# What is An Effective Way to Quantify the Drug Safety Profile When Background Event Rates Are Low with An Application to Diabetes Cardiovascular Studies

Haoda Fu, Ph.D.

Lilly Diabetes Medical Affairs,
Global Patient Outcomes & Real World Evidence

May 19, 2015

# Acknowledgements


- David Manner
- L.J. Wei (Harvard)
- Lu Tian (Stanford)
- Hajime Uno (Harvard)

# Outlines

# Wow! 80,000 patients!!!

AstraZeneca United States

Investors | Media | Careers | Health care professionals | Patients

Other websites

Search Go

Home About us Medicines Responsibility

You are here Media / Press releases / AstraZeneca announces initiation of two additional global studies with BRILINTA (ticagrelor)

AstraZeneca

Media

Press releases

2013

2012

2011

2010

2009

2008

2007

Fast facts

## AstraZeneca announces initiation of two additional global studies with BRILINTA (ticagrelor)

Thursday, 14 November 2013

### SOCRATES compares ticagrelor versus aspirin for the prevention of major vascular events in patients with acute ischemic stroke or transient ischemic attack

### THEMIS compares ticagrelor versus placebo for the long-term prevention of major vascular events in patients with Type 2 diabetes at high cardiovascular risk

AstraZeneca today announced plans to conduct two new clinical studies as part of PARTHENON, AstraZeneca's largest clinical trial program involving over 80,000 patients. The studies are designed to build scientific understanding of BRILINTA® (ticagrelor) tablets in additional high-risk patient populations.

Stay in touch

Follow us on Twitter

Visit our blog

RSS Feed

Social media

Please note

These press releases were issued in the United States (US) and are posted here as reference information for US investors and journalists only. The information contained in each press release was accurate at the time of issuance, and AstraZeneca assumes no responsibility for

http://www.astrazeneca-us.com/media/press-releases/Article/20131114-astrazeneca-announces-initiation-of-two-additional

# T2DM: 16 CVOTs in $> 150,000$ patients

| | n | Intervention | Population | Primary outcome | Study status | Start and estimated end date | ClinicalTrials.gov identifier |
|---|---|---|---|---|---|---|---|
| TECOS | 14 000 | Sitagliptin versus placebo | T2DM; $HbA_{1c}$ 6·5–8·0%; ≥50 years; CVD history | CV death, MI, UA, or stroke | Ongoing, not recruiting | 12/2008–Q3/2014 | NCT00790205 |
| TOSCA IT | 3371 | Pioglitazone versus sulfonylurea | T2DM; $HbA_{1c}$ ≥7·0% and ≤9·0%; metformin monotherapy | Death, MI, stroke or coronary revascularisation | Recruiting | 09/2008–12/2018 | NCT00700856 |
| CANVAS | 4330 | Canagliflozin 100 mg versus canagliflozin 300 mg versus placebo | T2DM; ≥30 years; $HbA_{1c}$ 7·0–10·5%; History of/high risk of CVD | CV death, MI, UA, or stroke | Ongoing, not recruiting | 12/2009–03/2017 | NCT01032629 |
| ELIXA | 6000 | Lixisenatide versus placebo | T2DM; $HbA_{1c}$ 5·5–11·0%; ACS | CV death, MI, UA, or stroke | Ongoing, not recruiting | 06/2010–01/2015 | NCT01147250 |
| EXSCEL | 14 000 | Exenatide once weekly versus placebo | T2DM; $HbA_{1c}$ 6·5–10·0%; CVD in about 60% | CV death, MI, or stroke | Recruiting | 06/2010–12/2017 | NCT01144338 |
| BI 10773 trial | 7000 | Empagliflozin 10 mg versus empagliflozin 25 mg versus placebo | T2DM; ≥18 years; $HbA_{1c}$ 7·0–10·0%; (7·0–9·0% drug naive); high CV risk | CV death, MI, or stroke | Ongoing, not recruiting | 07/2010–03/2018 | NCT01131676 |
| LEADER | 9340 | Liraglutide versus placebo | T2DM; $HbA_{1c}$ ≥7·0%; ≥50 years+CVD; ≥60 years+CV risk factors | CV death, MI, or stroke | Ongoing, not recruiting | 08/2010–10/2015 | NCT01179048 |
| CAROLINA | 6000 | Linagliptin versus glimepiride | T2DM; $HbA_{1c}$ 6·5–8·5%; 40–85 years; CVD/CV risk factors/diabetes end organ damage | CV death, MI, UA, or stroke | Ongoing, not recruiting | 10/2010–09/2018 | NCT01243424 |
| REWIND | 9622 | Dulaglutide versus placebo | T2DM; $HbA_{1c}$ ≤9·5%; 50 years+CVD; 55 years+subclinical CVD; ≥60 years+CV risk factors | CV death, MI, or stroke | Ongoing, not recruiting | 07/2011–04/2019 | NCT01394952 |
| MK-3102 trial | 4000 | MK-3102 versus placebo | T2DM; CVD history | CV death, MI, UA, or stroke | Recruiting | 10/2012–10/2017 | NCT01703208 |
| SUSTAIN6 | 3260 | Semaglutide 0·5 mg versus semaglutide 1·0 mg versus placebo | T2DM; $HbA_{1c}$ ≥7·0%; age ≥50 years+CVD; age ≥60 years+subclinical CVD | CV death, MI, or stroke | Ongoing, not recruiting | 02/2013–01/2016 | NCT01720446 |
| ITCA650 trial | 2000 | ITCA 650 (exenatide in DUROS) versus placebo | T2DM; $HbA_{1c}$ >6·5%; CVD history | CV death, MI, UA, or stroke | Recruiting | 03/2013–07/2018 | NCT01455896 |
| DECLARE-TIMI 58 | 17 150 | Dapagliflozin 10 mg versus placebo | T2DM; ≥40 years; high CV risk | CV death, MI, or stroke | Recruiting | 04/2013–04/2019 | NCT01730534 |
| CARMELINA | 8300 | Linagliptin versus placebo | T2DM; $HbA_{1c}$ 6·5–10%; 18 years; microalbuminuri or macroalbuminuria and previous macrovascular disease; impaired renal function with predefined UACR | CV death, MI, UA, or stroke | Recruiting | 07/2013–01/2018 | NCT01897532 |
| DEVOTE | 7500 | Insulin degludec versus insulin glargine | T2DM; $HbA_{1c}$ ≥7·0%; or HbA1c ≤7·0% and insulin treatment; ≥50 years and CV or renal disease or ≥60 years and CV risk factors | CV death, MI or stroke | Recruiting | 10/2013–11/2018 | NCT01959529 |
| Ertugliflozin trial | 3900 | Ertugliflozin 5 mg versus ertugliflozin 15 mg versus placebo | T2DM; $HbA_{1c}$ 7·0 –10·5%; CVD history | CV death, MI, or stroke | Recruiting | 11/2013–06/2020 | NCT01986881 |

T2DM=type 2 diabetes. CVD=cardiovascular disease. CV=cardiovascular. MI=myocardial infarction. UA=unstable angina. ACS=acute coronary syndrome.

***Table 2:*** **Ongoing cardiovascular outcome trials of glucose-lowering drugs or strategies (in order of starting date)**

(Holman et al., 2014, Page 2009)

# Some thoughts after impressed by numbers ...

- Is it ethical to enroll so many patients in these studies? What if the drug is not safe?
- It takes so long to complete. Do we feel that the public would like to know the results quicker if the drug is not safe?
- How can we control operation bias for such large trials?
- Does the high cost of such trials become a barrier for drug companies to develop new treatments for patients?
- Do we have alternatives ...

# Outline

# Examples of risk ratio (R) and risk difference (D)

Risk ratio versus risk difference: which one is more informative?

- In a two arm, randomized control trial, for treatment arm A, we observed 5 adverse events out of 100 patients, and for treatment B, we observed 4 adverse events out of another 100 patients.

# Examples of risk ratio (R) and risk difference (D)

Risk ratio versus risk difference: which one is more informative?

- In a two arm, randomized control trial, for treatment arm A, we observed 5 adverse events out of 100 patients, and for treatment B, we observed 4 adverse events out of another 100 patients.
- It is easy to see that the event proportions for treatment A is 5/100, and for treatment B is 4/100.

