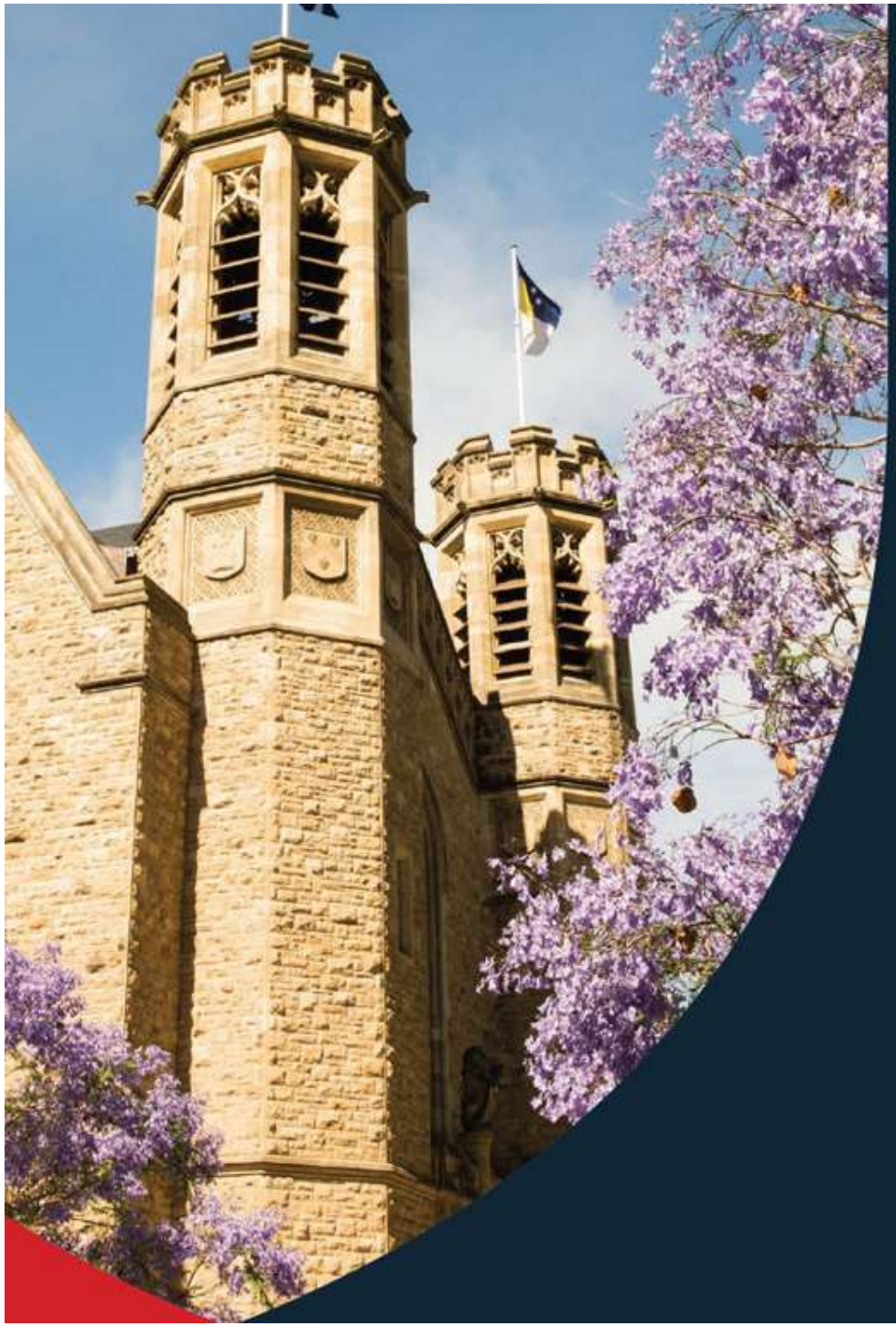


# Critical Appraisal: Clinical Trials

Danny Liew

THE UNIVERSITY of ADELAIDE

# Acknowledgement of Country

I acknowledge and pay my respects to the Kaurna people, the traditional custodians whose ancestral lands we gather on.

I acknowledge the deep feelings of attachment and relationship of the Kaurna people to country and I respect and value their past, present and ongoing connection to the land and cultural beliefs.

