AstraZeneca

Design and analysis of studies based on hierarchical composite endpoints: Insights from the DARE-19 trial

Samvel B. Gasparyan

28 April 2022

PSI Webinar: Novel composite estimands and their analysis

Disclaimer

- The author of this presentation is an AstraZeneca employee, but the contents of this presentation do not necessarily reflect the viewpoints of AstraZeneca.

The presentation is based on

- *Gasparyan SB, Buenconsejo J, Kowalewski EK, Oscarsson J, Bengtsson OF, Esterline R, Koch GG, Berwanger O, Kosiborod MN "**Design and analysis of studies based on hierarchical composite endpoints: Insights from the DARE-19 trial**". Under review.*

Publications relevant to this presentation

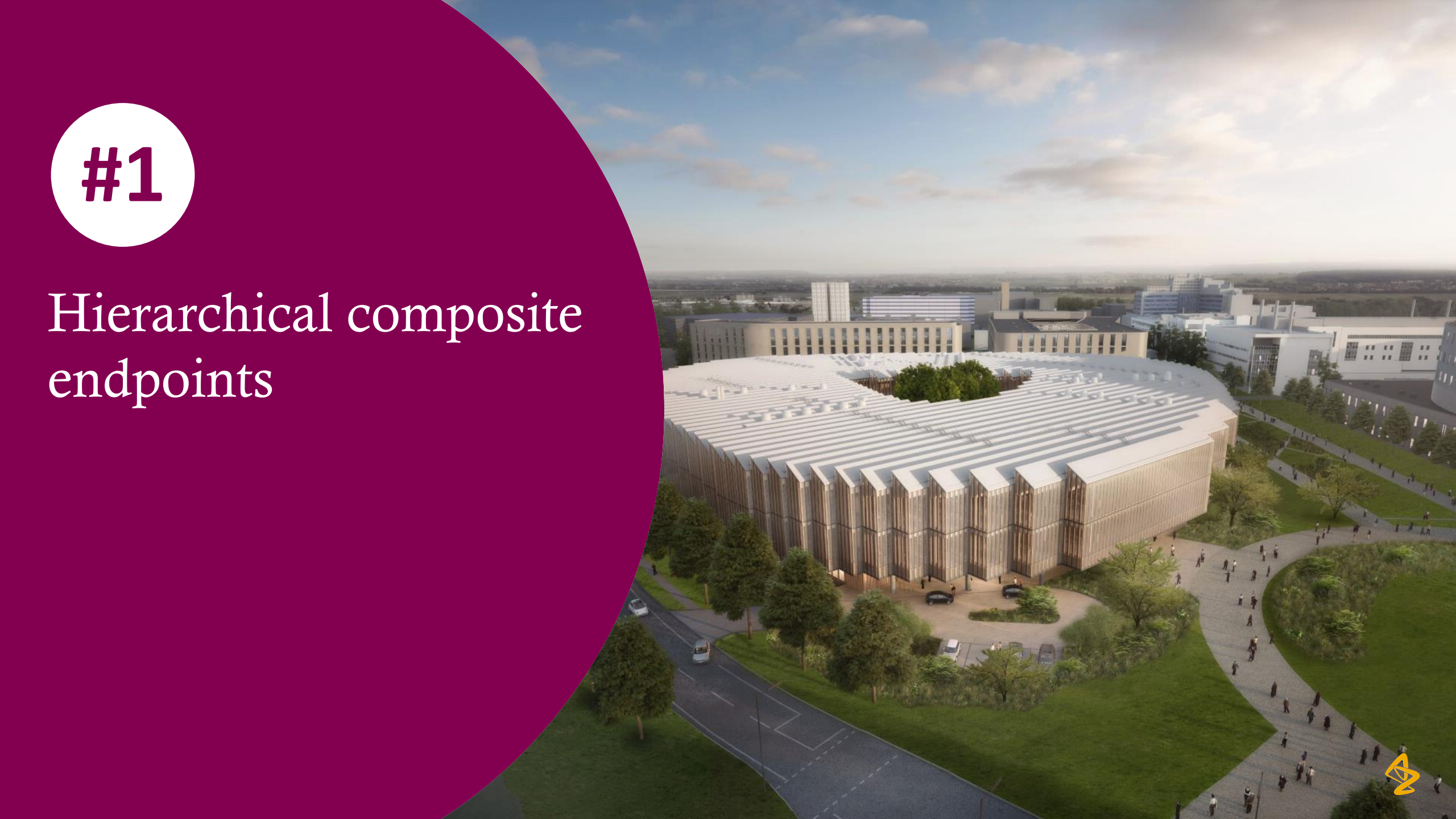
- *Gasparyan SB, Kowalewski EK, Folkvaljon F, Bengtsson O, Buenconsejo J, Adler J, Koch GG "**Power and sample size calculation for the win odds test: application to an ordinal endpoint in COVID-19 trials**". Journal of Biopharmaceutical Statistics, (Sep 2021), https://doi.org/10.1080/10543406.2021.1968893*
- *Gasparyan SB, Kowalewski EK, Koch GG "**Comments on 'Sample size formula for a win ratio endpoint' by R.X. Yu and J. Ganju**". Statistics in Medicine. 2022; 1-3. https://doi.org/10.1002/SIM.9379*
- *Kosiborod MN, Esterline R, Furtado RH, Oscarsson J, Gasparyan SB, Koch GG, Martinez F, Mukhtar O, Verma S, Chopra V, Buenconsejo J. "**Dapagliflozin in patients with cardiometabolic risk factors hospitalised with COVID-19 (DARE-19): a randomised, double-blind, placebo-controlled, phase 3 trial**". The Lancet Diabetes & Endocrinology, (2021 Sep 1); 9 (9): 586-94. https://doi.org/10.1016/S2213-8587(21)00180-7*

Agenda

1. Hierarchical composite endpoints
2. Analysis of HCE using win/Mann-Whitney odds
3. Designing trials with HCE and win odds
4. A novel HCE in COVID-19
5. Estimand in DARE-19: Dapagliflozin in inpatient COVID-19 setting

#1

Hierarchical composite endpoints



Hierarchical Composite Endpoints

- Hierarchical Composite Endpoints are composite endpoints which combine events of different clinical importance into an *ordinal* outcome.
- Each patient contributes to the analysis their most severe event.
- Unlike usual composite endpoints, the ordinal outcomes do not consider the components to be of equal importance but relies on the severity of events for prioritization.
- HCE are implemented in different disease areas with different names (win ratio endpoint, DOOR – the desirability of outcome ranking and so on).
- We use HCE as the term for the endpoint and we want to separate *the definition of the endpoint from the analysis method* (win ratio, win odds, ordinal logistic regression, rank based methods or so on).

