

# The Health Gym



Kuo NIH, Polizzotto MN, Finfer S, Jorm L, Barbieri S, et al

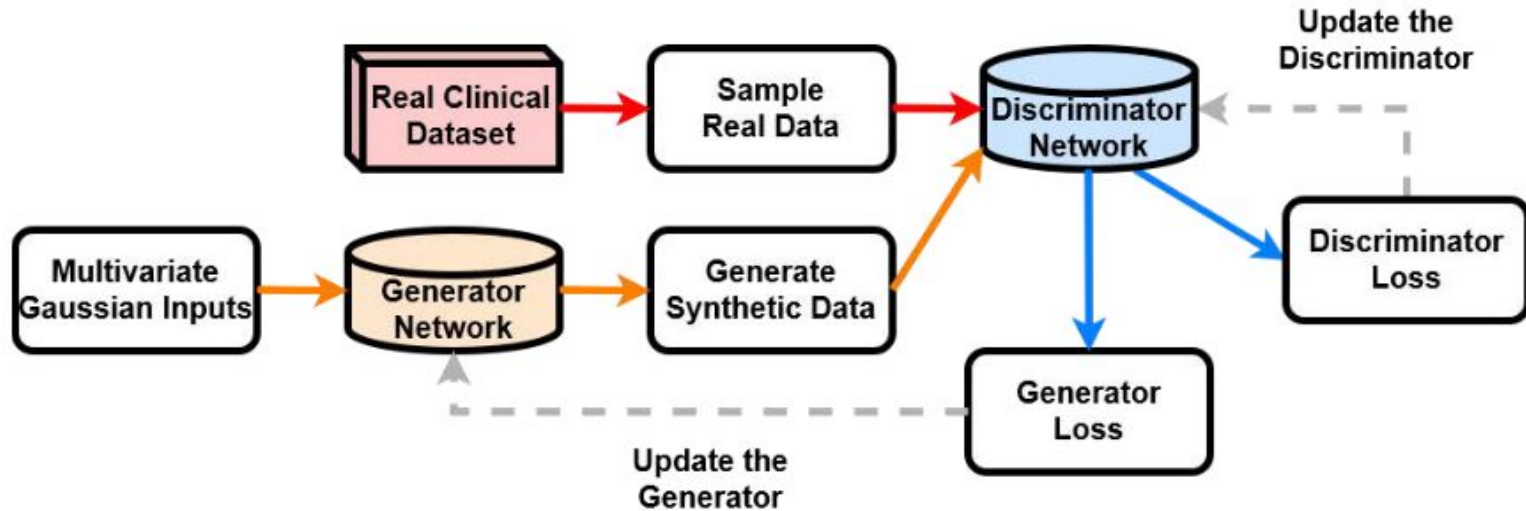


HealthGym.ai



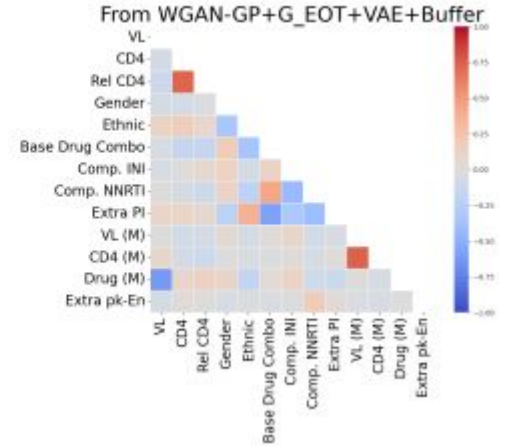
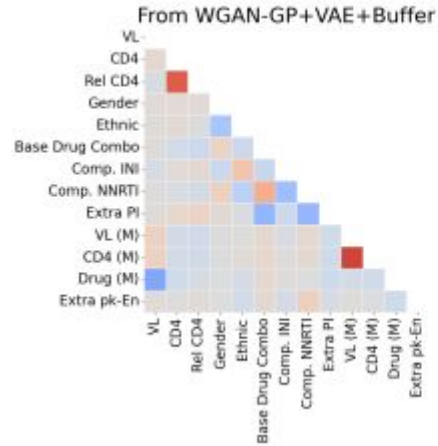
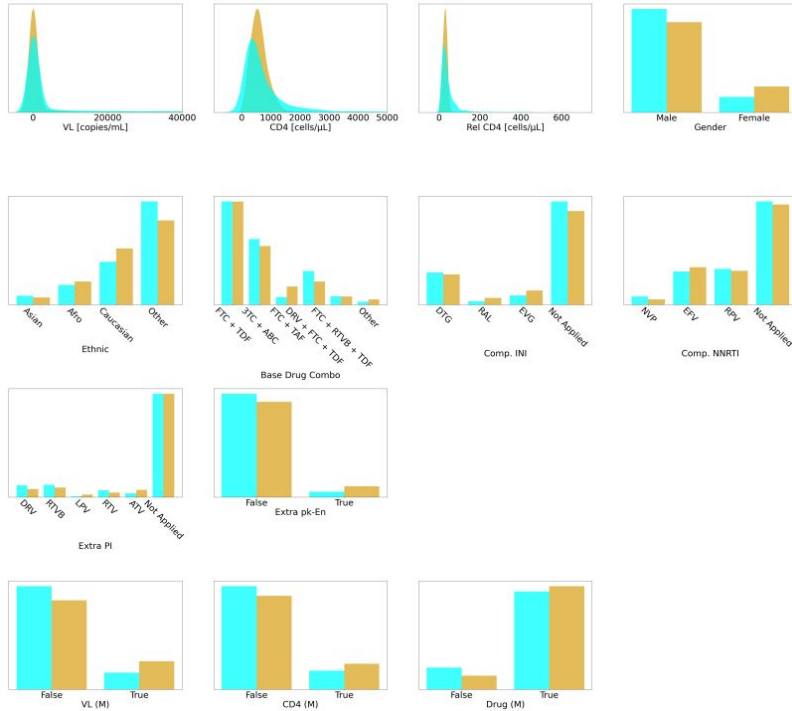
Kuo, Nicholas I-Hsien, et al. "The Health Gym: synthetic health-related datasets for the development of reinforcement learning algorithms." *Scientific Data* 9.1 (2022): 693.

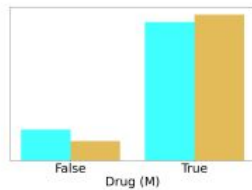
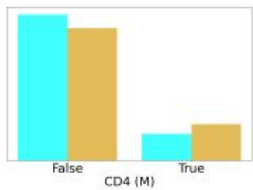