# Examples of risk ratio (R) and risk difference (D)

Risk ratio versus risk difference: which one is more informative?

- In a two arm, randomized control trial, for treatment arm A, we observed 5 adverse events out of 100 patients, and for treatment B, we observed 4 adverse events out of another 100 patients.
- It is easy to see that the event proportions for treatment A is 5/100, and for treatment B is 4/100.
- The risk difference (D) is 5/100-4/100=1/100 (i.e. 1%) with 95% exact confidence interval as (-6.6426%,8.2905%).

# Examples of risk ratio (R) and risk difference (D)

Risk ratio versus risk difference: which one is more informative?

- In a two arm, randomized control trial, for treatment arm A, we observed 5 adverse events out of 100 patients, and for treatment B, we observed 4 adverse events out of another 100 patients.
- It is easy to see that the event proportions for treatment A is 5/100, and for treatment B is 4/100.
- The risk difference (D) is 5/100-4/100=1/100 (i.e. 1%) with 95% exact confidence interval as (-6.6426%,8.2905%).
- The risk ratio (R) is (5/100)/(4/100)=5/4=1.25 with 95% exact confidence interval as (0.3352,4.8164).

# Examples of risk ratio (R) and risk difference (D)

Risk ratio versus risk difference: which one is more informative?

- In a two arm, randomized control trial, for treatment arm A, we observed 5 adverse events out of 100 patients, and for treatment B, we observed 4 adverse events out of another 100 patients.
- It is easy to see that the event proportions for treatment A is 5/100, and for treatment B is 4/100.
- The risk difference (D) is 5/100-4/100=1/100 (i.e. 1%) with 95% exact confidence interval as (-6.6426%,8.2905%).
- The risk ratio (R) is (5/100)/(4/100)=5/4=1.25 with 95% exact confidence interval as (0.3352,4.8164).
- It is worth to notice that the sample size 100 is not very useful when we calculate the risk ratio.

# Now we have a larger study...

- In a two arm, randomized control trial, for treatment arm A, we observed 5 adverse events out of 1000 patients, and for treatment B, we observed 4 adverse events out of another 1000 patients.
- It is easy to see that the event proportions for treatment A is 5/1000, and for treatment B is 4/1000.
- The risk difference (D) is 5/1000-4/1000=1/1000 (i.e. 0.1%) with 95% exact confidence interval as (-0.6984%,0.8505%).
- The risk ratio (R) is (5/1000)/(4/1000)=5/4=1.25 with 95% exact confidence interval as (0.3240,4.9863).
- The confidence interval for risk difference shrinks a lot but the confidence interval for risk ratio is virtually unchanged.

# What did we see here?

| Sample size | Trt A | Trt B | D | R |
|---|---|---|---|---|
| 100 | 5 | 4 | 1/100 | 5/4 |
| 1000 | 5 | 4 | 1/1000 | 5/4 |
| 10000 | 5 | 4 | 1/10000 | 5/4 |

# What did we see here?

| Sample size | Trt A | Trt B | D | R |
|---|---|---|---|---|
| 100 | 5 | 4 | 1/100 | 5/4 |
| 1000 | 5 | 4 | 1/1000 | 5/4 |
| 10000 | 5 | 4 | 1/10000 | 5/4 |

- As sample size increases, the risk difference becomes smaller and smaller, but the risk ratio is unchanged.
- The risk ratio calculation ignore the sample size which makes useful information unused.
- For rare events, due to variability, the risk ratio estimation is not stable.
- When lacking the background rate information, the risk ratio may be misleading. From 50% to 60% is 20% risk increase, and it is the same amount of incrase from 1% is 1.2%.

# D and R: which one contains more information?

## Frequentist point of view

From frequentist point, we would like to estimate a parameter, and a good estimate should have a smaller variance which can provide more accurate estimate to what we would like to know. We show that risk difference can be estimated more accurately.

## Bayesian point of view

From Bayesian perspective, we compare the difference between prior and posterior distributions updated by the observed data. A measurement contained more information will update the prior more. We show that under a reasonable non-informative prior, risk difference contains more information (because it takes sample size into account!)

If you believe in these two points, you can skip some technical details by

⏩ click here!

# Which one contains more information?

## Frequentist

$$\begin{aligned} \hat{D} &\sim N\left\{D, \frac{p_1(1-p_1)+p_2(1-p_2)}{n}\right\}, \\ \hat{R} &\sim N\left\{R, \frac{p_1^2+p_2^2-p_1p_2(p_1+p_2)}{np_2^3}\right\}. \end{aligned}$$

It is easy to see that if $p_1 \approx p_2 = p$, the variance of $\hat{D}$ goes to zero with a rate about $p/n$, but the variance of $\hat{R}$ goes to zero with a rate about $1/(np)$. It is obvious that when $p$ is small, $\hat{D}$ can more quickly converge to the truth than $\hat{R}$. In particular, the ratio of the variance is $V(\hat{D})/V(\hat{R}) = p^2 + o(p^2)$.

# Which one contains more information?

## Bayesian

For a probability density function $f(x)$, the entropy is defined as,

$$\begin{aligned} h(f) &= -\int f(x)\log\{f(x)\}dx \\ &= \log(u) - \int f(x)\log\frac{f(x)}{1/u}dx \end{aligned}$$

For a normal distribution, $h(f) = 1/2\log(2\pi\sigma^2) + 1/2$. Since $V(\hat{D})/V(\hat{R}) = p^2 + o(p^2)$, $\hat{D}$ contains more information than $\hat{R}$ (higher entropy means less information). This also has a Bayesian interpretation that if we have a flat prior about $D$, say $U(-1,1)$, and a flat prior about $R$ as $U(0,u)$, $\log(u) - h(f)$ can be viewed as how much information we gain from a uniform prior (Kullback-Leibler Divergence). For rare event situation, when prior for $R$ is between $(0,200)$, the $\hat{D}$ provides more information.

# Kullback-Leibler divergence is invariant

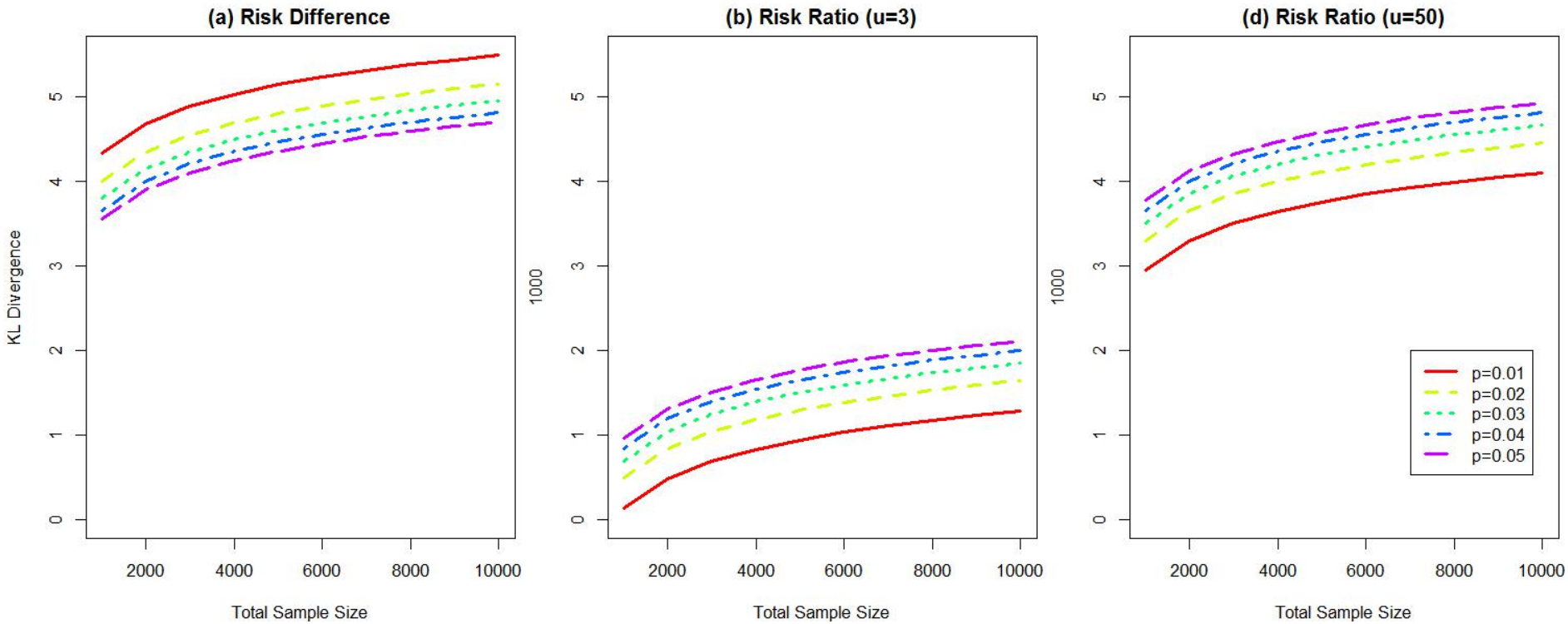
Let $f(x)$ and $g(x)$ be two density functions, and $x = h(y)$ be a transformation function wtih a non-vanishing Jacobian, i.e. $|\partial h(y)/\partial y| \neq 0$. The new transformed density functions are, $f_Y(y) = f\{h(y)\}|\partial h(y)/\partial y|$ and $g_Y(y) = g\{h(y)\}|\partial h(y)/\partial y|$, respectively. So we have,