THE UNIVERSITY of ADELAIDE

The NEW ENGLAND JOURNAL of MEDICINE

ESTABLISHED IN 1812 NOVEMBER 21, 2019 VOL. 381 NO. 21

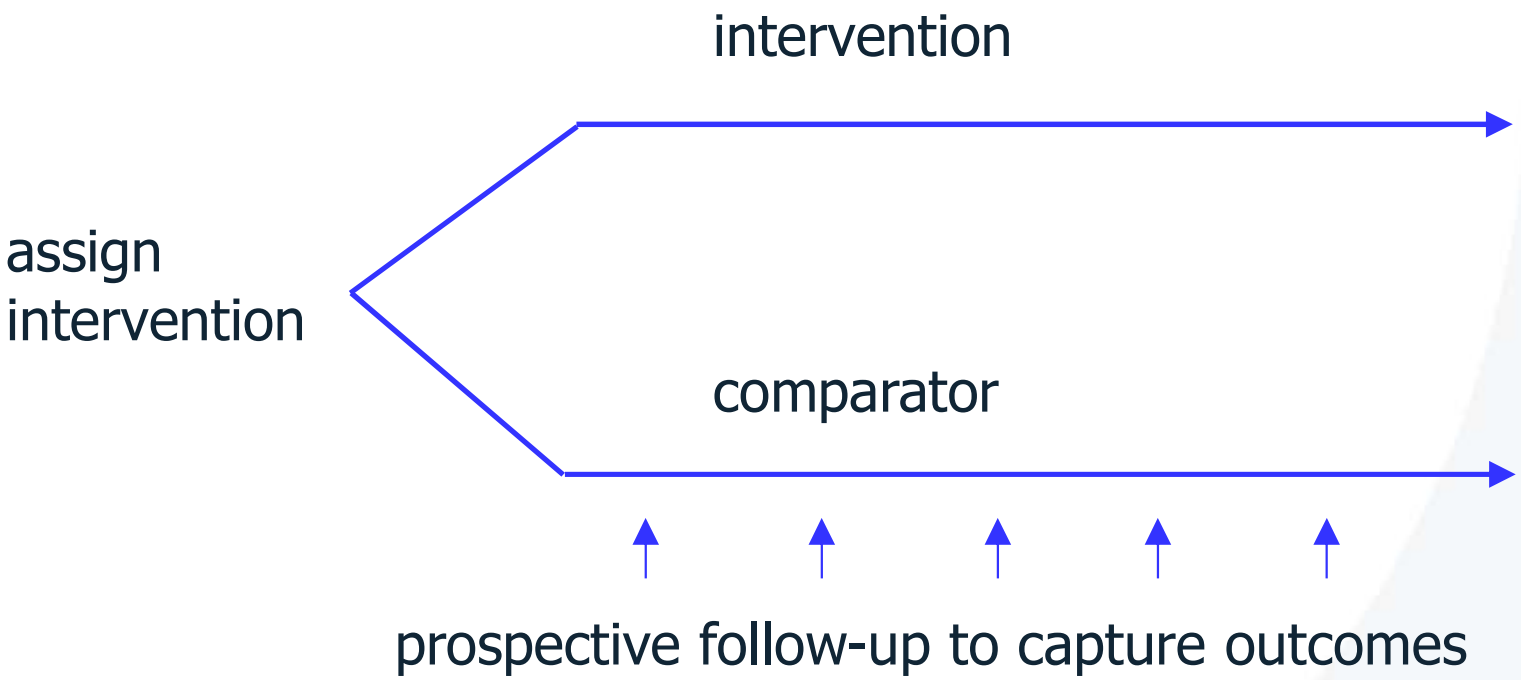
# Dapagliflozin in Patients with Heart Failure and Reduced Ejection Fraction

J.J.V. McMurray, S.D. Solomon, S.E. Inzucchi, L. Køber, M.N. Kosiborod, F.A. Martinez, P. Ponikowski, M.S. Sabatine, I.S. Anand, J. Bělohlávek, M. Böhm, C.-E. Chiang, V.K. Chopra, R.A. de Boer, A.S. Desai, M. Diez, J. Drozdz, A. Dukát, J. Ge, J.G. Howlett, T. Katova, M. Kitakaze, C.E.A. Ljungman, B. Merkely, J.C. Nicolau, E. O'Meara, M.C. Petrie, P.N. Vinh, M. Schou, S. Tereshchenko, S. Verma, C. Held, D.L. DeMets, K.F. Docherty, P.S. Jhund, O. Bengtsson, M. Sjöstrand, and A.-M. Langkilde, for the DAPA-HF Trial Committees and Investigators*

N Engl J Med 2019;381:1995-2008.

THE UNIVERSITY of ADELAIDE

# Clinical Trials (Classic Design)




THE UNIVERSITY of ADELAIDE

# Internal and External Validity

THE UNIVERSITY of ADELAIDE

# Internal Validity

- study rigour
- appropriate handling of limitations:
  - confounding
  - bias
  - statistical issues

THE UNIVERSITY of ADELAIDE

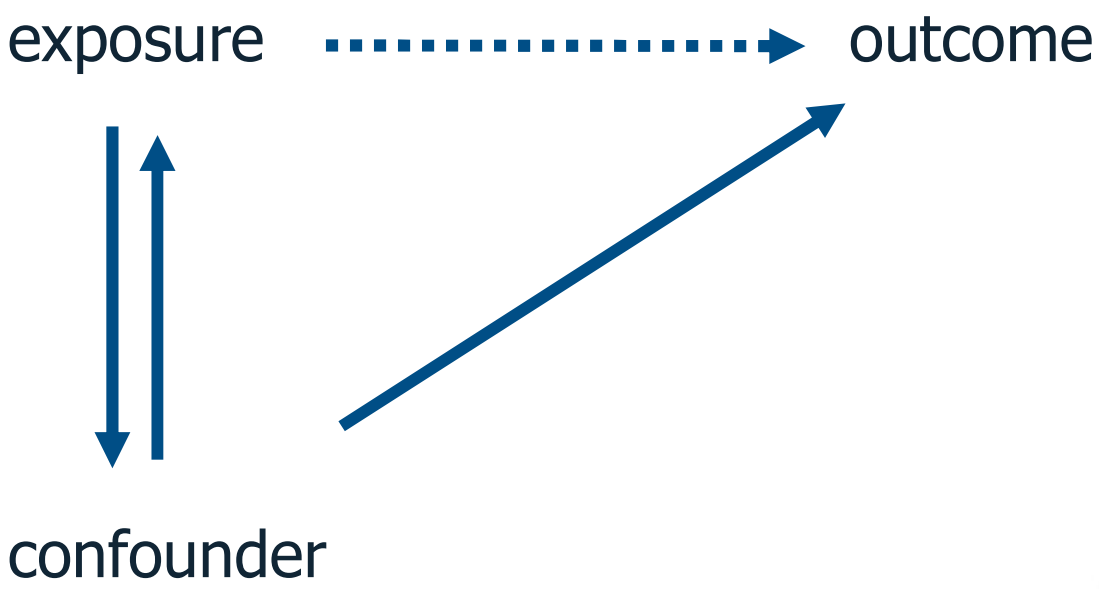
# External Validity

- applicability; generalisability; representativeness
- concordance of study and real-world settings re:
  - population
  - intervention
  - comparator
  - outcomes
  - timing

