

Handling missing data in exploratory analyses using RWD, “Collect the data you love or love the data you get” – *A case study*

Jean-Marc Ferran¹, Anne Helene Olsen², Navid Nedjatian³

PHUSE EU Connect 2023

RE05

8. November 2023

¹Qualiance ApS, Copenhagen, Denmark; ² Novo Nordisk A/S, Søborg, Denmark; ³Novo Nordisk Health Care AG, Zurich, Switzerland

Disclosures

- **Jean-Marc Ferran** is an external consultant at Qualiance ApS providing Statistical support to Novo Nordisk Health Care AG.
- **Anne Helene Olsen** is employed at Novo Nordisk A/S as a Senior Epidemiologist in the Real World Science department in Søborg, Denmark.
- **Navid Nedjatian** is employed at Novo Nordisk Health Care AG as a Global Medical Director in Medical Affairs department in Zurich, Switzerland.

This presentation explores statistical methods and approaches towards missing data from the following publication:



Reduced CV risk with long-term GH replacement in AGHD: data from two large observational studies

in Endocrine Connections

Authors: Charlotte Höybye , Beverly M K Biller, Jean-Marc Ferran, Murray B Gordon, Nicky Kelepouris, Navid Nedjatian, Anne H Olsen, and Matthias M Weber

View Less —

Correspondence should be addressed to C Höybye: charlotte.hoybye@regionstockholm.se

DOI: <https://doi.org/10.1530/EC-22-0267>

Article Type: Research Article

Volume/Issue: Volume 12: Issue 1

Online Publication Date: 13 Dec 2022

Article ID: e220267

Copyright: © The authors 2022

Agenda

- **Background**

- NordiNet[®] IOS and ANSWER Program overview
- Adult Growth Hormone Deficiency (AGHD) and Cardiovascular Risk

- **Handling missing data for our analyses**

- How to analyse CV risk and selection of "Brunner" Score?
- How to build a comparator group?
- How to ensure the data to be analysed does not define a specific population?
- How to ensure handling of missing data does not introduce bias?

- **Conclusion**

NordiNet[®] IOS and ANSWER Program overview

Non-interventional, multicentre studies evaluating long-term effectiveness and safety of Norditropin[®] (somatropin)* as prescribed by treating physicians in a real-life clinical setting¹

NordiNet[®] IOS (NCT00960128)

- In 23 countries (469 clinics)
- Paediatric and adult patients
- 2524 adults²

ANSWER Program (NCT01009905)

- In the USA (207 clinics)
- Paediatric and adult patients
- 966 adults²

Pooled effectiveness analysis set (n=1275)

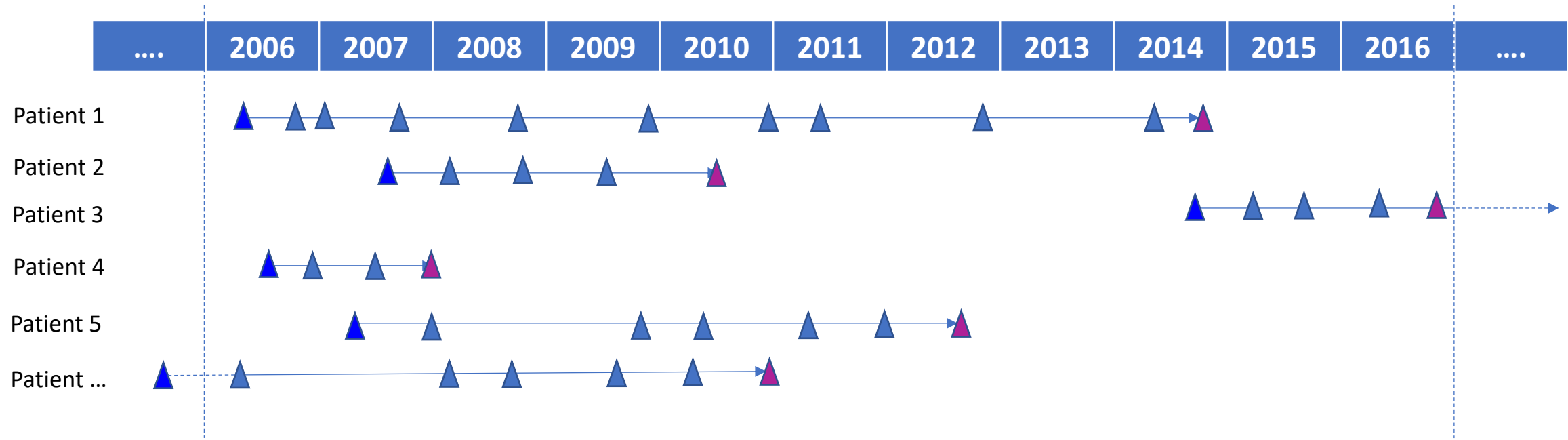
*Novo Nordisk A/S, Denmark

Effectiveness analysis set included patients who were GH-naïve at baseline and who received their first GH dose at age ≥20 years, for whom valid body mass index, age and dosing information was available at baseline