$$\begin{aligned}
\int f(x) \log \frac{f(x)}{g(x)} dx &= \int f\{h(y)\}|\partial h(y)/\partial y| \cdot \log \frac{f\{h(y)\}}{g\{h(y)\}} dy \\
&= \int f\{h(y)\}|\partial h(y)/\partial y| \cdot \log \frac{f\{h(y)\}|\partial h(y)/\partial y|}{g\{h(y)\}|\partial h(y)/\partial y|} dy \\
&= \int f_Y(y) \log \frac{f_Y(y)}{g_Y(y)} dy.
\end{aligned}$$

In particular, we have,

$$\begin{aligned}
\log(u) - h(f) &= \int f(x) \log \frac{f(x)}{1/u} dx \\
&= \int af(ax+b) \log \frac{af(ax+b)}{a/u} dx.
\end{aligned}$$

# Information changes




Conclusion: under reasonable priors, risk difference measured by a study with 1000 patients could be more informative than a 10,000 patients study measured by risk ratio.

For rare event situations, if we agree that risk difference contains more information, why we continue to provide risk ratios as less informative quantities to patients and public? We encourage people to think the change.

# Outline

# Logic follow

You will see that ...

- Using hazard ratio to measure rare events has a similar problem as using risk ratio.

# Logic follow

You will see that ...

- Using hazard ratio to measure rare events has a similar problem as using risk ratio.
- Actually, we have more concerns on using proportional hazard ratio model.

# Logic follow

You will see that ...

- Using hazard ratio to measure rare events has a similar problem as using risk ratio.
- Actually, we have more concerns on using proportional hazard ratio model.
- You show me the problems, but what is your solution?

# Logic follow

You will see that ...

- Using hazard ratio to measure rare events has a similar problem as using risk ratio.
- Actually, we have more concerns on using proportional hazard ratio model.
- You show me the problems, but what is your solution?
- Restricted mean event time enjoys similar benefit as ratio difference, and it has more merits!

# Logic follow

You will see that ...

- Using hazard ratio to measure rare events has a similar problem as using risk ratio.
- Actually, we have more concerns on using proportional hazard ratio model.
- You show me the problems, but what is your solution?
- Restricted mean event time enjoys similar benefit as ratio difference, and it has more merits!
- How does it apply to our diabetes CV outcome trials

# Hazard Ratio: Problems and Opportunities

- Traditionally, we use hazard ratio as a measurement of between treatment difference for event driven studies.

# Hazard Ratio: Problems and Opportunities

- Traditionally, we use hazard ratio as a measurement of between treatment difference for event driven studies.
- Hazard ratio estimator depends primarily on the observed number of events, NOT the sample size or duration of the study! (Kalbfleisch and Prentice, 2002).

# Hazard Ratio: Problems and Opportunities

- Traditionally, we use hazard ratio as a measurement of between treatment difference for event driven studies.
- Hazard ratio estimator depends primarily on the observed number of events, NOT the sample size or duration of the study! (Kalbfleisch and Prentice, 2002).
- In other words, we need certain number of events to complete the study.

# Hazard Ratio: Problems and Opportunities

- Traditionally, we use hazard ratio as a measurement of between treatment difference for event driven studies.
- Hazard ratio estimator depends primarily on the observed number of events, NOT the sample size or duration of the study! (Kalbfleisch and Prentice, 2002).
- In other words, we need certain number of events to complete the study.
- However, when the drug is very safe, the study could take for a long time without having enough number of events.

# Hazard Ratio: Problems and Opportunities

- Traditionally, we use hazard ratio as a measurement of between treatment difference for event driven studies.
- Hazard ratio estimator depends primarily on the observed number of events, NOT the sample size or duration of the study! (Kalbfleisch and Prentice, 2002).
- In other words, we need certain number of events to complete the study.
- However, when the drug is very safe, the study could take for a long time without having enough number of events.
- Intuitively, smaller number of events with a long follow up period means the drug works well.

# Hazard Ratio: Problems and Opportunities

- Traditionally, we use hazard ratio as a measurement of between treatment difference for event driven studies.
- Hazard ratio estimator depends primarily on the observed number of events, NOT the sample size or duration of the study! (Kalbfleisch and Prentice, 2002).
- In other words, we need certain number of events to complete the study.
- However, when the drug is very safe, the study could take for a long time without having enough number of events.
- Intuitively, smaller number of events with a long follow up period means the drug works well.
- Why hazard ratio can't capture this?! It is counter intuitive!

# Hazard Ratio: Problems and Opportunities

- Traditionally, we use hazard ratio as a measurement of between treatment difference for event driven studies.
- Hazard ratio estimator depends primarily on the observed number of events, NOT the sample size or duration of the study! (Kalbfleisch and Prentice, 2002).
- In other words, we need certain number of events to complete the study.
- However, when the drug is very safe, the study could take for a long time without having enough number of events.
- Intuitively, smaller number of events with a long follow up period means the drug works well.
- Why hazard ratio can't capture this?! It is counter intuitive!
- The simple reason is that the hazard ratio can't take the sample size and duration of exposure into account!

# Think carefully... Here is an example

- In a very extreme case ...(I know this is not real but I want to show you the point).

# Think carefully... Here is an example

- In a very extreme case ...(I know this is not real but I want to show you the point).
- A 2 year study to compare treatment A vs treatment B.

# Think carefully... Here is an example

- In a very extreme case ...(I know this is not real but I want to show you the point).
- A 2 year study to compare treatment A vs treatment B.
- We have 1000 patients per arm. Both arms do not have any events.

# Think carefully... Here is an example

- In a very extreme case ...(I know this is not real but I want to show you the point).
- A 2 year study to compare treatment A vs treatment B.
- We have 1000 patients per arm. Both arms do not have any events.
- Can we say that treatment A is better than treatment B? No! We need events to see a difference.

# Think carefully... Here is an example

- In a very extreme case ...(I know this is not real but I want to show you the point).
- A 2 year study to compare treatment A vs treatment B.
- We have 1000 patients per arm. Both arms do not have any events.
- Can we say that treatment A is better than treatment B? No! We need events to see a difference.
- Can we say that treatment A is as good as treatment B up to 2 years? Probably Yes! Duration and information about no events are taken into consideration.

# Think carefully... Here is an example

- In a very extreme case ...(I know this is not real but I want to show you the point).
- A 2 year study to compare treatment A vs treatment B.
- We have 1000 patients per arm. Both arms do not have any events.
- Can we say that treatment A is better than treatment B? No! We need events to see a difference.
- Can we say that treatment A is as good as treatment B up to 2 years? Probably Yes! Duration and information about no events are taken into consideration.
- No events situation can be translated to rare event cases in real trials.

# Think carefully... Here is an example

- In a very extreme case ...(I know this is not real but I want to show you the point).
- A 2 year study to compare treatment A vs treatment B.
- We have 1000 patients per arm. Both arms do not have any events.
- Can we say that treatment A is better than treatment B? No! We need events to see a difference.
- Can we say that treatment A is as good as treatment B up to 2 years? Probably Yes! Duration and information about no events are taken into consideration.
- No events situation can be translated to rare event cases in real trials.
- So, non-inferiority trial and superiority study might be different!