THE UNIVERSITY of ADELAIDE

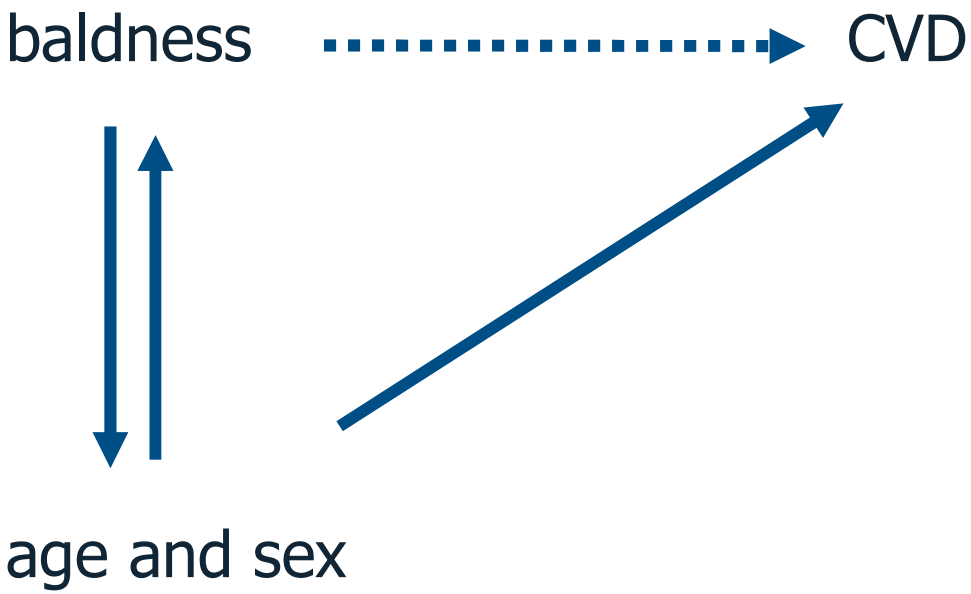
# Bias and Confounding

# Confounding



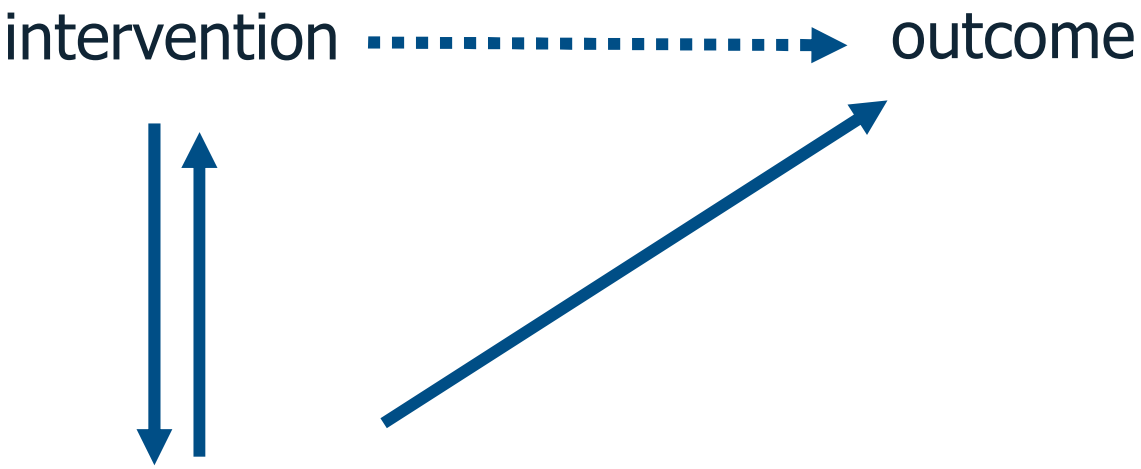

THE UNIVERSITY of ADELAIDE

# Confounding - Example




THE UNIVERSITY of ADELAIDE

# Confounding in Clinical Trials




THE UNIVERSITY of ADELAIDE

# Randomisation

- random allocation of subjects into each arm of a clinical trial
- objective: treatment groups identical in all aspects other than the intervention
- rationale: *reduce confounding*

THE UNIVERSITY of ADELAIDE

**Table 1. Characteristics of the Patients at Baseline.***

| Characteristic | Dapagliflozin (N=2373) | Placebo (N=2371) |
|---|---|---|
| Age — yr | 66.2±11.0 | 66.5±10.8 |
| Female sex — no. (%) | 564 (23.8) | 545 (23.0) |
| Body-mass index† | 28.2±6.0 | 28.1±5.9 |
| Race — no. (%)‡ | | |
| White | 1662 (70.0) | 1671 (70.5) |
| Black | 122 (5.1) | 104 (4.4) |
| Asian | 552 (23.3) | 564 (23.8) |
| Other | 37 (1.6) | 32 (1.3) |
| Region — no. (%) | | |
| North America | 335 (14.1) | 342 (14.4) |
| South America | 401 (16.9) | 416 (17.5) |
| Europe | 1094 (46.1) | 1060 (44.7) |
| Asia–Pacific | 543 (22.9) | 553 (23.3) |
| NYHA functional classification — no. (%) | | |
| II | 1606 (67.7) | 1597 (67.4) |
| III | 747 (31.5) | 751 (31.7) |
| IV | 20 (0.8) | 23 (1.0) |
| Heart rate — beats/min | 71.5±11.6 | 71.5±11.8 |
| Systolic blood pressure — mm Hg | 122.0±16.3 | 121.6±16.3 |
| Left ventricular ejection fraction — % | 31.2±6.7 | 30.9±6.9 |
| Median NT-proBNP (IQR) — pg/ml | 1428 (857–2655) | 1446 (857–2641) |
| Principal cause of heart failure — no. (%) | | |
| Ischemic | 1316 (55.5) | 1358 (57.3) |
| Nonischemic | 857 (36.1) | 830 (35.0) |
| Unknown | 200 (8.4) | 183 (7.7) |
| Medical history — no. (%) | | |
| Hospitalization for heart failure | 1124 (47.4) | 1127 (47.5) |
| Atrial fibrillation | 916 (38.6) | 902 (38.0) |
| Diabetes mellitus§ | 993 (41.8) | 990 (41.8) |

THE UNIVERSITY of ADELAIDE

N Engl J Med 2019;381:1995-2008.