AGHD, adult growth hormone deficiency; IOS, International Outcome Study

1. Höybye et al. Clin Epidemiol 2013;5:119–27;
2. Weber et al. Growth Horm IGF Res 2020;50:71–82

NordiNet[®] IOS and ANSWER Program Design



Average Duration: 4.5 years for Effectiveness Set

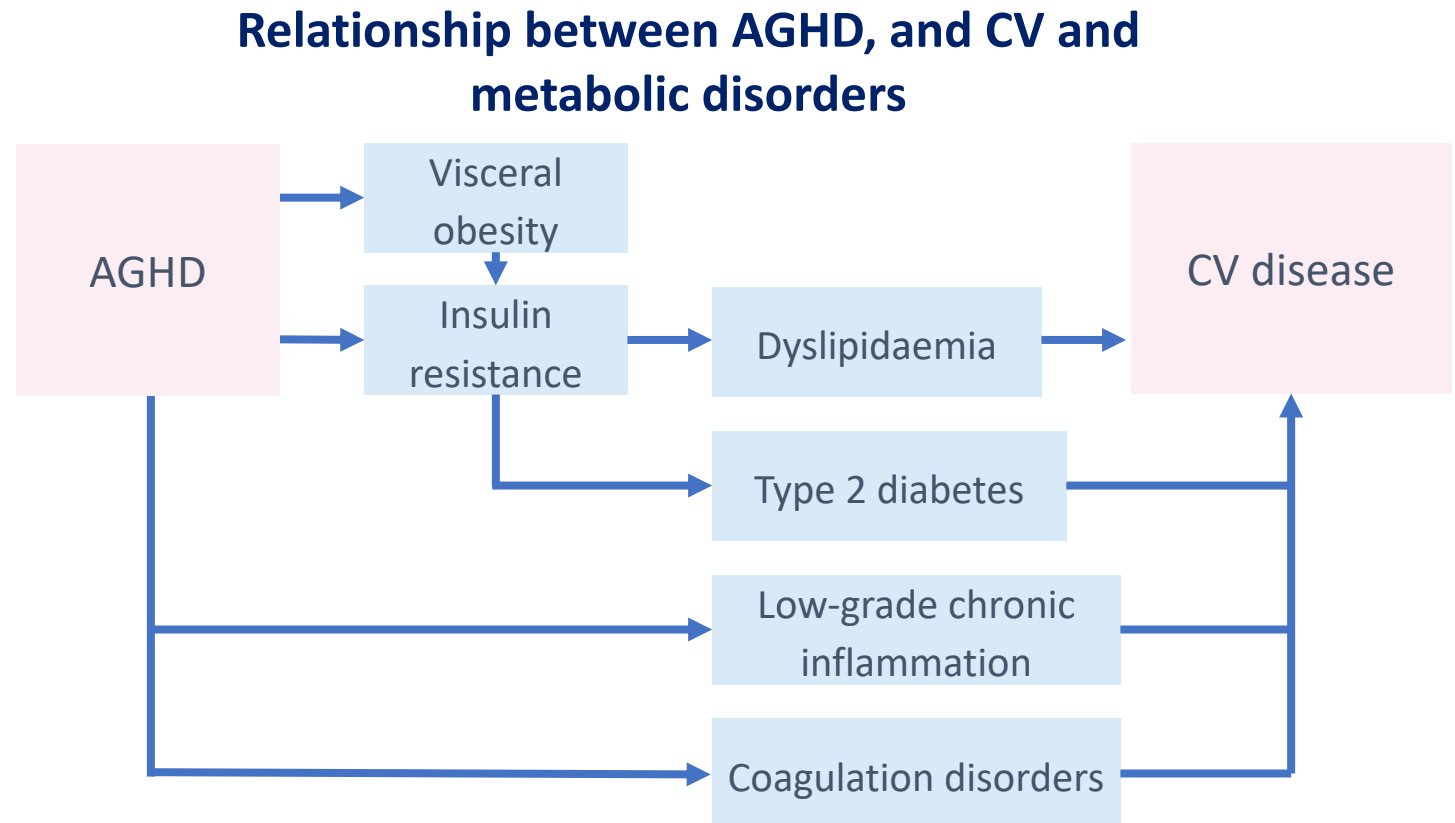
- ▲ Baseline: GH treatment start, Demographics, Baseline Measures, Medical History, Concomitant Medication, etc.
- ▲ Follow-up: GH treatment update, Follow-up measures, Adverse Events, Concomitant Medication, etc.
- ▲ Last visit: GH treatment update, Follow-up measures, Adverse Events, Concomitant Medication, Discontinuation Reason *if not lost to follow-up*, etc.