# HR may not always be the best choice

- A hazard ratio of 1.3 is difficult to interpret clinically due to lack of a reference absolute hazard concept. One cannot say that the worst case is 30% "RISK" increasing since the hazard is not a risk (probability) measure! (Klein and Moeschberger, 2003)

# HR may not always be the best choice

- A hazard ratio of 1.3 is difficult to interpret clinically due to lack of a reference absolute hazard concept. One cannot say that the worst case is 30% "RISK" increasing since the hazard is not a risk (probability) measure! (Klein and Moeschberger, 2003)
- The hazard ratio estimate is very unstable when the number of events is small so the confidence interval is large implying that there is lack of information to make a decision. However, when the two groups are identical and the study has a large number of patients and a long follow-up time, such information is misleading (Uno et al., 2014).

# HR may not always be the best choice

- A hazard ratio of 1.3 is difficult to interpret clinically due to lack of a reference absolute hazard concept. One cannot say that the worst case is 30% "RISK" increasing since the hazard is not a risk (probability) measure! (Klein and Moeschberger, 2003)
- The hazard ratio estimate is very unstable when the number of events is small so the confidence interval is large implying that there is lack of information to make a decision. However, when the two groups are identical and the study has a large number of patients and a long follow-up time, such information is misleading (Uno et al., 2014).
- The proportional hazard assumption is a strong assumption (the hazard ratio is constant over time). When the proportion hazard(PH) assumption is violated...

# When the PH assumption is violated...

- It is difficult to interpret which parameter we try to estimate; it is not an average of hazard ratio over time (Kalbfleisch and Prentice, 1981).
- The parameter estimated depends on censoring distributions !!
- Inference based on hazard ratio estimate (including the logrank test) are not efficient to detect the group difference.(Struthers and Kalbfleisch, 1986; Lin and Wei, 1989).

# Possible alternatives

- Median failure time (may not be estimable).
- t-year survival rate (not an overall measure).
- Number needed to treat (NNT) or number needed to harm (NNH) (not an overall measure and unstable for rare events).
- Restricted Mean Event Time (RMET).
- There is a long history of using RMET for time to event analysis. (Irwin, 1949; Karrison, 1987; Zucker, 1998; Murray and Tsiatis, 1999; Chen and Tsiatis, 2001; Andersen et al., 2004; Zhang and Schaubel, 2011; Royston and Parmar, 2011; Zhao et al., 2012; Tian et al., 2013, 2014).

.

# Outline

# In the rest of this section

You will see that ...

- An example of using RMET. It illustrates how to use RMET and how to interpret the results.

# In the rest of this section

You will see that ...

- An example of using RMET. It illustrates how to use RMET and how to interpret the results.
- Estimation methods and inference methods.

# In the rest of this section

You will see that ...

- An example of using RMET. It illustrates how to use RMET and how to interpret the results.
- Estimation methods and inference methods.
- Simulation to demonstrate the estimation is accurate, and type I errors are well controlled.

# In the rest of this section

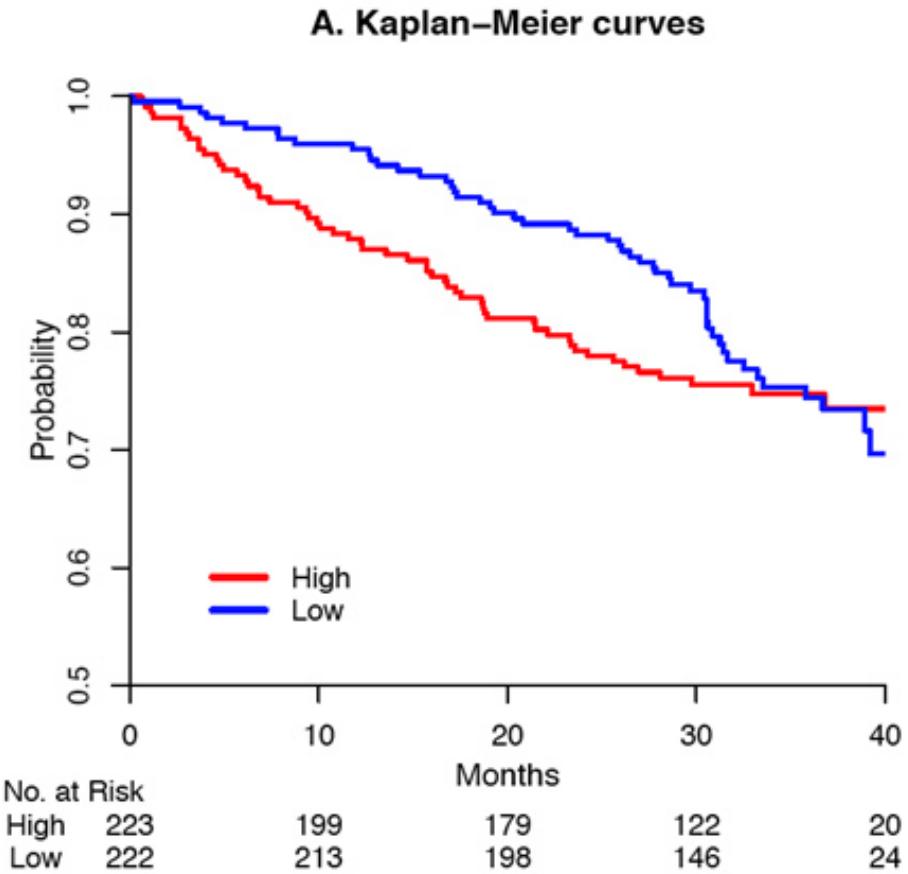
You will see that ...

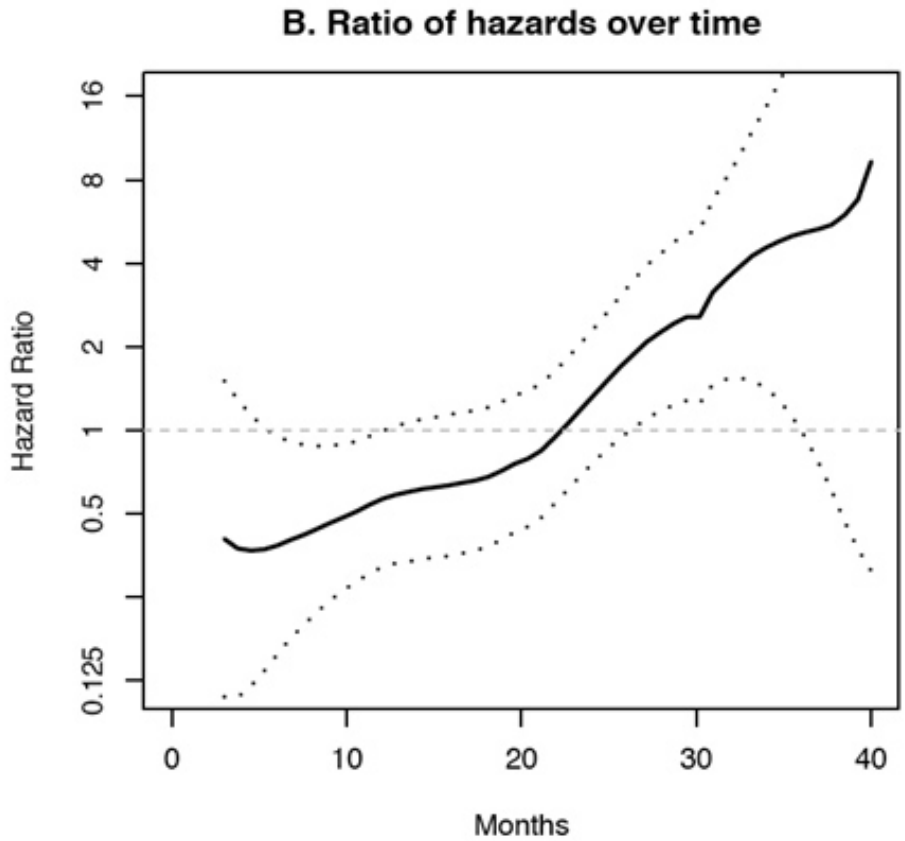
- An example of using RMET. It illustrates how to use RMET and how to interpret the results.
- Estimation methods and inference methods.
- Simulation to demonstrate the estimation is accurate, and type I errors are well controlled.
- We apply it to diabetes CV outcome trials, and how it will impact the sample size.