# Bias

- error (unintentional) $\rightarrow$ systematic difference between/among groups
- leads to under or over-estimation of true results

- two main types:
  - *selection bias*
  - *information (measurement) bias*

# Selection Bias (1)

- systematic difference in characteristics of people selected for study and those not selected

  (specifically, people whose data were used for analyses and people whose data were not)

- example: the 'worried well'

- observed result may not reflect the true situation, and/or may not be generalisable

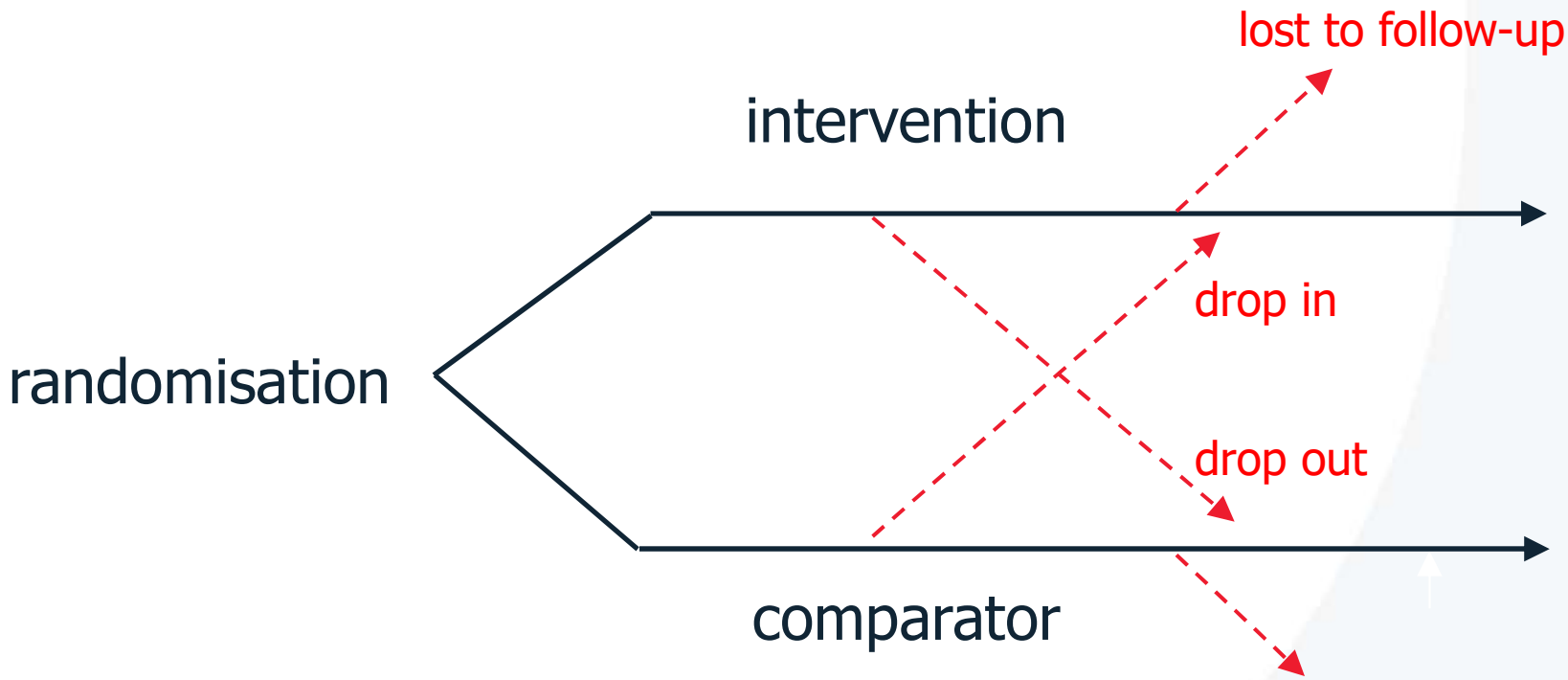
THE UNIVERSITY of ADELAIDE

# Selection Bias (2)

- systematic difference(s) in characteristics of subjects within groups being compared
- these differences are (partly) responsible for the observed study results

THE UNIVERSITY of ADELAIDE

# Cross-Over in (Parallel) Clinical Trials




THE UNIVERSITY of ADELAIDE

# Cross-Over in (Parallel) Clinical Trials

source of *selection bias* if significant and reasons likely to influence outcomes

eg, sick subjects cease active drug due to side effects

→ healthier group on active drug (less outcomes)