AGHD and CV disease

- Evidence suggests that adults with GHD exhibit metabolic syndrome and are at risk for type 2 diabetes

Characteristics of AGHD

- Alterations in body composition
- Decreased capacity for exercise
- Impaired quality of life
- Unfavourable risk factor profile for CV disease, and diabetes
 - Typical metabolic syndrome features and risk for type 2 diabetes



GH replacement therapy is the standard of care in treating AGHD

- Long-term GH replacement therapy offers improvements in:
 - ✓ Body composition: reduced fat mass, increased lean mass and muscle strength
 - ✓ Bone mineral density
 - ✓ Health-related quality of life
 - ✓ CV risk markers: increased HDL-C, reductions in LDL-C, C-reactive protein, diastolic blood pressure and carotid intima-media thickness
 - ✓ CV disease: reduced rate of myocardial infarction
- Concerns remain about effects on glucose metabolism and insulin sensitivity, i.e. metabolic syndrome features and type 2 diabetes risk factors

Study objective: to evaluate CV disease risk in patients with AGHD who received GH replacement therapy for up to 10 years as part of the NordiNet[®] IOS and ANSWER Program

How to analyse Cardiovascular Risk?

Analyze occurrence of CV events?

- AE under-reporting and too short follow-up periods in our studies does not make this approach viable...

Use a CV Risk Score?

- CV Risk scores* usually use Race and/or require full availability of different risk factors...
- Race not collected and high rate of missing data for smoking status in our data...

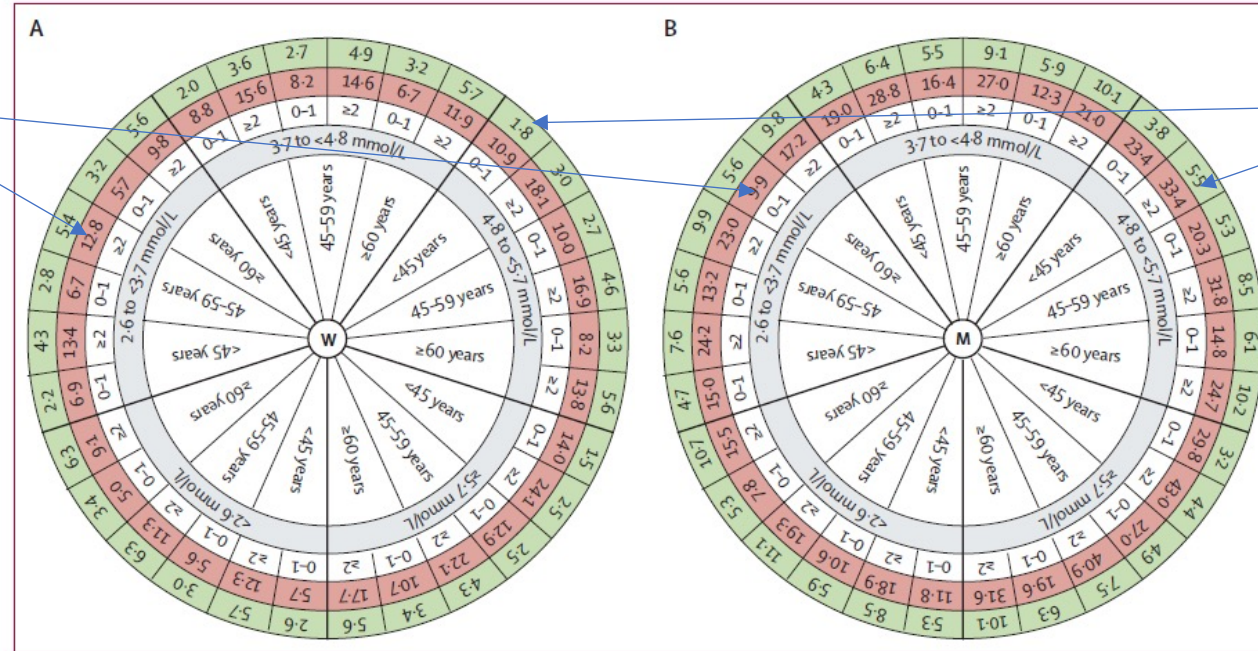
Use a feasible CV Risk Score?

- Brunner et al. (2019) published a CV risk score based on non-HDL cholesterol and an aggregated risk factor index...
- *But missing data and no obvious comparator to start with...*

*: ASCVD Risk Estimator: 2013 ACC/AHA Guideline on the Assessment of Cardiovascular Risk, Goff et al. (2014)
Comparisons of established risk prediction models for cardiovascular disease: systematic review, Siontis et al. (2012)

Individual risk of fatal or non-fatal cardiovascular disease by Age of 75 years – *The “Brunner Score”*

The probability (%) of cardiovascular disease by the age of 75 years.



The hypothetically achievable probability (%) for cardiovascular disease by the age of 75 years after 50% reduction of non-HDL cholesterol.