# An example of RMET

- E4A03 trial to compare low- and high-dose dexamethasone for patients with newly diagnosed multiple myeloma.
- One of the endpoints is overall survival $n = 445$.
- The trial stopped early at the second interim analysis; the low dose was superior.

# HR: 0.87 (0.60 to 1.27)



# Summary of HR analysis

- The proportional hazards assumption is not valid.
- The PH estimator is estimating a quantity which cannot be interpreted and, worse, depends on the study-specific censoring distributions.
- The logrank test is not powerful.
- In conventional analysis, we have Log-rank test: $p = 0.47$ and hazard ratio: HR=0.87 (0.60, 1.27).

# RMET and RMTL

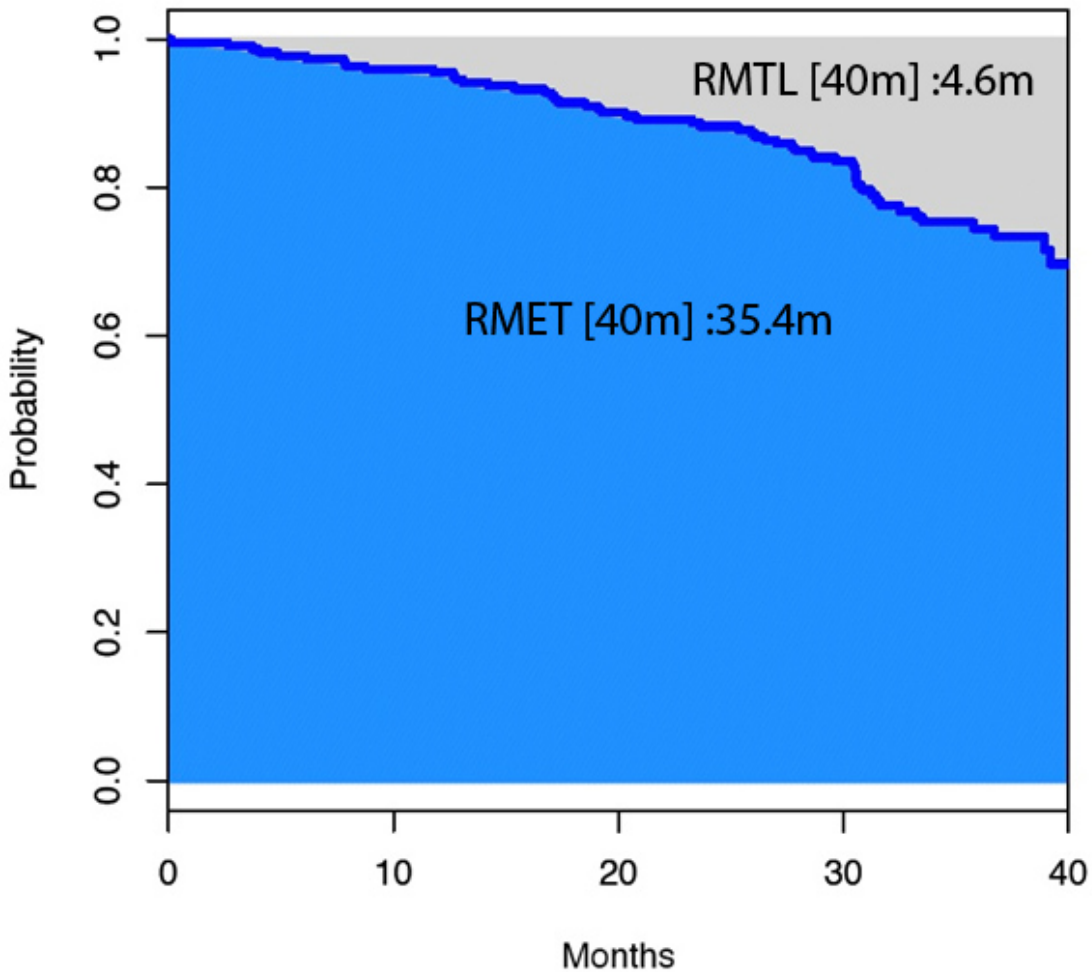



Figure: *Kaplan-Meier curve for low dose arm.* Restricted mean event time (RMET) and restricted mean time lost (RMTL)

# Summary of RMET analysis

- Restricted mean (up to 40 months).
- 35.4 months vs. 33.3 months.
- Difference = 2.1 (0.1, 4.2) months; $p = 0.04$.
- Ratio of RMET = 35.4/33.3 = 1.06 (1.00, 1.13).
- Ratio of time lost = 6.7/4.6 = 1.46 (1.02, 2.13).

# Question to think?

Can we use this approach for diabetes cardiovascular outcome studies? For a 3 year CVOTs, the RMET for a patient is 1062.8 days. If we run such a trial, what is the acceptable non-inferiority margin?

- 106 days $\approx 1062.8 * 10\%$.
- 53 days $\approx 1062.8 * 5\%$.
- 26 days $\approx 1062.8 * 2.5\%$.

We are going to show you that we have good methods to accurately estimate RMET, and they can control type I errors very well. If you are already convinced and would like to see the diabetes CV trial simulation results, you can ⏩ click here!

# Notations

- $T_i$: event time.
- $C_i^d$: dropout censoring time. $C_i^a$: administrative censoring time.
- $I(\cdot)$: indicator function.
- $C_i = C_i^d \wedge C_i^a$. $Y_i = T_i \wedge C_i$. $\Delta_i = I(T_i \leq C_i)$.
- $X_i$: a covariate vector.
- $(Y_i, \Delta_i, X_i)$: observed data.
- $\tau$: a time point where $P(Y_i > \tau) > 0$.
- $T_i^\tau = T_i \wedge \tau$: restricted event time.
- $Y_i^\tau = Y_i \wedge \tau$. $\Delta_i^\tau = I(T_i \wedge \tau \leq C_i)$.
- $(Y_i^\tau, \Delta_i^\tau, X_i)$: derived data based on $\tau$.

## RMET definition

$$\mu^{\tau}(X) = E(T_i^{\tau}|X) = \int_0^{\tau} S(t|X)dt, \quad (1)$$

where $S(t)$ is the survival function.

## Estimation

$$\tilde{\mu}^{\tau} = \int_0^{\tau} \hat{S}(t)dt,$$

where $\hat{S}$ is a Kaplan-Meier (KM) estimator for the survival function of $T$ based on $\{(Y_i, \Delta_i), i = 1, \ldots, n\}$.

# Nonparametric inference

Based on the martingale approach from Andersen (1993), we have,

**Variance estimator of $\tilde{\mu}^\tau$ (area under the KM curve)**

Let $t_1 < t_2 < \ldots < t_D$ are the unique events time, and $d_i$ is the number of events at $t_i$. We have,

$$\hat{V}(\tilde{\mu}^\tau) = \sum_{i=1}^{D} \left\{ \int_{t_i}^{\tau} \hat{S}(t)dt \right\}^2 \frac{d_i}{R(t_i)\{R(t_i) - d_i\}}, \tag{2}$$

where $R(t) = \sum_{i=1}^{n} I(Y_i \geq t)$.

Therefore, based on (Klein and Moeschberger, 2003), the confidence interval can be constructed as,

$$\tilde{\mu}^\tau \pm Z_{1-\alpha/2}\sqrt{\hat{V}(\tilde{\mu}^\tau)}.$$

# A simple, neat and elegant solution

## Logic review

$$\mu^{\tau}(X) \stackrel{?}{=} \beta^{T} X.$$

$$\Downarrow$$

$$\eta\{\mu^{\tau}(X)\} = \beta^{T} X.$$

$$\Downarrow$$

Estimating equation

$$\Downarrow$$

IPW estimate.

Now, we start from IPW estimate, then show the link function with estimating equation in the next a few slides ...