Packer M. ***"Development and evolution of a hierarchical clinical composite end point for the evaluation of drugs and devices for acute and chronic heart failure: a 20-year perspective".*** Circulation, 134(21), 1664-1678 (2016).

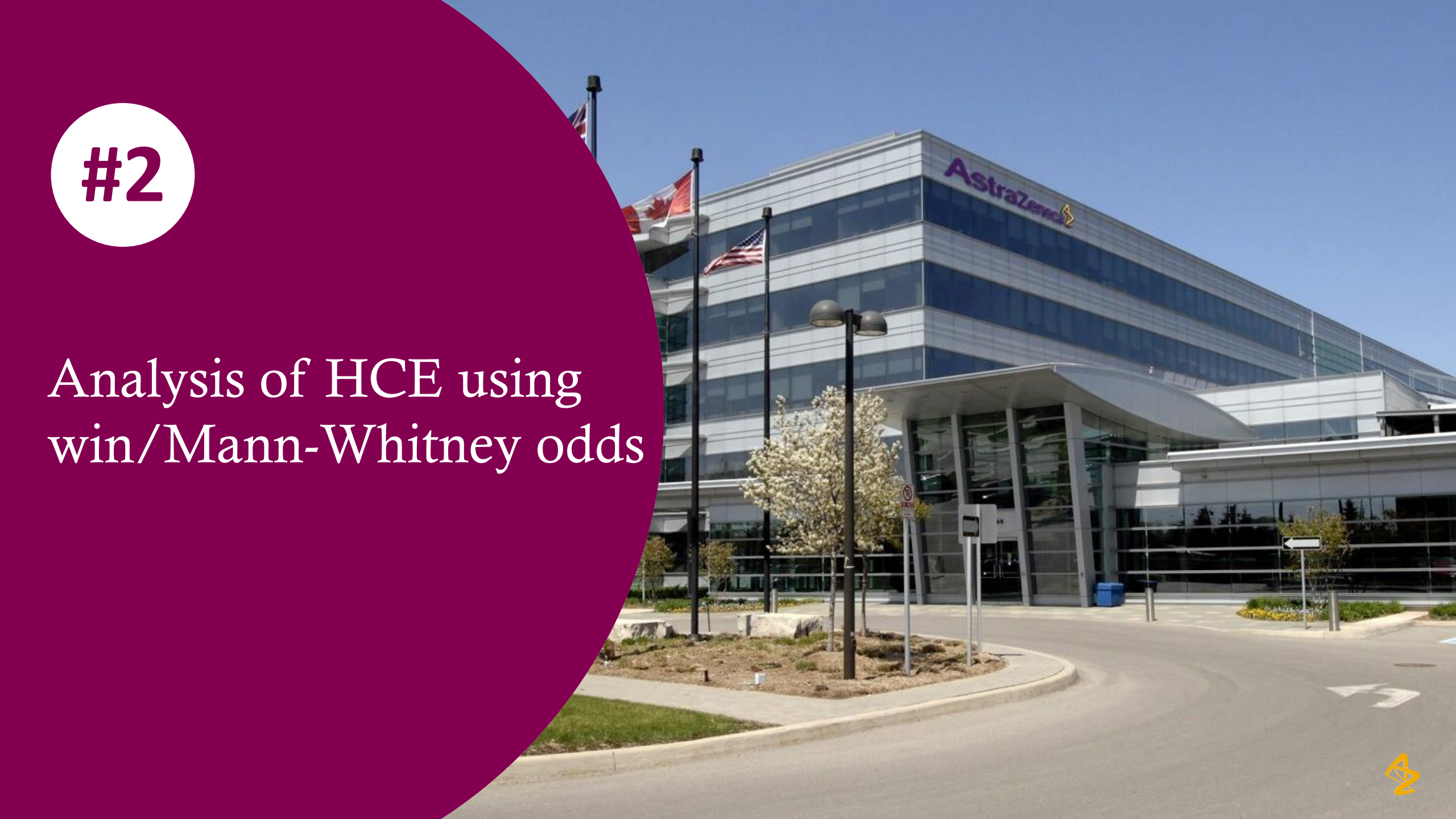
Example of a trial using HCE – EMPULSE trial

- The EMPULSE trial (acute HF, 530 patients) used an HCE combining ACM, total HFHs and change from baseline in KCCQ-TSS at Day 90.
- Analyzed using win ratio (stratified by de novo vs decompensated) it resulted in WR of 1.36 (1.09, 1.68), P = 0.0054.
- Empagliflozin was superior in 53.9% of paired comparisons and placebo was superior in 39.7%, whereas 6.4% of comparisons were tied.
- The time-to-event analysis for the composite of HFH and CVD resulted in HR of 0.69 (0.45, 1.08), therefore was statistically not significant.
- Hence combining KCCQ with HFH and CVD increased the power.
- Adjusted Mean difference between two groups in KCCQ-TSS change from baseline to day 90 was 4.45 (0.32–8.59).

Voors AA, Angermann CE, Teerlink JR, Collins SP, Kosiborod M, Biegus J, Ferreira JP, Nassif ME, Psotka MA, Tromp J, Borleffs C. ***"The SGLT2 inhibitor empagliflozin in patients hospitalized for acute heart failure: a multinational randomized trial"***. Nature Medicine. 2022 Feb 28:1-7.