→ perception that active drug is better than placebo

THE UNIVERSITY of ADELAIDE

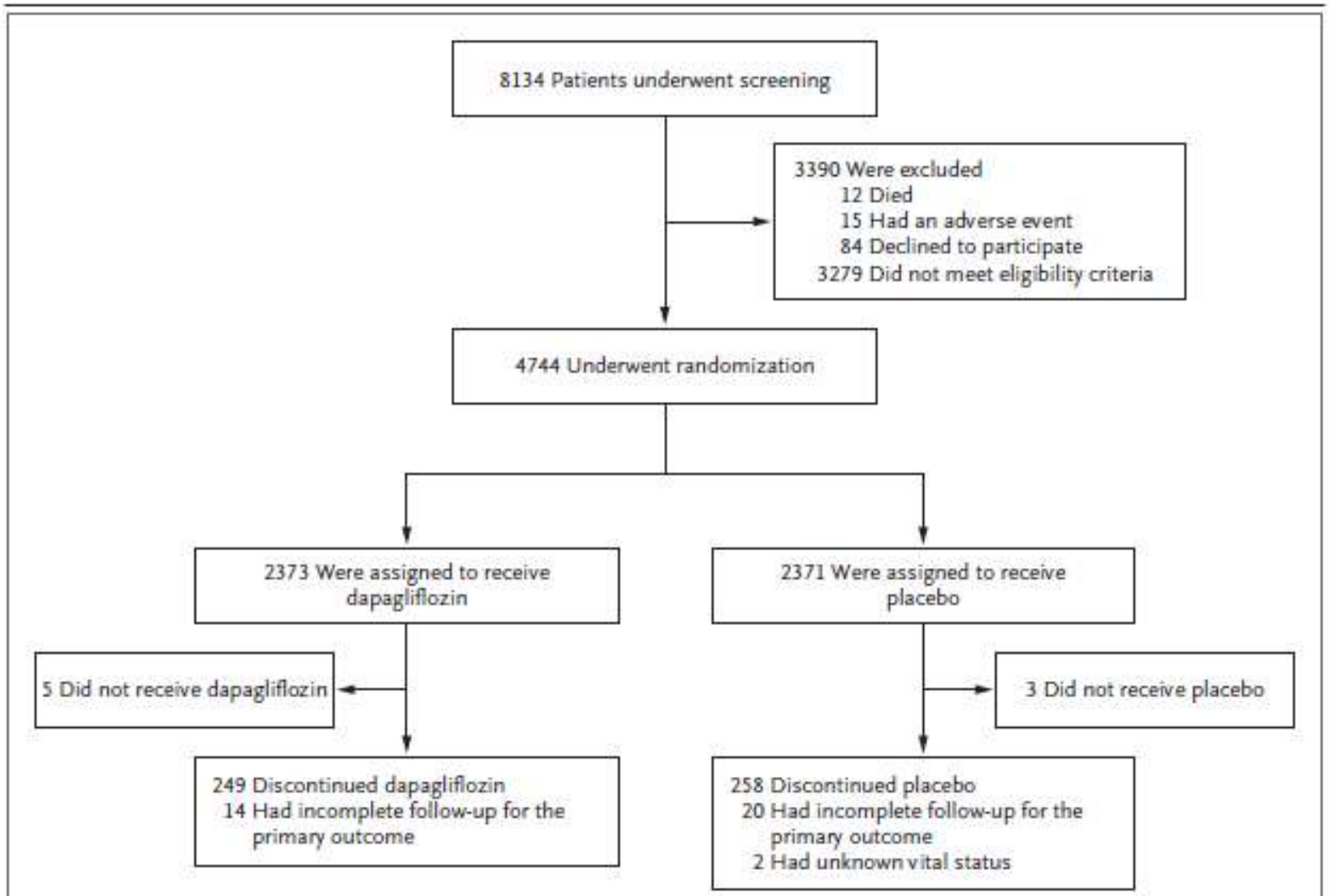



**Figure 1. Enrollment and Follow-up.**

All the patients who underwent randomization were included in the primary analysis. Patients who did not receive a dose of either dapagliflozin or placebo were excluded from the safety analysis.

N Engl J Med 2019;381:1995-2008.

# Intention-to-Treat Analysis

- assume that subjects remained in randomised group, regardless of cross-over
- rationale: reduce *selection bias*

THE UNIVERSITY of ADELAIDE

# Intention-to-Treat Analysis

- always *under-estimates any treatment effect*

  (ie, provides conservative estimate)

- reason: cross-over introduces overlap in

  treatment between groups, which is ignored

THE UNIVERSITY of ADELAIDE

We included data from all the patients who had undergone randomization in the analyses of the primary and secondary outcomes, according to the intention-to-treat principle. Baseline characteristics were summarized as means and stan-

N Engl J Med 2019;381:1995-2008.

THE UNIVERSITY of ADELAIDE

# Information Bias

- systematic difference(s) in the way information is collected between/among groups being compared
- differences are (partly) responsible for the observed study results
- arises when there is variability (especially subjectivity) in methods for collecting information

THE UNIVERSITY of ADELAIDE

# Blinding in Clinical Trials

- non-awareness of intervention allocation
- *single-blind:* subjects unaware (eg, use of placebo)
- *double-blind:* subjects and investigators unaware
- rationale: reduce *information (observer) bias*

THE UNIVERSITY of ADELAIDE

# Objective Outcome Ascertainment

- standardised criteria to define outcomes
- centralised ascertainment
- rationale: reduce *information (observer) bias*

THE UNIVERSITY of ADELAIDE

and death from any cause.[12] All outcomes were adjudicated by the members of a clinical-events committee, who were unaware of trial-group assignments, according to prespecified criteria (with definitions listed in the Supplementary Appendix).[15]

N Engl J Med 2019;381:1995-2008.

THE UNIVERSITY of ADELAIDE

# Statistical Errors

# Statistical Errors: Errors About Inference

| | | **True situation** | |
|---|---|---|---|
| | | No result present | Result present |
| **Conclusion from study** | No result present |  | Type II error (β) |
| | Result present | Type I error (α) |  |