Figure 4: Model of long-term cardiovascular disease risk prediction and the benefit of lipid reduction

Individual risk of fatal or non-fatal cardiovascular disease in women (A) and men (B) according to age, non-HDL cholesterol concentration, and the number of additional cardiovascular risk factors (daily smoking, arterial hypertension, diabetes, and obesity; white circle). The red circle represents the probability (%) of cardiovascular disease by the age of 75 years. The hypothetically achievable probability (%) for cardiovascular disease by the age of 75 years after 50% reduction of non-HDL cholesterol is given in the green circle. Corresponding 95% CIs are provided in the appendix (p 37).

Individual risk of fatal or non-fatal cardiovascular disease by Age of 75 years – *The “Brunner Score”*

Gender		Women											
2 levels		non-HDL-cholesterol category											
		<2.6 mmol/L		2.6-<3.7 mmol/L		3.7-<4.8 mmol/L		4.8-<5.7 mmol/L		≥5.7 mmol/L		Non-HDL Cholesterol	
Risk Factors		Age category	Risk factors	Risk	Therapy	Risk	Therapy	Risk	Therapy	Risk	Therapy	5 levels	
Diabetes Hypertension Obesity Daily Smoking	<45 years	0-1	0-1	5.7 (5.4, 6.1)	2.6 (2.4, 2.8)	6.9 (6.7, 7.3)	2.2 (2.2, 2.3)	8.8 (8.5, 9.1)	2.0 (2.0, 2.1)	10.9 (10.6, 11.3)	1.8 (1.8, 1.9)	14.0 (13.4, 14.7)	1.5 (1.5, 1.6)
			≥2	12.3 (11.1, 14.0)	5.7 (5.2, 6.6)	13.4 (12.8, 14.3)	4.3 (4.1, 4.6)	15.6 (14.9, 16.6)	3.6 (3.4, 3.8)	18.1 (17.1, 19.2)	3.0 (2.9, 3.2)	24.1 (22.6, 25.8)	2.5 (2.4, 2.7)
	45-59 years	0-1	0-1	5.6 (5.3, 6.0)	3.0 (2.9, 3.2)	6.7 (6.5, 7.0)	2.8 (2.7, 2.9)	8.2 (8.1, 8.5)	2.7 (2.6, 2.8)	10.0 (9.8, 10.3)	2.7 (2.6, 2.7)	12.9 (12.5, 13.3)	2.5 (2.4, 2.6)
			≥2	11.3 (10.5, 12.3)	6.3 (5.8, 6.8)	12.8 (12.4, 13.6)	5.4 (5.3, 5.8)	14.6 (14.2, 15.2)	4.9 (4.8, 5.1)	16.9 (16.5, 17.5)	4.6 (4.5, 4.8)	22.1 (21.4, 23.0)	4.3 (4.1, 4.4)
	≥60 years	0-1	0-1	5.0 (4.5, 5.7)	3.4 (3.1, 3.9)	5.7 (5.5, 6.0)	3.2 (3.1, 3.4)	6.7 (6.5, 6.9)	3.2 (3.1, 3.3)	8.2 (7.9, 8.4)	3.3 (3.3, 3.5)	10.7 (10.3, 11.1)	3.4 (3.3, 3.5)
			≥2	9.1 (7.9, 10.4)	6.3 (5.5, 7.2)	9.8 (9.3, 10.5)	5.6 (5.3, 6.0)	11.9 (11.5, 12.5)	5.7 (5.5, 6.0)	13.8 (13.3, 14.4)	5.6 (5.4, 5.9)	17.7 (17.1, 18.4)	5.6 (5.4, 5.9)