#2

Analysis of HCE using win/Mann-Whitney odds



Win ratio with ties (win/Mann-Whitney odds)

- The concept of win ratio was introduced by Pocock et al. in 2012, based on the original idea of Finkelstein & Schoenfeld (1999).
- It is a statistic for comparison of two independent ordinal endpoints.
- Each patient in the active group is compared to each patient in the placebo group to decide the "winner" (the endpoint with lower or higher ordinal value, depending on the convention).
- The total number of "wins" of the active group is divided by total number of "wins" of the placebo group forming the win ratio statistic.
- Ties were disregarded in calculating the win ratio statistic.
- Dong et al. (2020) suggested to include ties as half wins, forming the win odds statistic.
- Comparison of outcomes works without controversies if the follow-up is the same.

Finkelstein DM, Schoenfeld DA ***"Combining mortality and longitudinal measures in clinical trials"****. Statistics in medicine. 1999 Jun 15;18(11):1341-54.*

*Pocock SJ, Ariti CA, Collier TJ, Wang D "****The win ratio: A new approach to the analysis of composite endpoints in clinical trials based on clinical priorities****". European Heart Journal, (2012); 33 (2): 176–182.*

*Dong G, Hoaglin DC, Qiu J, Matsouaka RA, Chang YW, Wang J, Vandemeulebroecke M "****The win ratio: On interpretation and handling of ties****". Statistics in Biopharmaceutical Research, (2020); 12 (1): 99–106.*

Hypothesis test for the win odds

- It is more convenient to work with win odds, because
  - If the number of ties is high, it will have a lower value than the WR, hence in the analysis it will account for patients who do not experience improvement in the active group.
  - If we denote the win probability (WP) as the total number of wins, plus half of the ties of the active group divided by total number of comparison, then

$$WO = \frac{WP}{1-WP} \text{ or } WP = \frac{WO}{WO+1}$$

    and the alternative hypothesis can be easily written as WP>0.5 which corresponds to WO>1 (the null hypothesis is WO=1 or WP=0.5)
  - This gives the advantage of estimating WP (using U-statistics) and then transforming its SE to get the SE of $\log(\text{WO})$ as $\frac{SE}{WP*(1-WP)}$.
  - WO>1 corresponds to a shift of active distribution with respect to the placebo distribution, in the "right" direction (to the right if higher ordinal values correspond to a better outcome/winning, or to the left, in the opposite case).
- For the review of advantages of the win odds see Brunner (2021)

*Brunner EM, Vandemeulebroecke M, Tobias M "**Win odds: An adaptation of the win ratio to include ties**". Statistics in Medicine, (2021); 40 (14):3367–3384.*

*Hoeffding W "**A class of statistics with asymptotically normal distribution**". The Annals of Mathematical Statistics, (1948); 19 (3):293–325.*

#3

Designing trials with HCE and win odds

Designing trials based on the win odds

- Gasparyan et al. (2021) provided power and sample size calculation formulas for the win odds.
- It requires only specification of the alternative value for the win odds to calculate the sample size for given power.
- To infer the "true" WO to design the study simulations can be useful, if the endpoint contains components of different types and no prior data on the same endpoint are available.
- Gasparyan et al. (2022) provided a generalization of the initial formula for the cases of substantial number of ties.
- Under the assumption of alpha=5% (2-sided), 1:1 allocation, the total sample size is given below

Table 1. Sample size based on the win odds test.

Win odds	Win probability	Power = 80%	Power = 90%
1.10	0.5238	4616	6179
1.15	0.5349	2151	2879
1.20	0.5455	1267	1696
1.25	0.5556	848	1135
1.30	0.5652	616	824
1.35	0.5745	472	632
1.40	0.5833	377	505
1.45	0.5918	311	416
1.50	0.6000	262	351

*Gasparyan SB, Kowalewski EK, Folkvaljon F, Bengtsson O, Buenconsejo J, Adler J, Koch GG "**Power and sample size calculation for the win odds test: application to an ordinal endpoint in COVID-19 trials**" Journal of Biopharmaceutical Statistics, Sep 2021, https://doi.org/10.1080/10543406.2021.1968893*

*Gasparyan SB, Kowalewski EK, Koch GG **"Comments on 'Sample size formula for a win ratio endpoint' by R.X. Yu and J. Ganju".** Statistics in Medicine. 2022; 1-3. https://doi.org/10.1002/SIM.9379*

Power of WO as a function of sample size

Under the assumption of alpha=5% (2-sided), 1:1 allocation, the dependence of power on the total sample size is given below

