Advanced Analytics for Data-Driven Decision-Making and Action

Dr. Silviu Tomescu

Analytics, Right to Care (Group)

April 13, 2023

ASAP II

ACCELERATING SUPPORT TO ADVANCED LOCAL PARTNERS II

A FEW QUICK NOTES

1. Welcome Local Partners – **tell us where you're from in the chat.**
2. Please use the **Q&A box to ask any questions** and the chat box for answering questions asked by the presenters.
3. We have a few **polls** during the webinar today. They will pop up on your screen.
4. The presentation for today's webinar will be saved on ASAP's website at **www.intrahealth.org/asap-resources**

Rapidly prepare Local Partners to have the capabilities and resources to serve as Prime Partners for USAID/PEPFAR programming, in compliance with USAID and PEPFAR procedures, for PEPFAR program implementation in FY 2022 and 2023.

70% of USAID PEPFAR funding to local prime partners.

STRATEGIC OBJECTIVES

1. Strengthen Local Partners as they transition to receive PEPFAR funding as a USAID Prime Partner to comply with regulations.

2. Prepare Local Partners to directly manage, implement, and monitor PEPFAR programs, and maintain consistent PEPFAR program achievement and quality.

ASAP II

ACCELERATING SUPPORT TO ADVANCED LOCAL PARTNERS II

KEY RESULTS from ASAP I & II

ASAP has supported
126 local organizations
in **18** countries

113 local partner organizations

13 local government partners

ASAP II-SUPPORTED COUNTRIES

Angola

Cameroon

Côte d'Ivoire

DRC

eSwatini

Ethiopia

Lesotho

Malawi

Namibia

Nigeria

South Sudan

Uganda

Zimbabwe

ASAP I additional countries:

Kenya

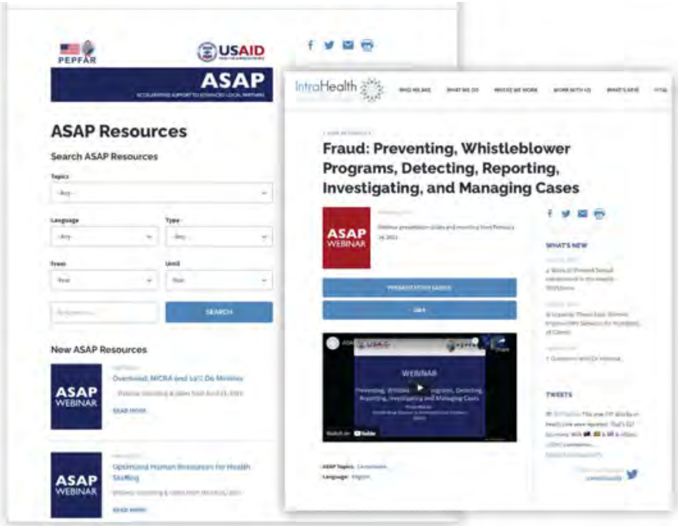
Mozambique

South Africa

Tanzania

Zambia

18 TOTAL COUNTRIES



ON-DEMAND WEBINARS

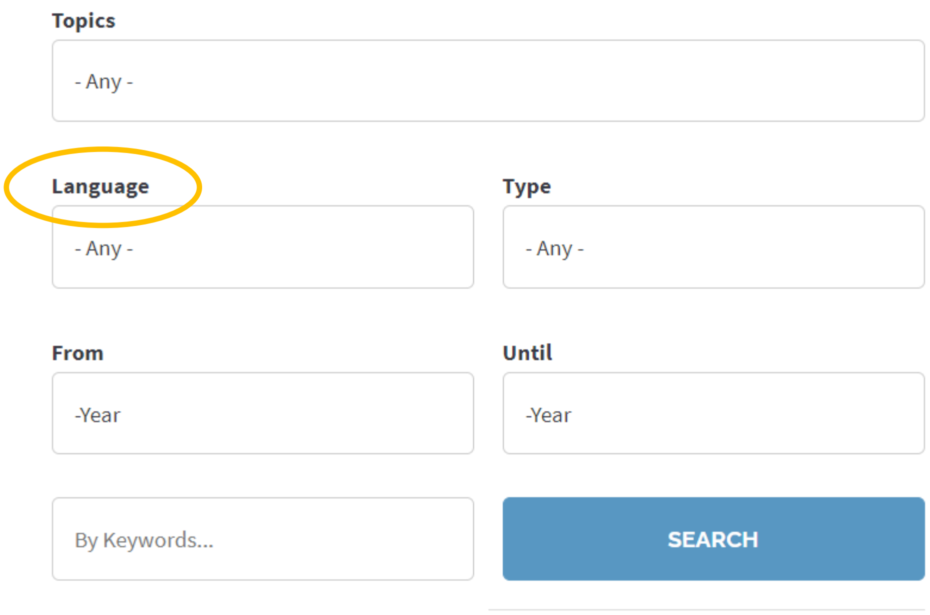
USAID/ASAP has broadcasted **89 webinars** for more than **20,000 attendees** in **76 countries**.