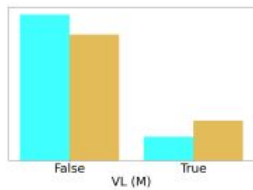
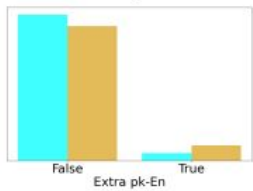
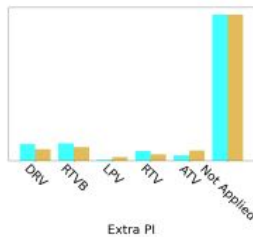
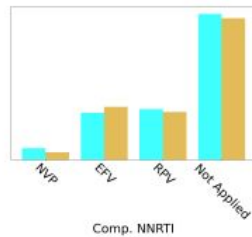
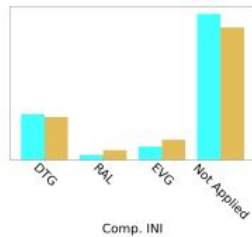
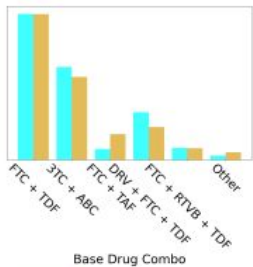
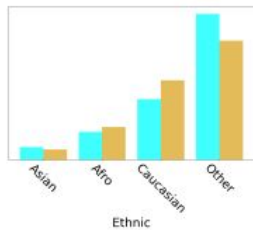
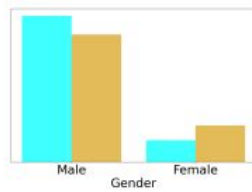
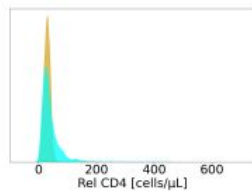
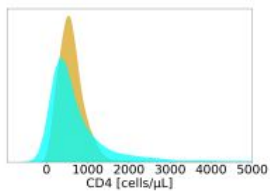
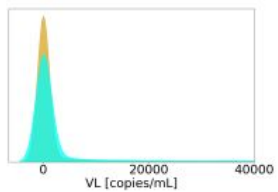
# Generative Adversarial Networks