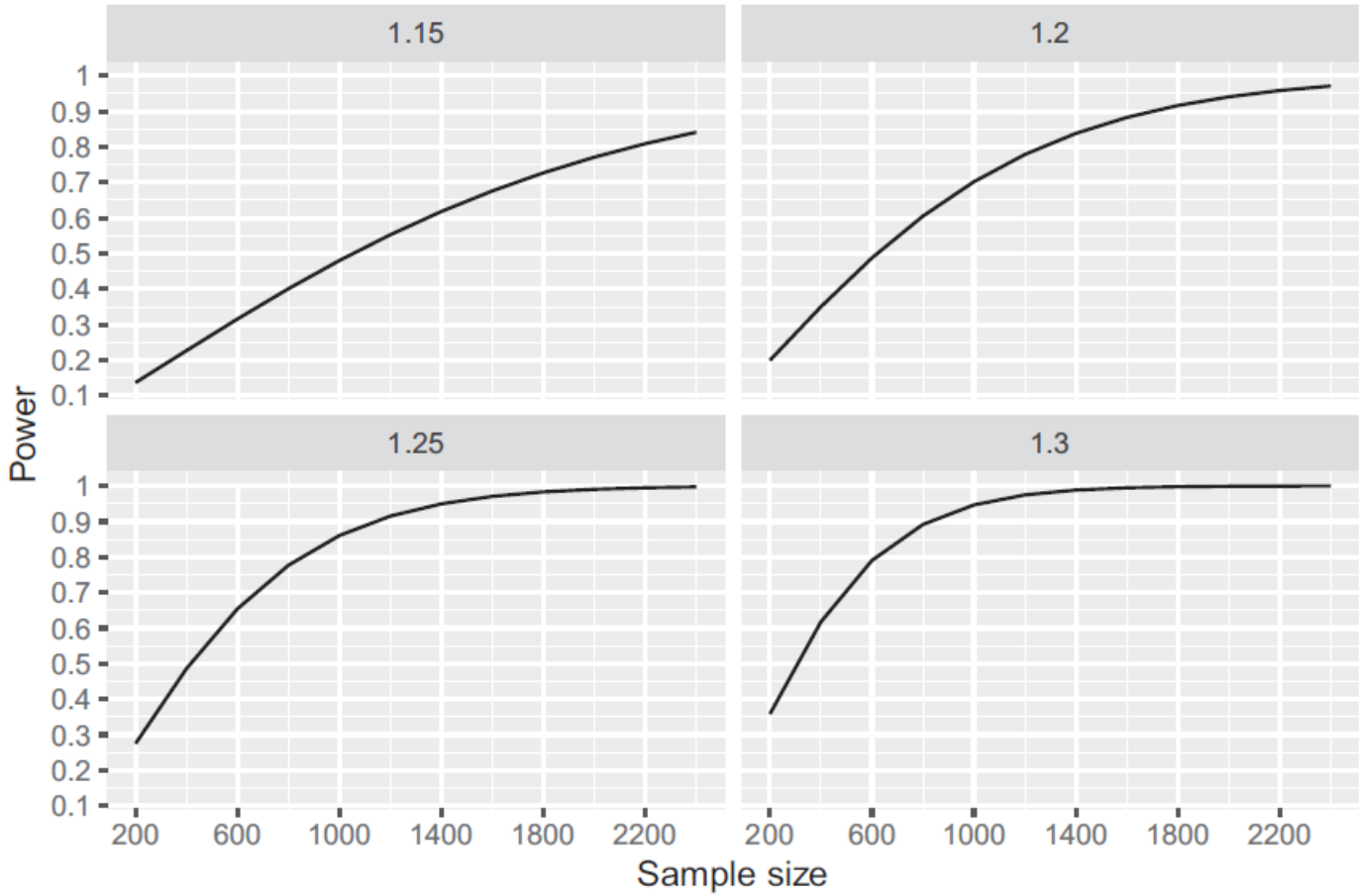


Figure 1. Win odds test power as a function of sample size, for the win odds values 1.15, 1.2, 1.25, 1.3.

*Gasparyan SB, Kowalewski EK, Folkvaljon F, Bengtsson O, Buenconsejo J, Adler J, Koch GG "**Power and sample size calculation for the win odds test: application to an ordinal endpoint in COVID-19 trials**" Journal of Biopharmaceutical Statistics, Sep 2021, https://doi.org/10.1080/10543406.2021.1968893*

Implementation in SAS

- Win odds is not implemented in SAS.
- The paper Gasparyan et al. (2021) showed theoretically how WO and its confidence intervals can be obtained from other statistics (Somers' D C/R and Fligner-Policello's statistic).
- Somers' D C/R is implemented in *proc freq*.
- Fligner-Policello's statistic is implemented in *proc npar1way*.
- Hence by simple programming WO and its CI can be obtained from these two procedures.
- Of course, it can be directly implemented as well (e.g. using *proc sql*).
- Or, again using the theory, can be directly implemented (more easily) using *proc rank*.

*Gasparyan SB, Kowalewski EK, Folkvaljon F, Bengtsson O, Buenconsejo J, Adler J, Koch GG "**Power and sample size calculation for the win odds test: application to an ordinal endpoint in COVID-19 trials**" Journal of Biopharmaceutical Statistics, Sep 2021, https://doi.org/10.1080/10543406.2021.1968893*