# Alternative estimate of RMET

## Estimation

$$\hat{\mu}^{\tau} = n^{-1} \sum_{i=1}^{n} \frac{\Delta_i^{\tau}}{\hat{G}(Y_i^{\tau})} Y_i^{\tau}, \tag{3}$$

where $\hat{G}(\cdot)$ is the KM estimator of the censoring time $C$ based on $\{(Y_i, 1 - \Delta_i), i = 1, \ldots, n\}$.

# Alternative estimate of RMET

## Estimation

$$\hat{\mu}^{\tau} = n^{-1} \sum_{i=1}^{n} \frac{\Delta_i^{\tau}}{\hat{G}(Y_i^{\tau})} Y_i^{\tau}, \tag{3}$$

where $\hat{G}(\cdot)$ is the KM estimator of the censoring time $C$ based on $\{(Y_i, 1 - \Delta_i), i = 1, \ldots, n\}$.

## Key message

$\hat{\mu}$ and $\tilde{\mu}$ are asymptotically equivalent at the $n^{-1/2}$ rate.

# Asymptotically equivalence between $\tilde{\mu}$ and $\hat{\mu}$

Based on the results from Satten and Datta (2001), we have,

$$\begin{aligned}
\hat{S}(t) &= 1 - \int_0^t \hat{S}(u^-)\frac{dN(u)}{R(u)} = 1 - \frac{1}{n}\int_0^t \frac{dN(u)}{\hat{G}(u^-)} + O_p(n^{-1/2}) \\
&= 1 - \frac{1}{n}\sum_{i=1}^n \frac{I(Y_i \le t)\Delta_i^\tau}{\hat{G}(Y_i)} + O_p(n^{-1/2}) \\
&= \frac{1}{n}\sum_{i=1}^n \frac{I(Y_i > t)\Delta_i^\tau}{\hat{G}(Y_i)} + O_p(n^{-1/2})
\end{aligned}$$

where $N(t) = \sum_{i=1}^n I(Y_i \le t)\Delta_i$, $R(t) = \sum_{i=1}^n I(Y_i \ge t)$, and $\hat{\Lambda}(\cdot)$ is the Nelsen-Aalen estimator of the cumulative hazard function of $T$. Therefore,

$$\hat{\mu}^\tau - \tilde{\mu}^\tau = \hat{\mu}^\tau - \frac{1}{n}\sum_{i=1}^n \int_0^\tau \frac{I(Y_i > t)\Delta_i^\tau}{\hat{G}(Y_i)} dt + O_p(n^{-1/2}) = O_p(n^{-1/2}). \quad \square$$

# Link function

## Modified logic link

$$\begin{aligned} \eta\{\mu^\tau(X)\} &= \log \frac{\mu^\tau(X)}{\tau - \mu^\tau(X)} \\ &= \beta^T X. \end{aligned}$$

# Link function

## Modified logic link

$$\begin{aligned} \eta\{\mu^\tau(X)\} &= \log \frac{\mu^\tau(X)}{\tau - \mu^\tau(X)} \\ &= \beta^T X. \end{aligned}$$

Here, the regression coefficient of the treatment indicator is,

$$\log \left[ \frac{\mu^\tau(X^t) \cdot \{\tau - \mu^\tau(X^c)\}}{\mu^\tau(X^c) \cdot \{\tau - \mu^\tau(X^t)\}} \right],$$

which is an odds ratio to summarize the treatment difference.

# Regression model

## Estimating equation

$$S_n(\beta) = \sum_{i=1}^{n} \frac{\Delta_i^\tau}{\hat{G}(Y_i^\tau)} X_i \{Y_i^\tau - \eta^{-1}(\beta^T X_i)\}. \quad (4)$$

Let $\beta_0$ be the true parameter and we estimate $\beta_0$ by solving the estimating equation $S_n(\beta) = 0$. Under mild conditions, $n^{-1} S_n(\beta_0) \xrightarrow{\mathcal{P}} 0$. So the estimate $\hat{\beta}$ is consistent.

# Inference procedure

1. Calculate $\hat{B}$: Let $\hat{\xi}_i = \frac{\Delta_i^\tau}{\hat{G}(Y_i^\tau)} X_i \{ Y_i^\tau - \eta^{-1}(\hat{\beta}^T X_i) \}$, and we have $\hat{B} = n^{-1} \sum_{i=1}^n \hat{\xi}_i \hat{\xi}_i^T$.
2. Generate $\omega_j$: $\omega_j \sim N(0, \hat{B}^{-1}), \quad i = 1, \ldots, M.$
3. Evaluate $\psi_j$. $\psi_j = n^{-1/2} S_n(\hat{\beta} + n^{-1/2}\omega_j)$.
4. Estimate $\Sigma$: $\hat{\Sigma}^{-1} = M^{-1} \sum_{j=1}^M \psi_j \psi_j^T$.

# Inference procedure

1. Calculate $\hat{B}$: Let $\hat{\xi}_i = \frac{\Delta_i^\tau}{\hat{G}(Y_i^\tau)} X_i \{ Y_i^\tau - \eta^{-1}(\hat{\beta}^T X_i) \}$, and we have $\hat{B} = n^{-1} \sum_{i=1}^n \hat{\xi}_i \hat{\xi}_i^T$.
2. Generate $\omega_j$: $\omega_j \sim N(0, \hat{B}^{-1}), \quad i = 1, \ldots, M.$
3. Evaluate $\psi_j$. $\psi_j = n^{-1/2} S_n(\hat{\beta} + n^{-1/2}\omega_j)$.
4. Estimate $\Sigma$: $\hat{\Sigma}^{-1} = M^{-1} \sum_{j=1}^M \psi_j \psi_j^T$.

**A magic!**

$$n^{1/2}(\hat{\beta} - \beta_0) \xrightarrow{\mathcal{D}} N\{0, \Sigma\}$$

Because, we have the following expansion,
$n^{-1/2} S_n(\beta) = A n^{1/2}(\beta - \beta_0) + n^{-1/2} \sum_{i=1}^n \xi_i + o_p(1 + n^{1/2}||\beta - \beta_0||)$.

# Simulation studies

## Simulation model

$$\eta\{\mu^{\tau}(X)\} \quad = \quad a_0 + a_1 \mathsf{Trt} + a_2 U,$$

where Trt is treatment with value equal to 0 or 1, and $U$ is a uniform distribution with support $(0, 1)$.

Simulation studies include,

- Estimation of the parameters.
- The confidence interval coverage.
- Power and type I error control.
- Sample size calculation for Diabetes CVOTs.

# Simulation for parameter estimation

- True value $a_0 = -2.5$, $a_1 = 1$, and $a_2 = 1$.
- Constant hazard for events, uniform censoring time. Can you figure out how we simulate event times? :)
- $\tau = 24$.
- 500 patients per arm. 1000 simulation runs.
- Estimation results in mean(sd): $-2.499(0.107)$, $0.999(0.088)$, and $0.999(0.173)$.
- 95% confidence interval coverage: (0.949,0.955,0.952).

Conclusion: the estimation is robust and accurate in these simulation settings.