= 1-β
= power

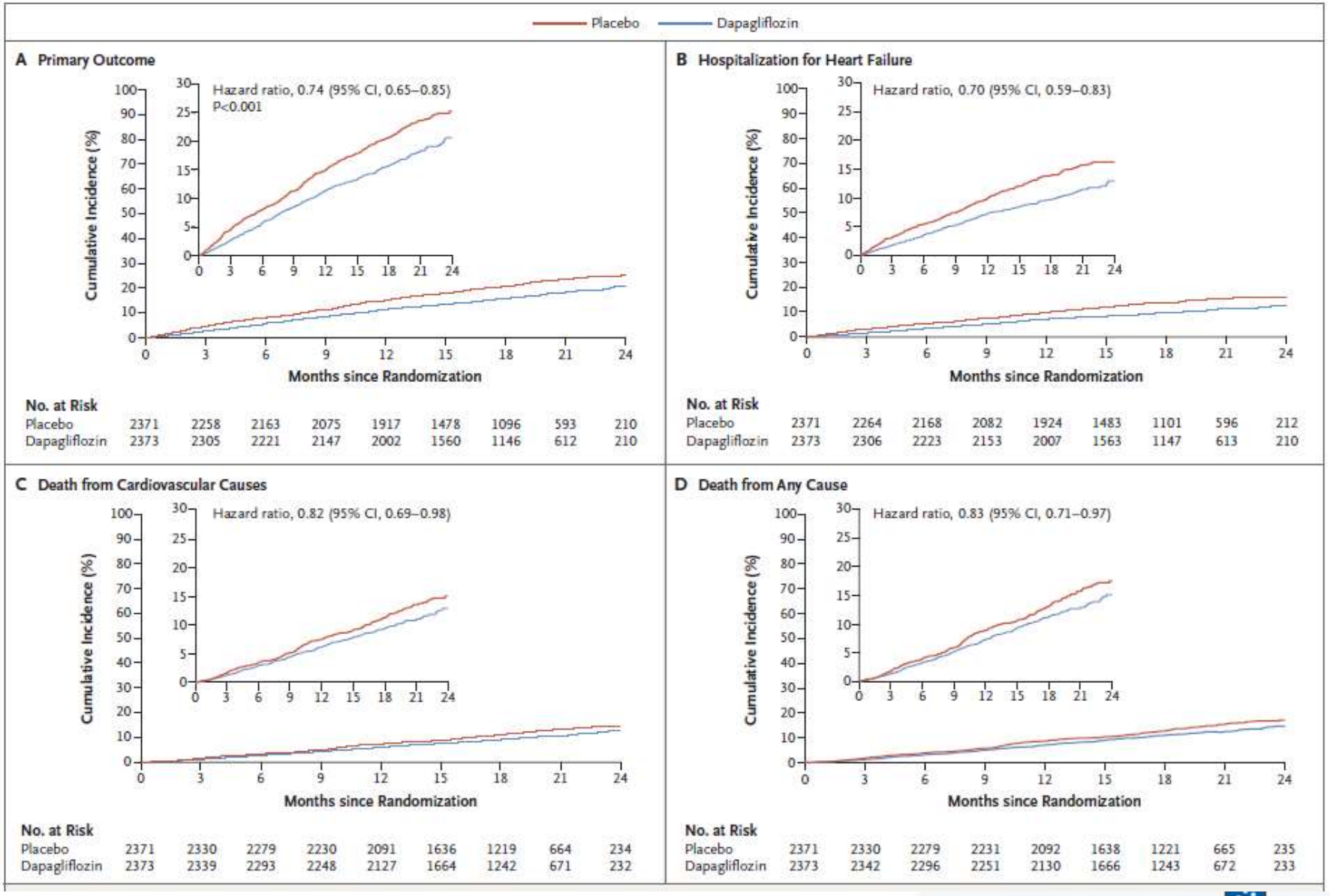
THE UNIVERSITY of ADELAIDE

**Table 2. Primary and Secondary Cardiovascular Outcomes and Adverse Events of Special Interest.***

| Variable | Dapagliflozin (N=2373) | | Placebo (N=2371) | | Hazard or Rate Ratio or Difference (95% CI) | P Value |
|---|---|---|---|---|---|---|
| | *values* | *events/100 patient-yr* | *values* | *events/100 patient-yr* | | |
| **Efficacy outcomes** | | | | | | |
| Primary composite outcome — no. (%)† | 386 (16.3) | 11.6 | 502 (21.2) | 15.6 | 0.74 (0.65 to 0.85) | <0.001 |
| Hospitalization or an urgent visit for heart failure | 237 (10.0) | 7.1 | 326 (13.7) | 10.1 | 0.70 (0.59 to 0.83) | NA |
| Hospitalization for heart failure | 231 (9.7) | 6.9 | 318 (13.4) | 9.8 | 0.70 (0.59 to 0.83) | NA |
| Urgent heart-failure visit | 10 (0.4) | 0.3 | 23 (1.0) | 0.7 | 0.43 (0.20 to 0.90) | NA |
| Cardiovascular death | 227 (9.6) | 6.5 | 273 (11.5) | 7.9 | 0.82 (0.69 to 0.98) | NA |
| Secondary outcomes | | | | | | |
| Cardiovascular death or heart-failure hospitalization — no. (%) | 382 (16.1) | 11.4 | 495 (20.9) | 15.3 | 0.75 (0.65 to 0.85) | <0.001 |
| Total no. of hospitalizations for heart failure and cardiovascular deaths‡ | 567 | — | 742 | — | 0.75 (0.65 to 0.88) | <0.001 |
| Change in KCCQ total symptom score at 8 mo§ | 6.1±18.6 | — | 3.3±19.2 | — | 1.18 (1.11 to 1.26) | <0.001 |
| Worsening renal function — no. (%)¶ | 28 (1.2) | 0.8 | 39 (1.6) | 1.2 | 0.71 (0.44 to 1.16) | NA |
| Death from any cause — no. (%) | 276 (11.6) | 7.9 | 329 (13.9) | 9.5 | 0.83 (0.71 to 0.97) | NA |
| **Safety outcomes‖** | | | | | | |
| Discontinuation due to adverse event — no./total no. (%) | 111/2368 (4.7) | — | 116/2368 (4.9) | — | — | 0.79 |
| Adverse events of interest — no./total no. (%) | | | | | | |
| Volume depletion | 178/2368 (7.5) | — | 162/2368 (6.8) | — | — | 0.40 |
| Renal adverse event | 153/2368 (6.5) | — | 170/2368 (7.2) | — | — | 0.36 |
| Fracture | 49/2368 (2.1) | | 50/2368 (2.1) | — | — | 1.00 |

N Engl J Med 2019;381:1995-2008.

THE UNIVERSITY of ADELAIDE




THE UNIVERSITY of ADELAIDE

N Engl J Med 2019;381:1995-2008.