#4

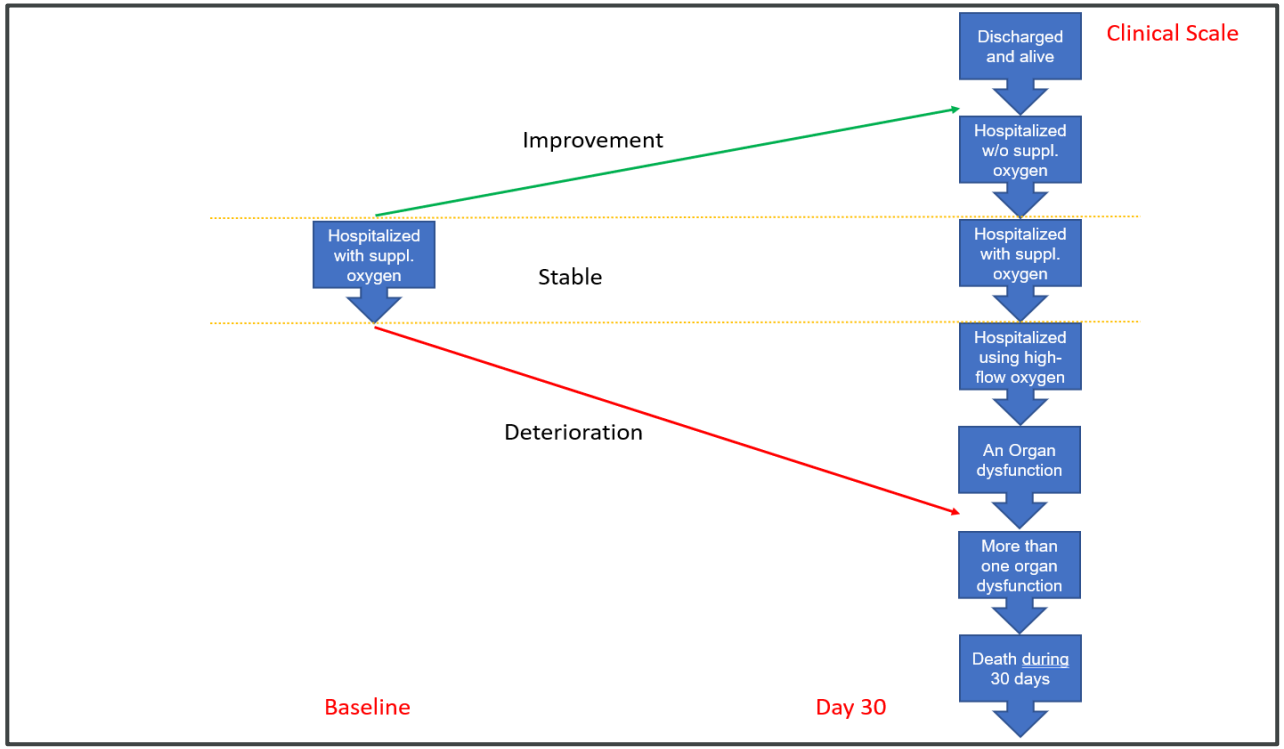
A novel HCE in COVID-19

Ordinal Scale and HCE

WHO 8-point ordinal scale

Not hospitalized, no limitations of activities	1
Not hospitalized, limitation of activities, home oxygen requirement, or both	2
Hospitalized, not requiring supplemental oxygen and no longer requiring ongoing	3
Medical care (used if hospitalization was extended for infection-control reasons)	
Hospitalized, not requiring supplemental oxygen but requiring ongoing medical care (Covid-19–related or other medical conditions)	4
Hospitalized, requiring any supplemental oxygen	5
Hospitalized, requiring noninvasive ventilation or use of high-flow oxygen devices	6
Hospitalized, receiving invasive mechanical ventilation or extracorporeal membrane oxygenation (ECMO)	7
Death	8

The novel Hierarchical Composite Endpoint



*Kosiborod M, Berwanger O, Koch GG, Martinez F, Mukhtar O, Verma S, Chopra V, Javaheri A, Ambery P, Gasparyan SB, Buenconsejo J "**Effects of dapagliflozin on prevention of major clinical events and recovery in patients with respiratory failure because of COVID-19: Design and rationale for the DARE-19 study**". Diabetes, Obesity and Metabolism, (2021 Apr); 23(4): 886-96.*

Recovery HCE in DARE-19

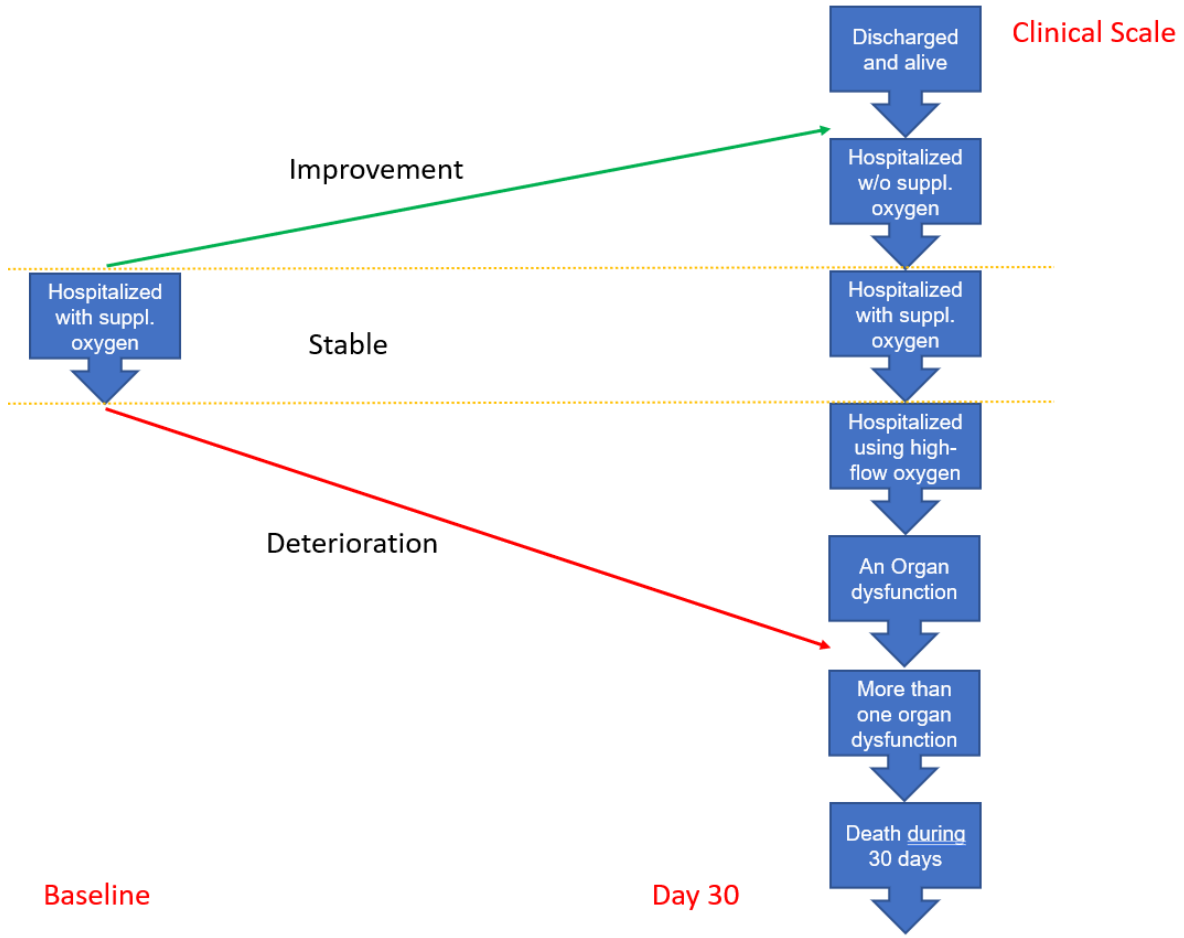
I. **Patients alive at the end of follow up (Day 30), without primary composite event and are discharged from hospital before Day 30** [*Ranking within this cohort will be based on the time to discharge, with patients being discharged later getting a worse rank*]