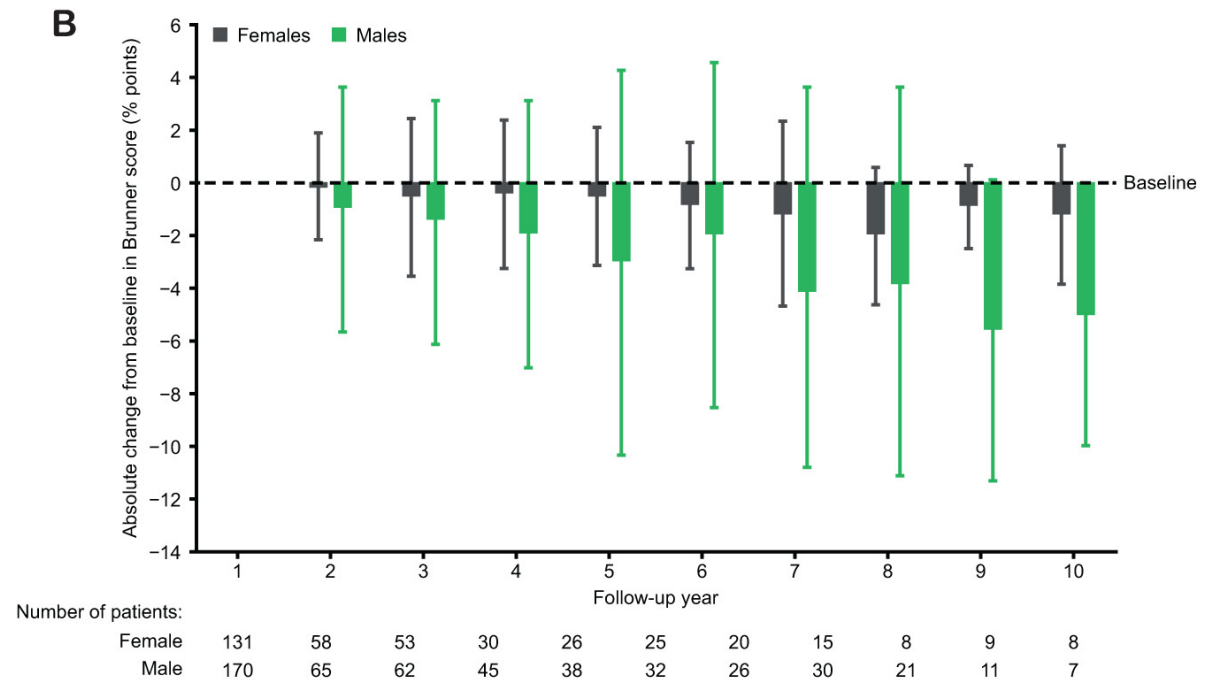
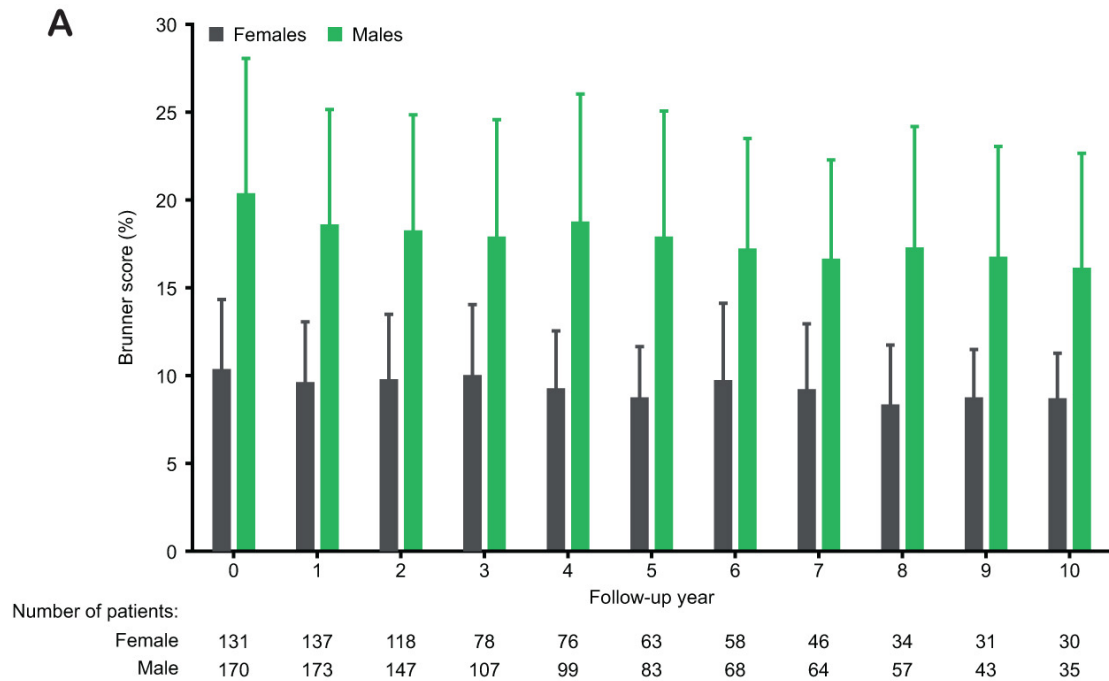
Brunner risk score (%) - safety population, naïve patients only

Cross-sectional data

Year	N	Mean	SD	Med	p10	p90	Patients with baseline and follow up data n
0	278	18.64	8.93	16.65	6.90	31.80	—
1	285	15.95	7.82	14.60	6.90	27.00	110
2	257	15.84	7.74	14.00	6.90	27.00	103
3	172	15.85	7.28	15.00	8.20	27.00	61
4	179	15.81	7.99	13.20	6.90	27.00	54
5	140	15.01	7.80	13.30	6.90	27.00	47
6	124	15.16	7.26	13.40	6.70	27.00	46
7	109	14.48	6.65	13.40	6.70	24.20	37
8	88	15.56	8.13	15.00	6.70	28.80	26
9	68	14.88	6.85	14.00	6.70	24.20	18
10	61	14.35	7.02	12.90	6.70	23.00	14

Brunner risk score (%) - safety population, naïve patients only

Cross-Sectional (A) and Longitudinal (B) data by Gender



How to build a comparator group of Untreated AGHD patients?

Propensity Score
Matching

- Not viable as all our patients are treated...

Use of an external
database?

- No database of untreated AGHD patients was available...
- Comparison with general population is not defensible as AGHD patients have an increased CV risk...