# Parameter estimation

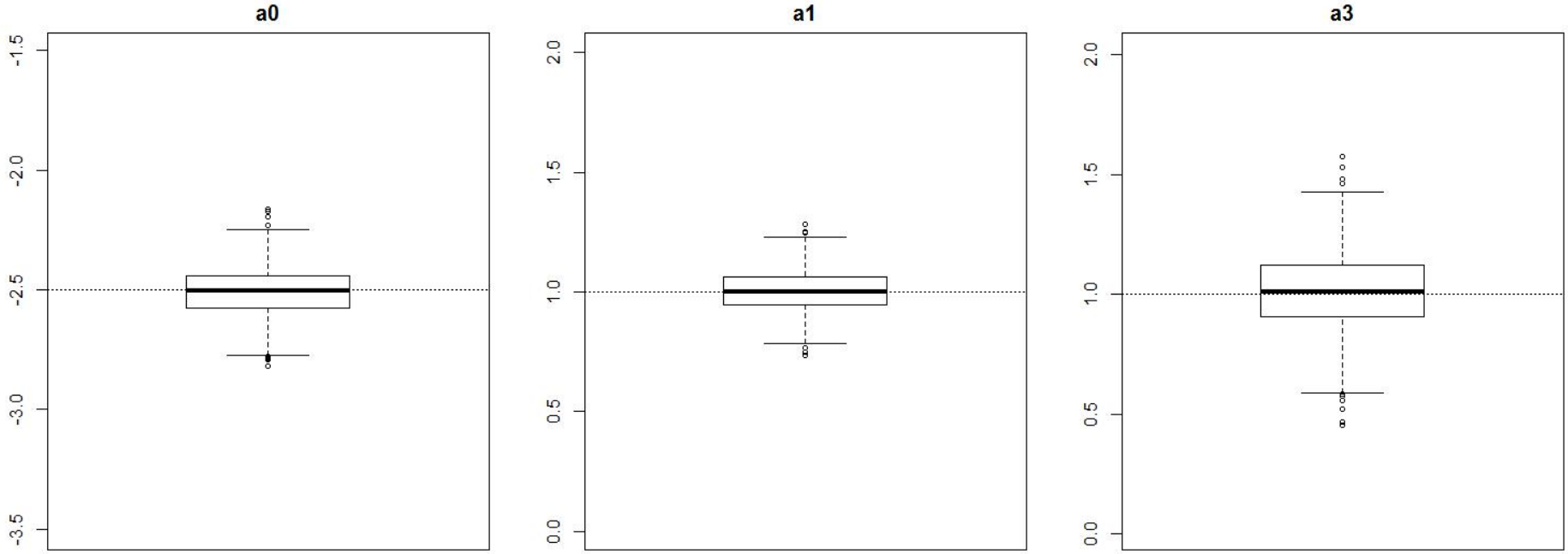
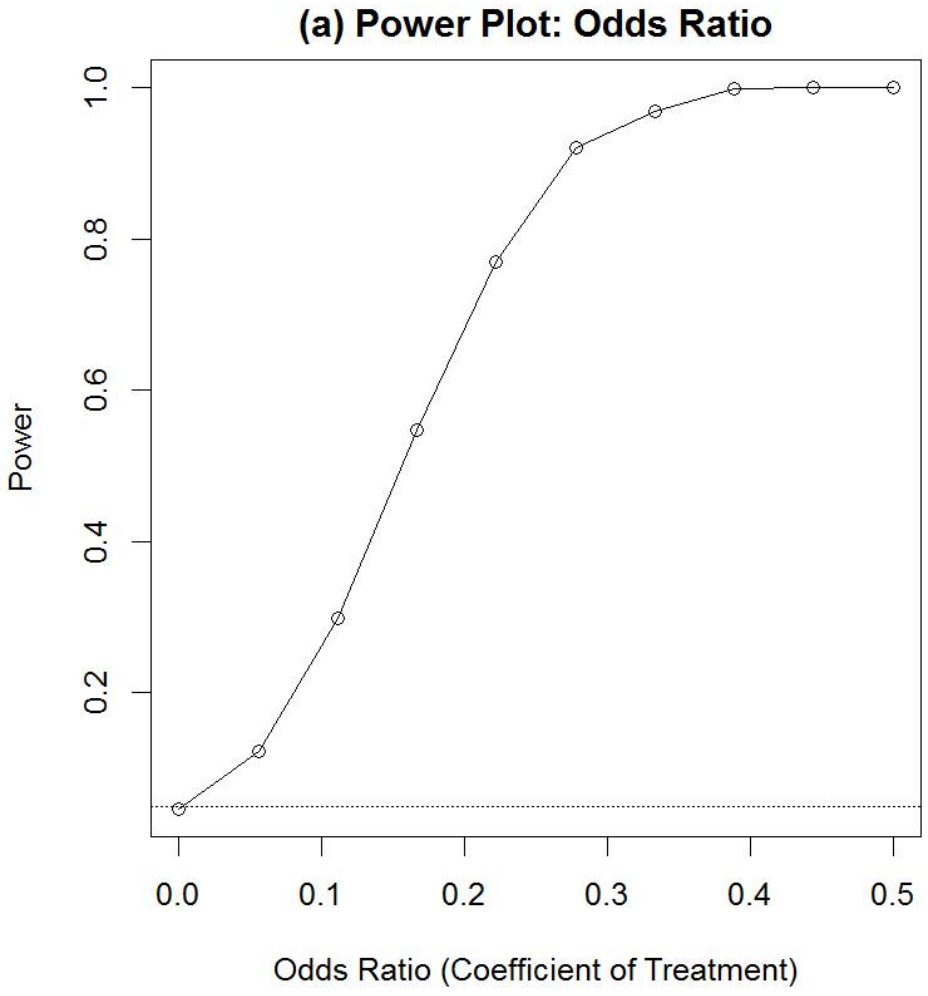



Figure: Boxplots of the estimates



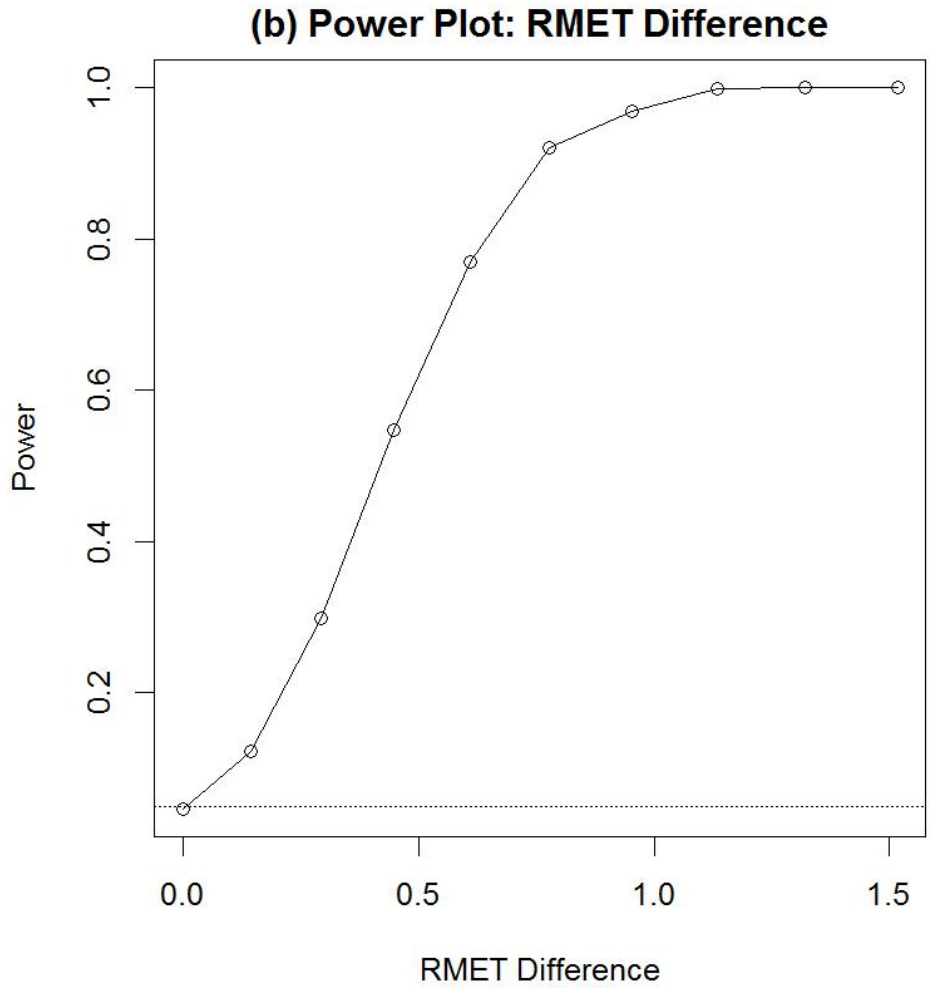
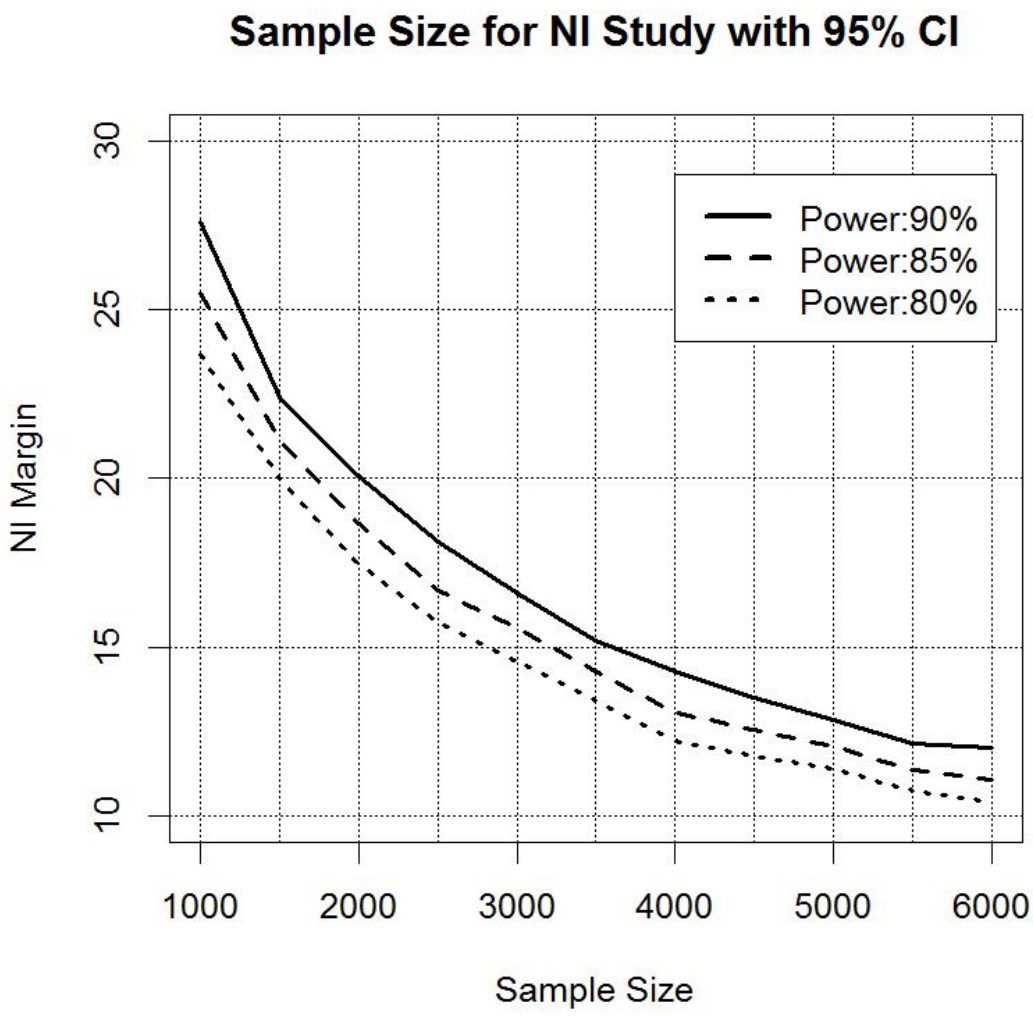