II. **Patients without primary composite event but hospitalized at the end of follow-up (Day 30),** [*Ranking within this cohort, from worse to better, includes patients on high-flow oxygen devices, patients requiring supplemental oxygen, and patients not requiring supplemental oxygen*]

III. **Patients who did not die but have more than one new or worsened organ dysfunction events**, [*Ranking within this cohort will be based on the number of events, with higher number getting a worse rank*]

IV. **Patients who did not die but have only one new or worsened organ dysfunction event,** [*Ranking within this cohort will be based on the timing of the event, with patients having the event sooner getting a worse rank. Type of organ dysfunction will not be considered*]

V. **Patients dying during the study**, [*Ranking within this cohort will be based on the timing of the event, with patients dying sooner getting a worse rank*]



*Kosiborod M, Berwanger O, Koch GG, Martinez F, Mukhtar O, Verma S, Chopra V, Javaheri A, Ambery P, Gasparyan SB, Buenconsejo J "**Effects of dapagliflozin on prevention of major clinical events and recovery in patients with respiratory failure because of COVID-19: Design and rationale for the DARE-19 study**". Diabetes, Obesity and Metabolism, (2021 Apr); 23(4): 886-96.*

Key Considerations on defining Recovery in DARE-19

Ordinal scale and "recovery" HCE

- Two important differences (based on the mode of action of dapagliflozin) between the ordinal scale endpoint and the suggested HCE are the following:

    1. The ordinal scale endpoints are assessed at a prespecified timepoint (for example, at Day 15), HCE uses severity of all events that a patient experiences during the 30 days of follow-up.
    2. The HCE takes into account in-hospital worsening of COVID-19 and not only the eventual discharge from hospital.

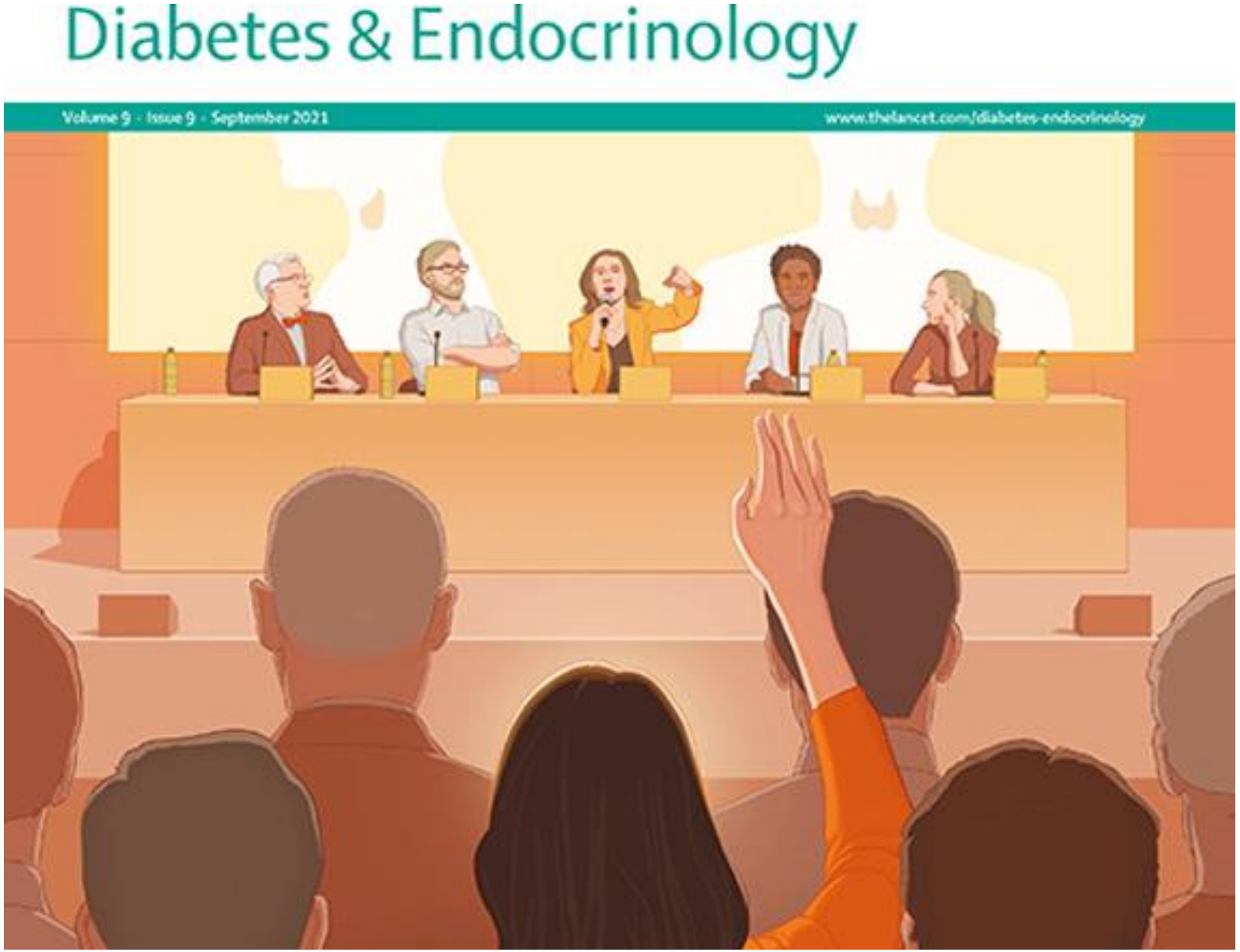
Therefore, the HCE is a recovery endpoint with a stricter definition of recovery. ***Recovery*** is represented on the clinical scale as ***improvement in clinical status*** compared to baseline:

1. Discharge from hospital before day 30 without in-hospital worsening and alive at Day 30; or
2. Still in hospital at Day 30, but without in-hospital worsening during the 30 days of hospitalization and without oxygen support.