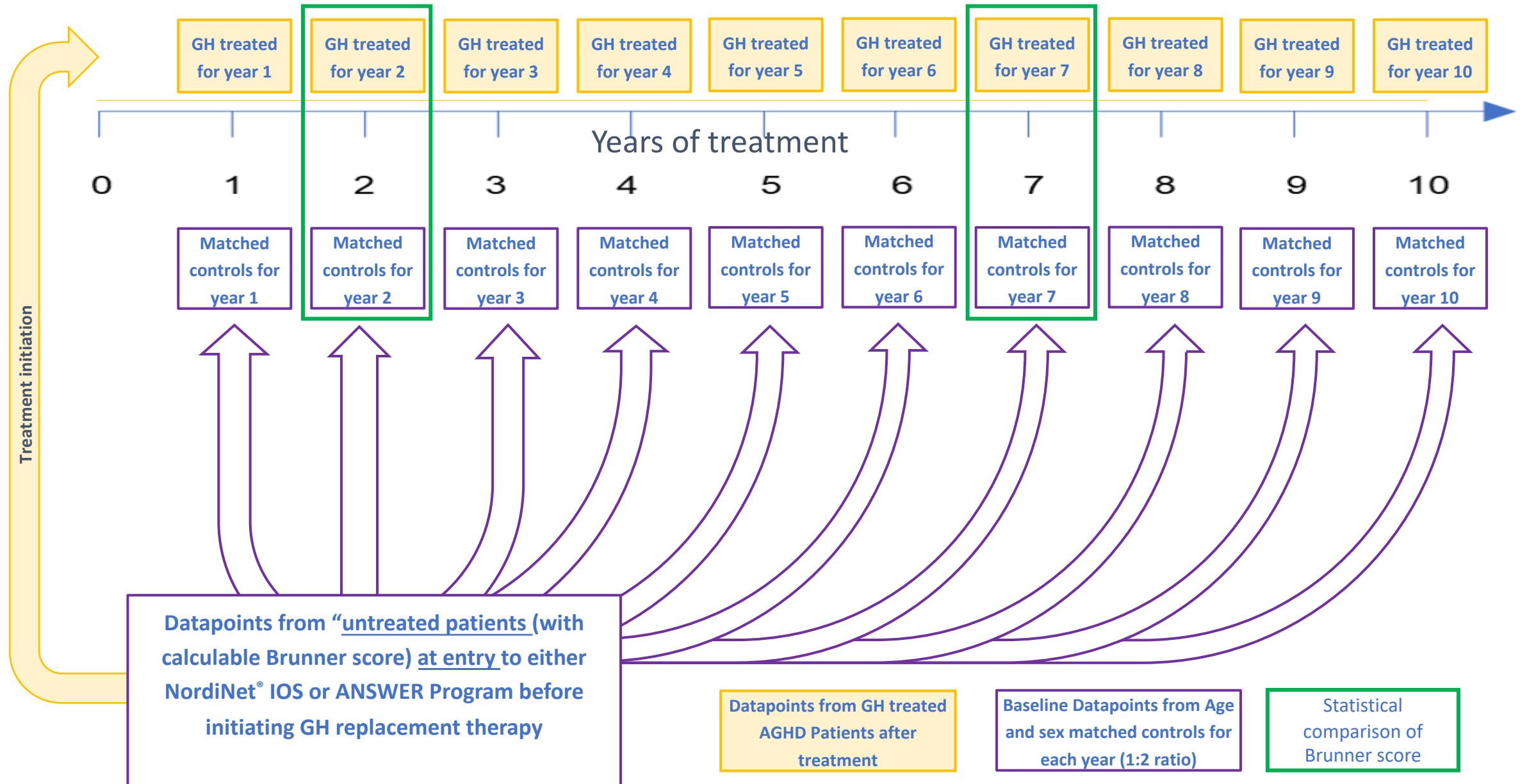
Solely use change
from baseline?

- Would not replace a comparator group...
- And Brunner score decreases with age at equal non-HDL and risk levels...

Build a
comparator group
from our own
study population?

- Possible using our GH-naïve/untreated patients at baseline who are also AGHD...

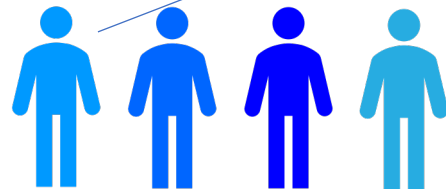
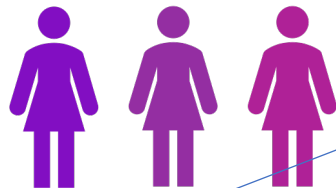
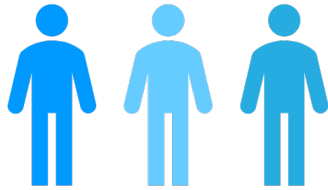
Build a comparator group of Untreated AGHD patients from our own study population.