(a) An Overview

# Antiretroviral Therapy for HIV (ART for HIV)







# Why ART for HIV?

[https://www.youtube.com/watch?v=cSNaBui2IM8&ab\\_channel=ITPCGlobal](https://www.youtube.com/watch?v=cSNaBui2IM8&ab_channel=ITPCGlobal)



# Why is ART for HIV difficult?

[https://www.youtube.com/watch?v=VCVjHSuYqto&ab\\_channel=WorldHealthOrganization%28WHO%29](https://www.youtube.com/watch?v=VCVjHSuYqto&ab_channel=WorldHealthOrganization%28WHO%29)



# ART for HIV

Number of Patients: 8,916

Time-Series: 60 units per patient

Mixed-Type Variables:

Numeric: Viral Load (VL), Cluster of Differentiation 4 (CD4)

Binary: Gender,  
Pharmacokinetic Enhancer (pk-En)  
VL Measured (VL (M)), CD4 (M), Drug (M)

Categorical: Ethnicity,  
Base Drug Combination (Base Drug Combo),  
Complementary INI (Comp. INI),  
Comp. NNRTI, Extra PI



Variable Name	Data Type	Unit	Valid Categorical Options
Viral Load (VL)	numeric	copies/mL	- -
Absolute Count for CD4 (CD4)	numeric	cells/ $\mu$ L	- -
Relative Count for CD4 (Rel CD4)	numeric	cells/ $\mu$ L	- -
Gender	binary	- -	Male; Female
Ethnicity (Ethnic)	categorical	- -	Asian; African; Caucasian; Other
Base Drug Combination (Base Drug Combo)	categorical	- -	FTC + TDF; 3TC + ABC; FTC + TAF; DRV + FTC + TDF; FTC + RTVB + TDF; Other
Complementary INI (Comp. INI)	categorical	- -	DTG; RAL EVG; Not Applied
Complementary NNRTI (Comp. NNRTI)	categorical	- -	NVP; EFV RPV; Not Applied
Extra PI	categorical	- -	DRV; RTVB; LPV; RTV; ATV; Not Applied
Extra pk Enhancer (Extra pk-En)	binary	- -	False; True
VL Measured (VL (M))	binary	- -	False; True
CD4 (M)	binary	- -	False; True
Drug Recorded (Drug (M))	binary	- -	False; True






# ART for HIV





 - Patient demographics;  - Medication combinations

 - Data measured (opposite of missingness)

50 medication combinations spanning 21 medications:

Numeric: Viral Load (VL), Cluster of Differentiation 4 (CD4)

Binary:  Gender,  
 Pharmacokinetic Enhancer (pk-En)  
 VL Measured (VL (M)),  CD4 (M),  Drug (M)

Categorical:  Ethnicity,  
 Base Drug Combination (Base Drug Combo),  
 Complementary INI (Comp. INI),  
 Comp. NNRTI, Extra PI