Figure: Type I error control and power plots

# Diabetes CVOTs settings

- Annual event rate is 2%.
- Annual dropout rate is 1%.
- Study period for each patient is 36 months.
- RMET is for 36 months.
- Non-inferiority study assumes both arm has the same event rate.
- Non-inferiority margin is based on 95% confidence interval.
- The event rate and dropout rate assumption are translated to 1062.8 days RMET for this patient population.

# CVOTs sample size




Example: If we choose the non-inferiority margin as 20 days, 2000 patients will provide 90% power.

# Considerations for CVOTs

- We must pre-define the time period for which the RMET will be calculated.
- How we can choose the patients from standard care patients? this would prevent someone choosing a very low CV risk population- prevent "gaming the design".
- We need to pre-define non-inferiority margin in days. So that we have to gain clinical consensus on what is a meaningful difference in days.

Notes: by the assumption, 3 year event rate is 5.75%. With 2000 patients, we expect to observe 114.9 events with standard deviation equal to 10.4. A 95% confidence interval for the observed 3 year event rate is [4.72%,6.76%].

# Conclusion:

- There are 3 major limitations of using hazard ratio and proportion hazard models
  - It can't well take duration of exposure and sample size into consideration.
  - Superiority studies and non-inferiority studies might be different.
  - There are many issues when proportional hazard assumption is not valid.
- RMET is a good alternative.
- Through the estimation and inference methods for RMET is robust. In particular, our proposed method can well control the type I error rate.
- When it is appropriately used, it has a great potential to significantly reduce the sample size for diabetes CVOTs.
- We need to work together to change the paradigm including influencing regulatory agents.
- There are strong support from external thought leaders.
- We believe that there is great opportunities to shift the paradigm which is eventually good for the patients.

I believe that
*every complex problem has a simple, neat and elegant solution ...*

# References

Andersen, P. K. (1993), *Statistical models based on counting processes*, Springer series in statistics, New York: Springer-Verlag.

Andersen, P. K., Hansen, M. G., and Klein, J. P. (2004), "Regression analysis of restricted mean survival time based on pseudo-observations," *Lifetime data analysis*, 10, 335–350.

Chen, P.-Y. and Tsiatis, A. A. (2001), "Causal inference on the difference of the restricted mean lifetime between two groups," *Biometrics*, 57, 1030–1038.

Holman, R. R., Sourij, H., and Califf, R. M. (2014), "Cardiovascular outcome trials of glucose-lowering drugs or strategies in type 2 diabetes," *The Lancet*, 383, 2008–2017.

Irwin, J. (1949), "The standard error of an estimate of expectation of life, with special reference to expectation of tumourless life in experiments with mice," *J Hyg*, 47, 188–9.

Kalbfleisch, J. D. and Prentice, R. L. (1981), "Estimation of the average hazard ratio," *Biometrika*, 68, 105–112.

— (2002), *The statistical analysis of failure time data*, Wiley series in probability and statistics, Hoboken, N.J.: J. Wiley, 2nd ed.

Karrison, T. (1987), "Restricted mean life with adjustment for covariates," *Journal of the American Statistical Association*, 82, 1169–1176.

Klein, J. P. and Moeschberger, M. L. (2003), *Survival analysis : techniques for censored and truncated data*, Statistics for biology and health, New York: Springer, 2nd ed.

Lin, D. Y. and Wei, L.-J. (1989), "The robust inference for the Cox proportional hazards model," *Journal of the American Statistical Association*, 84, 1074–1078.

Murray, S. and Tsiatis, A. A. (1999), "Sequential Methods for Comparing Years of Life Saved in the Two-Sample Censored Data Problem," *Biometrics*, 55, 1085–1092.

Royston, P. and Parmar, M. K. (2011), "The use of restricted mean survival time to estimate the treatment effect in randomized clinical trials when the proportional hazards assumption is in doubt," *Statistics in medicine*, 30, 2409–2421.

Satten, G. A. and Datta, S. (2001), "The Kaplan–Meier estimator as an inverse-probability-of-censoring weighted average," *The American Statistician*, 55, 207–210.

Struthers, C. A. and Kalbfleisch, J. D. (1986), "Misspecified proportional hazard models," *Biometrika*, 73, 363–369.

Tian, L., Zhao, L., and Wei, L. (2013), "On the restricted mean event time in survival analysis," *Harvard University Biostatistics Working Paper Series. Working Paper 156.*

— (2014), "Predicting the restricted mean event time with the subject's baseline covariates in survival analysis," *Biostatistics*, 15, 222–233.

Uno, H., Claggett, B., Tian, L., Inoue, E., Gallo, P., Miyata, T., Schrag, D., Takeuchi, M., Uyama, Y., Zhao, L., et al. (2014), "Moving beyond the hazard ratio in quantifying the between-group difference in survival analysis," *Journal of Clinical Oncology*, 32, 2380–2385.

Zhang, M. and Schaubel, D. E. (2011), "Estimating differences in restricted mean lifetime using observational data subject to dependent censoring," *Biometrics*, 67, 740–749.

Zhao, L., Tian, L., Uno, H., Solomon, S. D., Pfeffer, M. A., Schindler, J. S., and Wei, L. J. (2012), "Utilizing the integrated difference of two survival functions to quantify the treatment contrast for designing, monitoring, and analyzing a comparative clinical study," *Clinical Trials*, 9, 570–577.

Zucker, D. M. (1998), "Restricted mean life with covariates: modification and extension of a useful survival analysis method," *Journal of the American Statistical Association*, 93, 702–709.