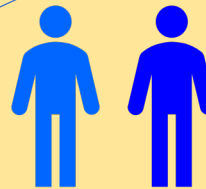
1:2 Ratio "Greedy" Random Matching

Patients
at Year N

Naïve Patients a baseline
matched on Age and Gender



Final Set of Matched patients for Year N



Brunner Risk Score

GH-treated vs. Matched control groups

Table 3 Brunner risk score.

GH-treated patients				Matched control group ^a			GH-treated patients with baseline and follow-up data		
Years of treatment	<i>n</i>	Brunner risk score, mean %	Brunner risk score, s.d. %	<i>n</i>	Brunner risk score, mean %	Brunner risk score, s.d. %	<i>n</i>	Relative change from baseline, mean %	Relative change from baseline, s.d. %
0	301	16.04	8.02	–	–	–	–	–	–
1	310	14.66	6.97	263	16.30	7.96	123	–1.26	21.34
2^b	265	14.51	6.90	285	16.15	8.07	115	–3.54	23.24
3	185	14.59	6.90	253	16.71	8.18	75	–4.11	25.27
4	175	14.66	7.52	251	16.60	8.35	64	–5.58	27.18
5	146	13.95	7.30	224	16.39	7.30	57	–6.09	25.30
6	126	13.78	6.61	193	16.19	7.57	46	–12.40	26.07
7^b	110	13.53	6.13	178	16.81	8.18	45	–14.40	27.65
8	91	13.97	7.24	153	16.61	8.37	29	–19.10	21.31
9	74	13.40	6.48	132	16.51	8.72	20	–15.00	19.14
10	65	12.69	6.30	109	16.27	7.69	15	–4.77	31.74

The risk of having CV disease by age 75 years was calculated using the Multinational CV Risk Consortium model, as described by Brunner *et al.* (14). A lower number/score indicates lower risk of CV disease.

^aAge- and sex-matched untreated patients at study enrolment of NordiNet[®] IOS or the ANSWER Program before initiating GH treatment

^bStatistical comparison was only performed between treated and matched untreated patients at Year 2 and Year 7.

CV, cardiovascular; GH, growth hormone; IOS, International Outcome Study; *n*, number of participants.

- Sensitivity Analysis: To assess the possible impact of the random matching algorithm on the results, the matching was repeated 20 times for years 2 and 7 and consistency of results evaluated

Are patients with 7 years of data not specifics?

- Each population with a specific number of years of follow-up may be considered as not fully representative of the general AGHD population.
- We have tried to limit the potential impact of this by comparing treated patients with age- and sex-matched controls both at shorter and longer-term follow-up.
 - We planned statistical comparisons at Year 2 and Year 7, which were before and after the mean follow-up year of the original studies (4.5 years), to address the concern of a possible selection bias for the patients with a follow-up duration longer than that of the mean.

Comparison of GH-treated vs. Matched control groups

Table 2 Patient characteristics of patients included in statistical comparison of GH-treated and matched control patients at Year 2 and Year 7.

	Year 2		Year 7	
	GH treated	Matched control group ^a	GH treated	Matched control group ^a
Number of patients, <i>n</i>	265	285	110	178
Female, <i>n</i> (%)	118 (44.5)	126 (44.2)	46 (41.8)	72 (40.5)
Mean age, years (s.d.)	49.5 (13.7)	47.6 (13.8)	52.9 (12.9)	51.7 (12.3)
Mean GH dose, mg/day (s.d.)	0.32 (0.22) (<i>n</i> = 258)	–	0.36 (0.17) (<i>n</i> = 104)	–
Mean IGF-I SDS	0.45 (1.33) (<i>n</i> = 260)	–1.25 (1.46) (<i>n</i> = 274)	0.42 (1.14) (<i>n</i> = 109)	–1.05 (1.40) (<i>n</i> = 172)
Mean BMI, kg/m ² (s.d.)	29.2 (6.4) (<i>n</i> = 259)	28.5 (5.9) (<i>n</i> = 285)	29.2 (6.8) (<i>n</i> = 107)	29.0 (5.6) (<i>n</i> = 178)
Onset of GHD, <i>n</i> (%)				
Childhood	12 (4.5)	16 (5.6)	7 (6.4)	4 (2.3)
Adult	253 (95.5)	269 (94.4)	103 (93.6)	174 (97.8)
Patients with risk factors, <i>n</i> (%)				
Obesity	96 (36.2)	88 (30.9)	33 (30.0)	61 (34.3)
Missing	6 (2.3)	0 (0.0)	3 (2.7)	0 (0.0)
Hypertension	74 (27.9)	77 (27.0)	34 (30.9)	55 (30.9)
Missing	8 (3.0)	4 (1.4)	4 (3.6)	1 (0.6)
Diabetes	33 (12.5)	37 (13.0)	15 (13.6)	24 (13.5)
Missing	6 (2.3)	10 (3.5)	3 (2.7)	9 (5.1)
Identified as a daily smoker	16 (6.0)	3 (1.1)	7 (6.4)	2 (1.1)
Missing	71 (26.8)	260 (91.2)	19 (17.3)	162 (91.0)
Non-HDL cholesterol category, <i>n</i> (%)				
<2.6 mmol/L	16 (6.0)	18 (6.3)	7 (6.4)	10 (5.6)
2.6 to >3.7 mmol/L	92 (34.7)	74 (26.0)	49 (44.6)	39 (21.9)
3.7 to >4.8 mmol/L	99 (37.4)	96 (33.7)	35 (31.8)	60 (33.7)
4.8 to >5.7 mmol/L	36 (13.6)	61 (21.4)	15 (13.6)	39 (21.9)
≥5.7 mmol/L	22 (8.3)	36 (12.6)	4 (3.6)	30 (16.9)