# Lack of Statistical Significance

1 of 2 possible reasons:

i. no real result exists

ii. a real result exists, but the study lacked power to detect it (Type II error)

NB:

- studies (especially trials) are designed around, and powered for, primary outcome(s) only
- be wary of over-interpreting the findings for outcomes around which studies were *not* designed

THE UNIVERSITY of ADELAIDE

# Multiple Hypotheses Testing

- convention: set probability of type I error at 0.05

- in 1 analysis, chance of type I error = 0.05

  in 2 analyses, chance of *any* type 1 error = 1 - (0.95)^2 = 0.098

  in 20 analyses, chance of *any* type 1 error = 1 - (0.95)^20 = 0.642

- if planning to undertake multiple analyses, need to adjust cut-off for significance of p-value

  eg, Bonferroni correction: divide 0.05 by number of analyses

THE UNIVERSITY of ADELAIDE

# Power and Sample Size

# Power

- power = chance of detecting a result if it truly exists
- usually pre-determined at 80% or 90%
- insufficient power: study may fail to demonstrate an result even if a true and important one exists

THE UNIVERSITY of ADELAIDE

## Sample Size

during study design, required sample size determined by:

- Type I error (α) - usually 0.05
- power = 1 minus Type II error (β) - usually 0.8 or 0.9
- outcome of interest - likelihood and variability
- size of the effect to be detected: *minimal clinically important difference, MCID*

THE UNIVERSITY of ADELAIDE

## STATISTICAL ANALYSIS

We calculated that 844 primary outcome events would provide the trial with a power of 90% to detect a hazard ratio of 0.80 for the comparison between dapagliflozin and placebo, using a two-sided alpha level of 0.05. With an expected annual event incidence of 11% in the placebo group, we estimated that the enrollment of approximately 4500 patients would provide the required number of primary events, based on an anticipated recruitment period of 18 months and an average follow-up period of approximately 24 months.

N Engl J Med 2019;381:1995-2008.

THE UNIVERSITY of ADELAIDE

**Table 2. Primary and Secondary Cardiovascular Outcomes and Adverse Events of Special Interest.***

| Variable | Dapagliflozin (N=2373) | | Placebo (N=2371) | | Hazard or Rate Ratio or Difference (95% CI) | P Value |
|---|---|---|---|---|---|---|
| | *values* | *events/100 patient-yr* | *values* | *events/100 patient-yr* | | |
| **Efficacy outcomes** | | | | | | |
| Primary composite outcome — no. (%)† | 386 (16.3) | 11.6 | 502 (21.2) | 15.6 | 0.74 (0.65 to 0.85) | <0.001 |
| Hospitalization or an urgent visit for heart failure | 237 (10.0) | 7.1 | 326 (13.7) | 10.1 | 0.70 (0.59 to 0.83) | NA |
| Hospitalization for heart failure | 231 (9.7) | 6.9 | 318 (13.4) | 9.8 | 0.70 (0.59 to 0.83) | NA |
| Urgent heart-failure visit | 10 (0.4) | 0.3 | 23 (1.0) | 0.7 | 0.43 (0.20 to 0.90) | NA |
| Cardiovascular death | 227 (9.6) | 6.5 | 273 (11.5) | 7.9 | 0.82 (0.69 to 0.98) | NA |
| Secondary outcomes | | | | | | |
| Cardiovascular death or heart-failure hospitalization — no. (%) | 382 (16.1) | 11.4 | 495 (20.9) | 15.3 | 0.75 (0.65 to 0.85) | <0.001 |
| Total no. of hospitalizations for heart failure and cardiovascular deaths‡ | 567 | — | 742 | — | 0.75 (0.65 to 0.88) | <0.001 |
| Change in KCCQ total symptom score at 8 mo§ | 6.1±18.6 | — | 3.3±19.2 | — | 1.18 (1.11 to 1.26) | <0.001 |
| Worsening renal function — no. (%)¶ | 28 (1.2) | 0.8 | 39 (1.6) | 1.2 | 0.71 (0.44 to 1.16) | NA |
| Death from any cause — no. (%) | 276 (11.6) | 7.9 | 329 (13.9) | 9.5 | 0.83 (0.71 to 0.97) | NA |
| **Safety outcomes‖** | | | | | | |
| Discontinuation due to adverse event — no./total no. (%) | 111/2368 (4.7) | — | 116/2368 (4.9) | — | — | 0.79 |
| Adverse events of interest — no./total no. (%) | | | | | | |
| Volume depletion | 178/2368 (7.5) | — | 162/2368 (6.8) | — | — | 0.40 |
| Renal adverse event | 153/2368 (6.5) | — | 170/2368 (7.2) | — | — | 0.36 |
| Fracture | 49/2368 (2.1) | | 50/2368 (2.1) | — | — | 1.00 |

N Engl J Med 2019;381:1995-2008.

THE UNIVERSITY of ADELAIDE

# Equivalence and Non-Inferiority Trials

# Equivalence and Non-Inferiority Trials

- comparison of new intervention against current best practice (usually active)
- to demonstrate:
  - *equivalence* - not more *and* not less efficacious
  - *non-inferiority* - not less efficacious

THE UNIVERSITY of ADELAIDE

# Equivalence and Non-Inferiority Trials

- advantage(s) of new intervention in terms of factors other than efficacy

  eg, adverse effects, costs, pharmacokinetics

- to join an existing market

THE UNIVERSITY of ADELAIDE

# Per-Protocol Analysis

- grouping according to actual treatment
- reason: intention-to-treat analysis always *under-estimates any treatment effect*
- therefore, not conservative against a hypothesis of equivalence