- This HCE is essentially designed to capture any increase in the number of patients who recover/leave the hospital without complications in the active group compared to placebo, as well as reduction in the time to recovery. ***It combines disease specific events and events related to dapagliflozin's mode of action.***

*Kosiborod M, Berwanger O, Koch GG, Martinez F, Mukhtar O, Verma S, Chopra V, Javaheri A, Ambery P, Gasparyan SB, Buenconsejo J "**Effects of dapagliflozin on prevention of major clinical events and recovery in patients with respiratory failure because of COVID-19: Design and rationale for the DARE-19 study**". Diabetes, Obesity and Metabolism, (2021 Apr); 23(4): 886-96.*

#5
Estimand in DARE-19: Dapagliflozin in inpatient COVID-19 setting



Power for the HCE analysis in DARE-19

- Dapagliflozin in patients with cardiometabolic risk factors hospitalised with COVID-19 (DARE-19) was a randomized, double-blind, placebo-controlled, phase 3 trial with sample size of 1250.
- DARE-19 had dual primary endpoints of prevention, analyzed as a time-to-event endpoint, and a novel HCE analyzed using win odds (WO).
- The study was powered for WO=1.23 and it was estimated that approximately 1200 patients will be needed to attain 80% power, at a two-sided alpha of 2.5%.
- The minimal detectable WO was 1.15.

*Kosiborod MN, Esterline R, Furtado RH, Oscarsson J, Gasparyan SB, Koch GG, Martinez F, Mukhtar O, Verma S, Chopra V, Buenconsejo J. "**Dapagliflozin in patients with cardiometabolic risk factors hospitalised with COVID-19 (DARE-19): a randomised, double-blind, placebo-controlled, phase 3 trial**" The Lancet Diabetes & Endocrinology, (2021 Sep 1); 9 (9): 586-94, https://doi.org/10.1016/S2213-8587(21)00180-7*

Estimands (ICH E9 R1)

- An estimand is a ***precise description of the treatment*** effect reflecting the clinical question posed by a given ***clinical trial objective***.
- The targets of estimation are to be defined in advance of a clinical trial (*The design of a trial needs to be aligned to the estimands that reflect the trial objectives.*)
- Once defined, a trial can be designed to enable reliable estimation of the targeted treatment effect.
- The description of an estimand involves precise specifications of certain ***attributes***.
  - The **treatment** condition of interest
  - The **population** of patients targeted by the clinical question
  - The **variable** (or **endpoint**) to be obtained for each patient that is required to address the clinical question. *The specification of the variable might include whether the patient experiences an **intercurrent event**.*
  - The **population-level summary** for the variable should be specified, providing a basis for comparison between treatment conditions.

Estimands (layman's wording)

- Estimand says that the treatment effect is not a philosophical concept ("the drug works" or "the drug does not work" is meaningless).
- Efficaciousness of the intervention can be argued if
  - How to use the intervention? The intervention is clearly defined, including the dose and the frequency of use (The **treatment** attribute)
  - Who will benefit from the intervention? (The **population** attribute)
  - What the intervention will be affecting? (The **variable** attribute*)
  - How much the effect of the intervention will be? (The **population-level summary)**

Are there any events that will affect the measurements? (Intercurrent** events)*

DARE-19 Primary Objective and Estimand

Objective - To determine whether dapagliflozin 10 mg is superior to placebo, in terms of reducing complications or all-cause mortality, or improving clinical recovery in patients hospitalized with COVID-19.

Attributes of the estimand:

- **Treatment** = Dapagliflozin 10 mg once daily + Standard of Care (30 days of treatment, including after discharge)
- **Population** = Hospitalized patients (mild to severe) respiratory failure due to COVID-19 and cardio-renal-metabolic risk factors
- **The Endpoint** = Hierarchical Composite Endpoint of Recovery (which measures change in clinical status compared to baseline)
- **Population-level summary** = win odds (the odds that a randomly selected patient in the dapagliflozin group will have a better change in clinical status than a randomly selected patient in the placebo group)

Discharge from hospital is an important intercurrent event, since organ dysfunction is not defined after discharge (only death after discharge is included in the composite). This intercurrent event is handled using the composite strategy by including it in the composite.

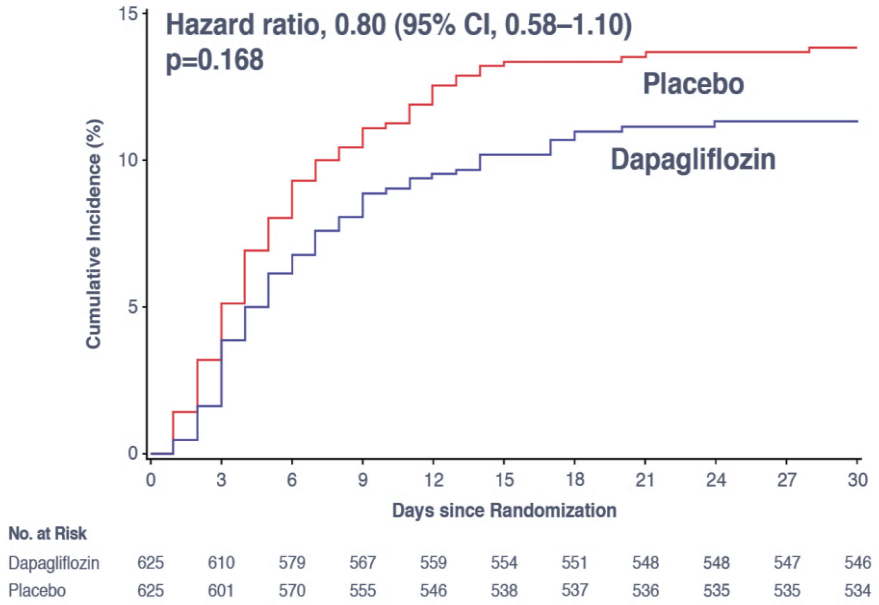
Intercurrent events for study drug discontinuation or initiation of concomitant medication were disregarded according to the treatment policy strategy.