Patients were identified as hypertensive when their SBP was ≥140 mmHg or when their SBP was <140 mmHg with reported concomitant use of an antihypertensive medication. Patients were considered positive for daily smoking if they were identified as a regular smoker in the lifestyle section of the study case reports. Patients were identified as having diabetes when their FPG was ≥126 mg/dL, or their HbA1c was ≥47.5 mmol/mol (6.5%) or when either FPG or HbA1c was below the defined threshold with reported concomitant use of antidiabetic medications. Patients were identified as obese if their BMI was ≥30 kg/m².

^aAge- and sex-matched untreated patients at study enrolment of NordiNet[®] IOS or the ANSWER Program before initiating GH treatment.

BMI, body mass index; FPG, fasting plasma glucose; GH, growth hormone; GHD, growth hormone deficiency; HbA1c, glycated haemoglobin; HDL, high-density lipoprotein; IGF-I, insulin-like growth factor-I; IOS, International Outcome Study; *n*, number of participants; SBP, systolic blood pressure.

~90% of missing data on Smoking Status at baseline

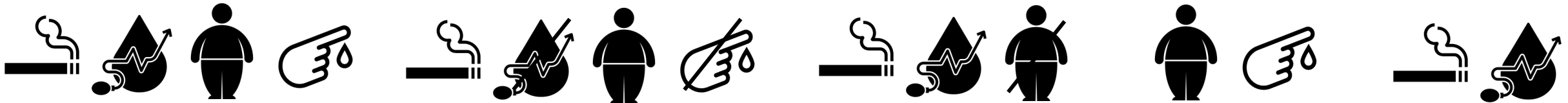
Issue with high rate of missing data at baseline?

Was our baseline worst as a result? (1)

- There are only two Risk levels based on following risk markers:
 - Diabetes, Obesity, Smoking Status and Hypertension:
 - 0-1 risk marker:



- 2-4 risk markers:



Achieving "0-1" level requires at least 3 data point while "2-4" requires only 2....

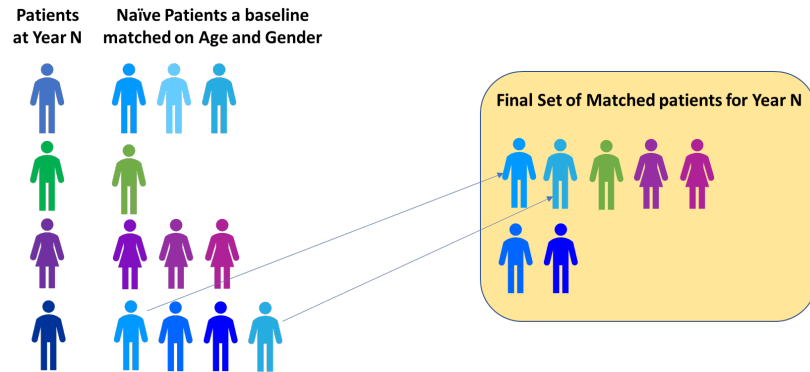
Issue with high rate of missing data at baseline?

Was our baseline worst as a result? (2)

Method

Matching algorithm rerun with:

- (1) Only patients with 3 risk markers available
- (2) Only patients with 4 risk markers available



Results

Comparison:

- (1) At year 2: **Significant** difference in same direction
At year 7: **Significant** difference in same direction
- (2) At year 2: **Close to Significant** difference in same direction
At year 7: **Significant** difference in same direction

The results of the sensitivity analyses showed similar trends of reduced CV risk in GH-treated patients compared with matched control groups.

Conclusions

- “Love the data you get” ...It may not be perfect but with a critical mindset and adequate measures, hurdles can be overcome.
 - The cross-functional effort we have had between Medical Advisors, Epidemiologists, Data Scientists and KOLs was instrumental in progressing with confidence on this manuscript.
 - Every step of the analyses was challenged and numerous sensitivity analyses were carried out to check if any bias was introduced along the way.
- Limitations:
 - The data regarding the **time from diagnosis of AGHD to starting GH replacement therapy are not captured for all patients** within NordiNet® IOS and the ANSWER Program and, accordingly, accurate duration of the disease could not be matched between the GH-treated and control groups.
 - **Results could not be reproduced using other CV Risk Scores** due to Race not being collected and the high rate of missing data in certain risk factors