THE UNIVERSITY of ADELAIDE

# External Validity

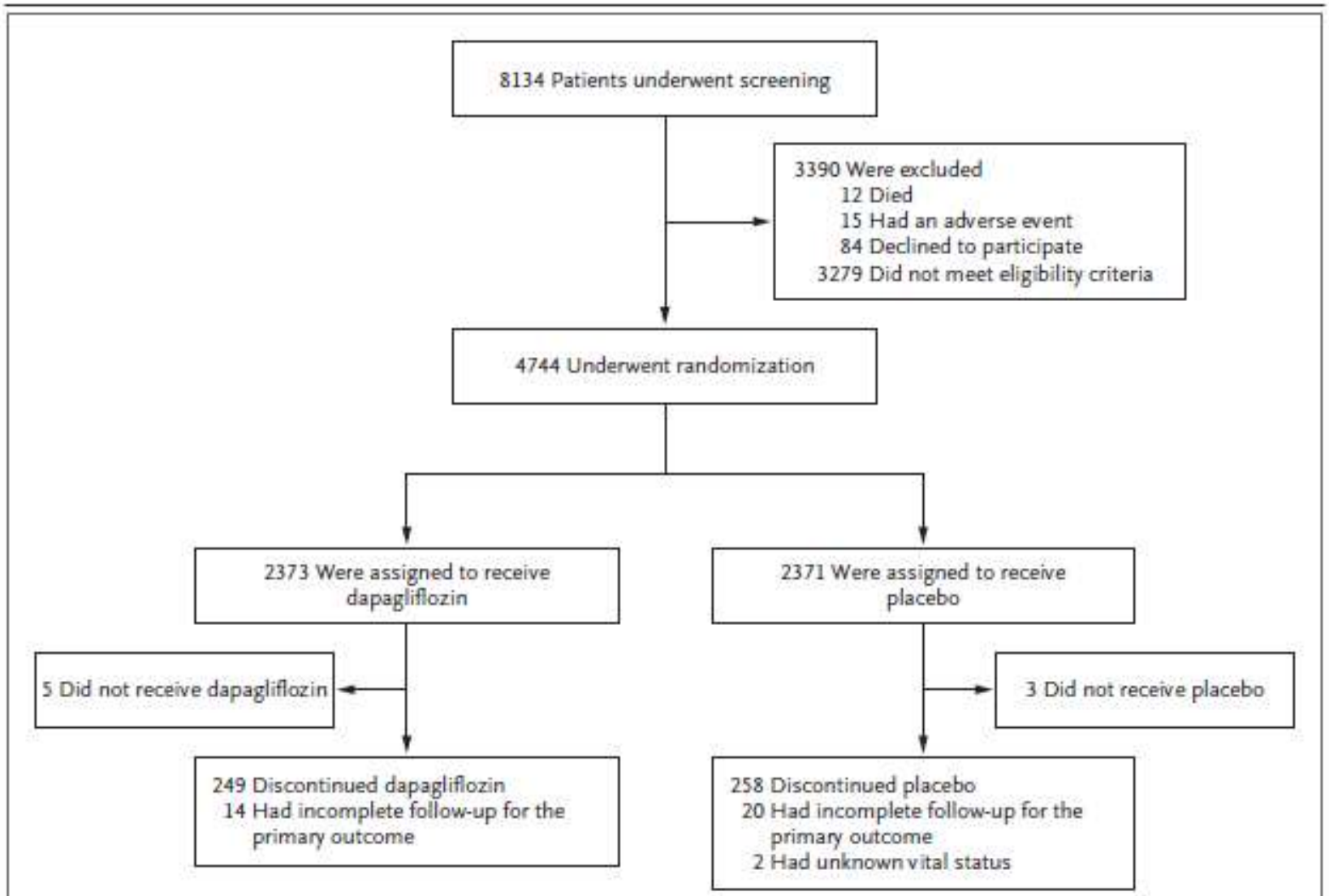



**Figure 1. Enrollment and Follow-up.**

All the patients who underwent randomization were included in the primary analysis. Patients who did not receive a dose of either dapagliflozin or placebo were excluded from the safety analysis.

N Engl J Med 2019;381:1995-2008.

**Table 1. Characteristics of the Patients at Baseline.***

| Characteristic | Dapagliflozin (N=2373) | Placebo (N=2371) |
|---|---|---|
| Age — yr | 66.2±11.0 | 66.5±10.8 |
| Female sex — no. (%) | 564 (23.8) | 545 (23.0) |
| Body-mass index† | 28.2±6.0 | 28.1±5.9 |
| Race — no. (%)‡ | | |
| White | 1662 (70.0) | 1671 (70.5) |
| Black | 122 (5.1) | 104 (4.4) |
| Asian | 552 (23.3) | 564 (23.8) |
| Other | 37 (1.6) | 32 (1.3) |
| Region — no. (%) | | |
| North America | 335 (14.1) | 342 (14.4) |
| South America | 401 (16.9) | 416 (17.5) |
| Europe | 1094 (46.1) | 1060 (44.7) |
| Asia–Pacific | 543 (22.9) | 553 (23.3) |
| NYHA functional classification — no. (%) | | |
| II | 1606 (67.7) | 1597 (67.4) |
| III | 747 (31.5) | 751 (31.7) |
| IV | 20 (0.8) | 23 (1.0) |
| Heart rate — beats/min | 71.5±11.6 | 71.5±11.8 |
| Systolic blood pressure — mm Hg | 122.0±16.3 | 121.6±16.3 |
| Left ventricular ejection fraction — % | 31.2±6.7 | 30.9±6.9 |
| Median NT-proBNP (IQR) — pg/ml | 1428 (857–2655) | 1446 (857–2641) |
| Principal cause of heart failure — no. (%) | | |
| Ischemic | 1316 (55.5) | 1358 (57.3) |
| Nonischemic | 857 (36.1) | 830 (35.0) |
| Unknown | 200 (8.4) | 183 (7.7) |
| Medical history — no. (%) | | |
| Hospitalization for heart failure | 1124 (47.4) | 1127 (47.5) |
| Atrial fibrillation | 916 (38.6) | 902 (38.0) |
| Diabetes mellitus§ | 993 (41.8) | 990 (41.8) |

THE UNIVERSITY of ADELAIDE

N Engl J Med 2019;381:1995-2008.