the clinical meaning or the usefulness of question where the competing event could have been prevented/eliminatated, that is the controlled direct effect. For example Thernau, author of the R "survival" package, formulated that "\_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\_". Likewise Andersen et al. adviced strongly to "stick to this world" in their paper on xx.
* These cautionary recommendations make sense, up to the point that some may ask, why did we ever consider this as a useful estimand? To understand this, it is key to go back to the origins of the study of competing events, which goes back to the 18th century. Back in those days, there was a smallpox pandemic, and as such, several attempts to develop treatments where under study. One of them was innoculation of smallpox. In different countries, several physicians tested if smallpox innoculation could prevent the disease, though in extremely unethical and violent circumstances (such as testing in slave's children) and it was very controversial, given that individuals may also die from smallpox after innoculation.
* As such, Bernoulli, interested in assessing the risks and benefits of innoculation, proposed different counterfactual scenarios to assess if innoculation would reduce the risk of death by smallpox, and He created life tables to count how many people died from smallpox and from innoculation. Then described a counterfactual scenario where death due to smallpox could have been eliminated. His underlying question was, how many deaths could be prevented with inoculation? One of the problems wit
* This counterfactual scenario makes sense for smallpox and any other infectious disease, currently the WHO is still trying to erradicate infectious disease that are present in low-middle income countries and low income countries, so imagining a world where death from ID are prevented, looks more like looking at the WHO charts from high income countries than considering survival after a zombie apocalipse. In fact, in the current state of the world, we could even ask what would be the risk of dementia if we could have prevented all deaths caused by Covid-19?
* At that time, Bernoulli's theory and calculations caused some negative reactions by others, such as D'Alembert. He believed that mathematics and probability should be applied to real-world scenarios, and critized Bernoulli's assumptions (his work on the \_art of conjecture\_") where \_conjecture\_ means . I am still fascinated with this story, because the most trascending debates in causal inference are related to how unrealistic are the identifiability assumptions - or conjectures - and conclusions about it are most of times subjective and challenge researchers to dare to believe or imagine, where in reality \_data\_ is not necessary equivalent to \_information\_, and I don't think the discussion has moved forward that much, specially considering the time.
* One point stands out, there is no one right answer when it comes to competing, and choosing between estimands (or neither) should be assessed on a case-by-case setting.
* Future directions
* Understanding the effect of social determinants of health
* We need to disentangle how does a social determinant affect mortality and dementia, separately.
* This gives room for decomposition, not just between the exposure and the social determinant, but also conceptualize how to decompose the effect of mortality and death.
* To this matter I hope to understand and learn more about the work of John Jackson,.
* Maybe mention 2 clear questions: such as the question related to gender gap in education and the risk of dementia and
* How much would estrict control in blood pressure would reduce the disparities in the risk of dementia across diferent racial and ethnic populations.
* I would also love to understand in more depth the nuances about life expectancy for causal inference. It already has a lot of assumptions just a descriptive measure, but it is even more challenging to conceptualize life-expency as counterfactuals related to an intervention, specially when conceptualizing cause-specific mortality and life-expectancy and age-specific life-expectancy. Robins and others have already
* Keep generating educational resources for a better understanding of why competing events matter
* I look very much forward to learning about social determinants of health, critical race theory, the critical epidemiology conceptual ideas from Latin America, and hope to service my community with my understanding and experience in research methods in epidemiology.

Conclusion

* Causal inference framework help us prevent many sources of bias
* It also help us frame and answer clearer questions, and get clear results and useful interpretations
* This brings an opportunity to change the narrative related to when we say “we need more research”. Questions should be also about, if we could, what would we really want to measure.
* Large sources of data are necessary, yes, but only if we use it right. Which means that we should be familiarized with the data generating mechanisms in every way. That is, understand what each variable means from an applied expertice, how was this variable measured and collected over time (what are the key features of each data source?).
* From the question to the answer we are making decisions all the time, a clean code and a workflow should show each of these steps.
* Garbage in, garbage out.
* What is truly the question that we would want to answer? Does the variable I have in my dataset reflect what I would have wanted to measure? How can I argument that what I have measured is a good proxy of what I would have really wanted to measure? This sounds more of a deductive process. While on the other hand, starting from the available measured data, looking at associations, and then argumenting all the potential mechanisms that could be related to that variable in the first place and develop a theory seems more of inductive approach to doing science.