We selected a cohort of 8,916 people (with 332,800 records) from the EuResist database (Zazzi et al., 2012) using published inclusion/exclusion criteria (Kuo et al., 2022). There are 3 numeric, 5 binary, and 5 categorical variables, as listed in Table 1. The numeric variables – VL, CD4, and Rel CD4 – are indicative of the patient’s health status. The HIV treatment regimens are deconstructed as Base Drug Combo, Complimentary (Comp.) INI, Comp. NNRTI, Extra PI, and Extra pk-En. The medication classes are: nucleoside reverse transcriptase inhibitors (NRTIs), nucleotide reverse transcriptase inhibitors (NtRTIs), non-nucleotide reverse transcriptase inhibitors (NNRTIs), integrase inhibitor (INI), and protease inhibitors (PIs). The base drug combo mainly comprises NRTIs + NtRTIs; see Tang et al. (2012) for more details on the individual medications. There are 50 medication combinations in the real dataset spanning 21 medications.

The original EuResist database also contains a considerable proportion of missing data. Such information can be highly informative, indicating *e.g.*, the need for specific laboratory tests; and hence we include the variables with suffix (M) to indicate **if measurements are taken**. Measurements are taken (*i.e.*, entry with True) 24.27% for VL (M), 22.21% for CD4 (M), and 85.13% for Drug (M) in the real dataset  $\mathcal{D}_{\text{real}}$ .

# Data organisation

Stored as a Comma-Separated Values (CSV) file

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	VL	CD4	Rel CD4	Gender	Ethnic	Base Drug Combo	Comp. INI	Comp. NNRTI	Extra PI	Extra pk-En	VL (M)	CD4 (M)	Drug (M)	PatientID	Timestep
2	29.944271	793.4583	30.834505	1.0	3.0	0.0	0.0	3.0	5.0	0.0	0.0	1.0	1.0	0	0
3	29.24198	467.4189	30.35598	1.0	3.0	0.0	0.0	3.0	5.0	0.0	0.0	0.0	1.0	0	1
4	28.748991	465.12485	30.40532	1.0	3.0	0.0	0.0	3.0	5.0	0.0	0.0	0.0	1.0	0	2
5	28.101835	692.0069	30.248816	1.0	3.0	0.0	0.0	3.0	5.0	0.0	0.0	0.0	1.0	0	3
6	28.813837	641.75714	29.944712	1.0	3.0	0.0	0.0	3.0	5.0	0.0	0.0	0.0	1.0	0	4
7	28.033257	444.34845	30.647573	1.0	3.0	0.0	0.0	3.0	5.0	0.0	0.0	0.0	1.0	0	5
8	27.235563	650.4474	30.108711	1.0	3.0	0.0	0.0	3.0	5.0	0.0	0.0	0.0	1.0	0	6
9	26.021847	745.2554	30.728205	1.0	3.0	0.0	0.0	3.0	5.0	0.0	0.0	1.0	1.0	0	7
10	26.717731	436.72992	30.792599	1.0	3.0	0.0	0.0	3.0	5.0	0.0	0.0	0.0	1.0	0	8



```
# Gender
#       1 = Male, 2 = Female
# Ethnic
#       1 = Asian, 2 = Afro, 3 = Caucasian, 4 = Other
# Base Drug Combo
#       0 = FTC + TDF
#       1 = 3TC + ABC
#       2 = FTC + TAF
#       3 = DRV + FTC + TDF
#       4 = FTC + RTVB + TDF
#       5 = Other
# Comp. INI
#       0 = DTG
#       1 = RAL
#       2 = EVG
#       3 = Not Applied
# Comp. NNRTI
#       0 = NVP
#       1 = EFV
#       2 = RPV
#       3 = Not Applied
```

```
# Extra PI
#       0 = DRV
#       1 = RTVB
#       2 = LPV
#       3 = RTV
#       4 = ATV
#       5 = Not Applied
# Extra pk-En
#       0 = False
#       1 = True
# VL (M)
#       0 = False
#       1 = True
# CD4 (M)
#       0 = False
#       1 = True
# Drug (M)
#       0 = False
#       1 = True
```

# Questions?

:)