Therefore, the ***estimand*** of the DARE-19 trial associated with the novel HCE can be written as (including the attributes and strategies of handling intercurrent events)

Dapagliflozin (10 mg once daily plus standard of care) improves recovery in adults with cardio-metabolic-renal risk factors and hospitalized with severe respiratory failure due to COVID-19, irrespective of concomitant treatment or study drug discontinuation.

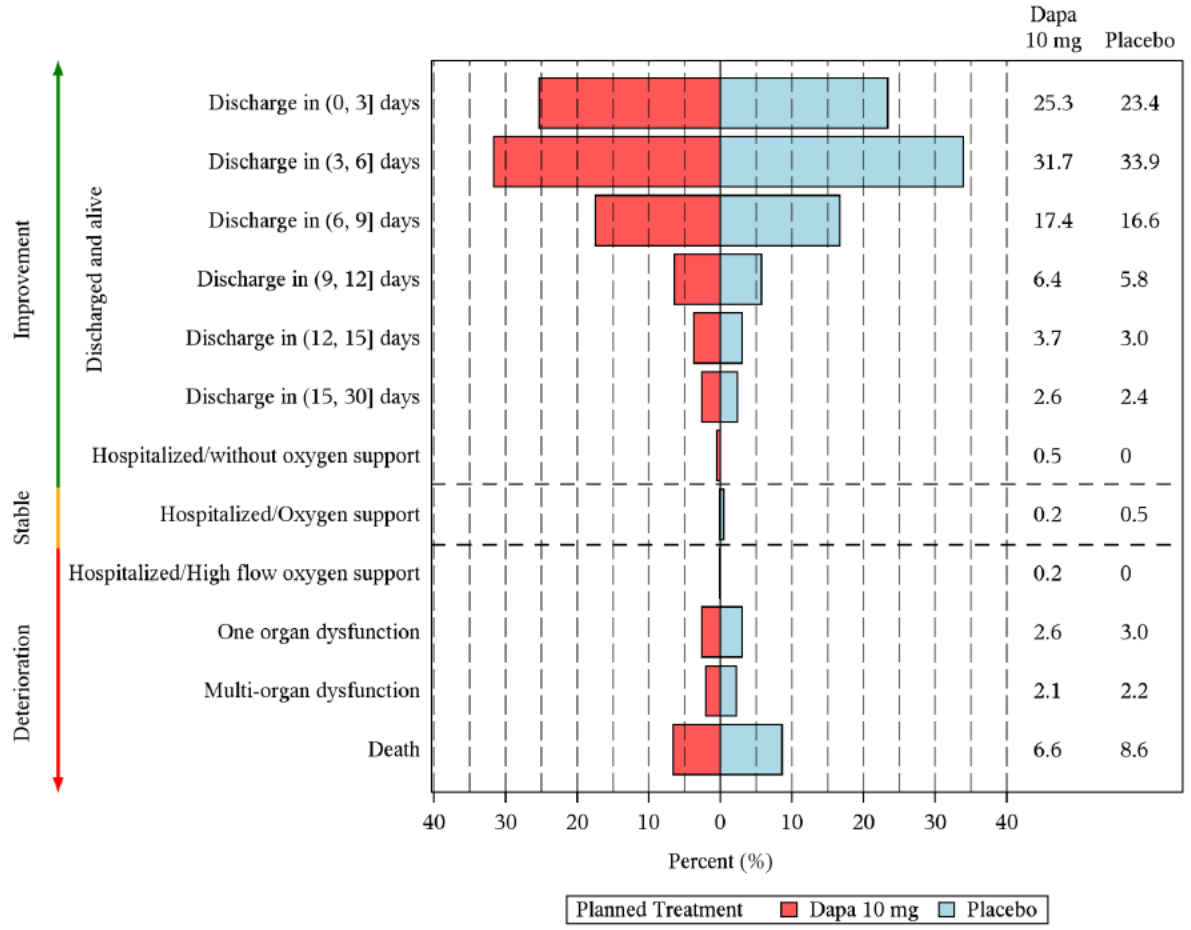
DARE-19 – Dual Objectives – Prevention and Recovery

Prevention endpoint



Recovery hierarchical composite endpoint

WO: 1.09 (95% CI 0.97, 1.22) P=0.136



*Kosiborod MN, Esterline R, Furtado RH, Oscarsson J, Gasparyan SB, Koch GG, Martinez F, Mukhtar O, Verma S, Chopra V, Buenconsejo J. "**Dapagliflozin in patients with cardiometabolic risk factors hospitalised with COVID-19 (DARE-19): a randomised, double-blind, placebo-controlled, phase 3 trial**" The Lancet Diabetes & Endocrinology, (2021 Sep 1); 9 (9): 586-94, https://doi.org/10.1016/S2213-8587(21)00180-7*

Interpretation

- The effect of dapagliflozin on change in clinical status of patients hospitalized with COVID-19 is characterized by a WO 1.09 (95% CI 0.97, 1.22) .
- Hence the odds that a randomly selected patient in the dapagliflozin group will have a better change in clinical status than a randomly selected patient in the placebo group is 1.09, which is not significantly different from WO=1.
- The minimal detectable WO was 1.15, which was considered the clinically meaningful threshold (corresponding to NNT=15 – Number of patients needed to be treated by dapagliflozin compared to being treated with placebo to get one patient with a better change in clinical status).
- Therefore, dapagliflozin did not show an improvement on the recovery in the COVID-19 setting

Recovery is improvement in clinical status compared to baseline.

Conclusion

- HCE are flexible endpoints that can be constructed in different disease areas and for drugs with different mode of actions, as a clinically meaningful measure of patient's condition.

- COVID-19 is a recent example where HCE can be used to construct a measure of patient’s clinical status.

- HCE can be analyzed using win/Mann-Whitney odds (win ratio with ties), which does not require distributional assumptions for estimation (including the proportionality assumption).

- Win odds can be used for designing new trials and provides a clinically meaningful treatment effect estimate as defined by the estimand framework.