Find past webinars on ASAP's web page **www.intrahealth.org/asap-resources**

AVAILABLE IN 3 LANGUAGES

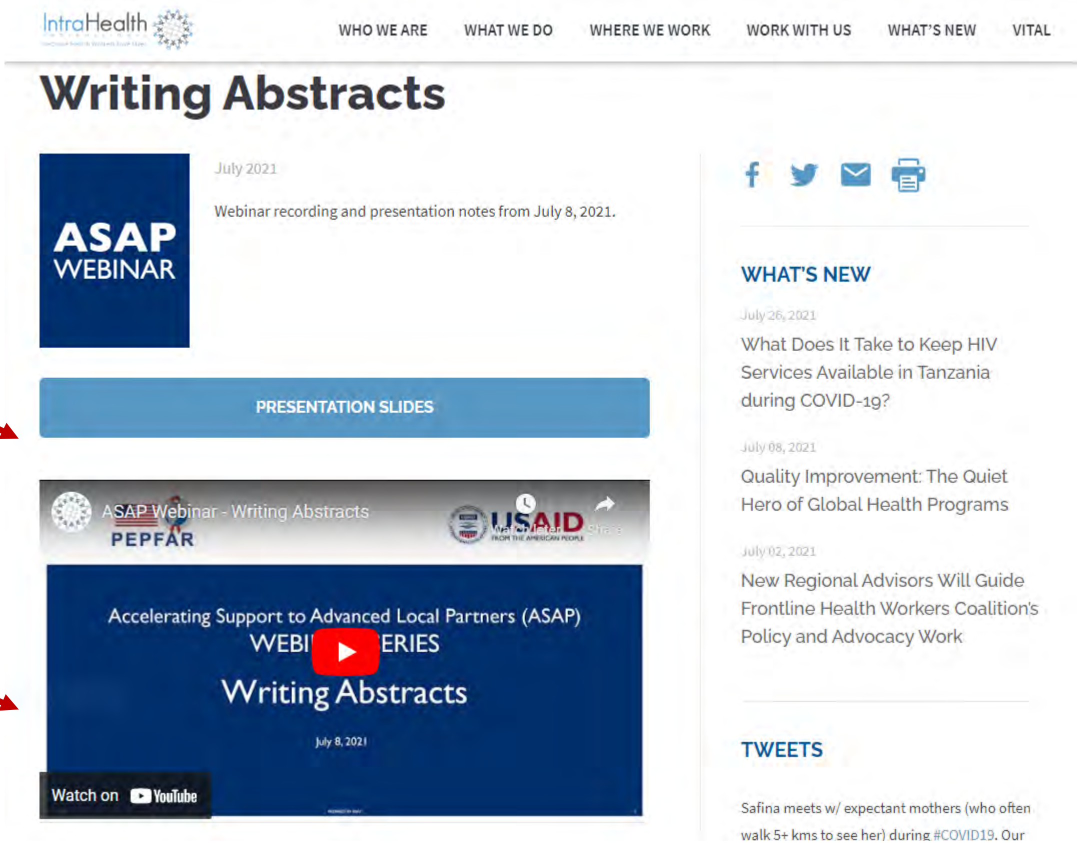
Choose your **language or topic.**

Featuring webinars in **French, English, and Portuguese.**



Download a pdf
of the presentation.

Watch a recording
of the webinar.



UPCOMING USAID/ASAP II WEBINARS

Path to Prime Webinar Series

Practical Application GIS Methods
& Tools to Guide Spatial Targeting & Micro-Planning

English-language webinar - ***April 20***

www.intrahealth.org/upcoming-asap-webinars

TODAY'S PRESENTER

Dr. Silviu Tomescu

Analytics

Right to Care, a South Africa-based USAID Local Partner and ASAP II Consortium Partner

What will be covered

- Overview of Data Analytics
  - Descriptive
  - Advanced (Predictive)
- Abstraction of modelling for predictions
- Data Process Overview
- Requirements for Advanced Analytics and Benefits
- Translating Models into Action (Examples)
- Case Study
- Demo

Why the need for Data Analytics: Value add to programs

- PEPFAR "ensures that every dollar is optimally focused for impact through data-driven policies"
- Data Analytics is not a nice to have, it is a need for more effective budget allocation, planning and programming

Data Analytics

Generate insight and guidance using descriptive, diagnostic, predictive, and prescriptive analytics

- **Descriptive [Dashboard, Slides, Reports]**
  - What happened in the past?: Monitor trend/pattern of program performance
  - Why did it happen (diagnostic / deep-dive / advanced descriptive analytics)
  - Where are adjustments or resources needed

- **Advanced [Predictive and Prescriptive Analytics (Modelling)]**
  - Identify causal factors and contribution (e.g., to IIT, VL non-suppression, HIV positivity)
  - Useful when there are many factors/explanatory variables available to analyse:
    - E.g. age, sex, location, distance from facility, employment, education level, side effects, co-morbidities, mental disorders, average number of days late, number of times late/IIT, cohabiting/marital status, religion, transport
  - Good at explaining the combination of reasons why something may be happening.
    - The why can then be translated into action

Descriptive Analytics: automated reporting on MER/non-MER indicators

- All programs are judged/scored based on achieving a particular set of indicators – MER Guideline
  - Two levels of MER:
    - 1: Non-composite
      - HTS_TST, HTS_TST_POS, TX_NEW, TX_CURR
    - 2: Composite: Need calculation
      - Yield = HTS_TST_POS/HTS_TST * 100
      - Linkage = TX_NEW/HTS_TST_POS * 100
      - TX_PVLS = # Suppressed ART patients / # Patients with a VL result documented within last 12 months * 100

- Required vs not required daily/weekly/monthly
  - Indicators not required give insights into program gaps
    - E.g. Required: TX_PVLS (%) is required. Not required: Gaps in the underlining indicator, how many of those on ART were tested (the denominator).

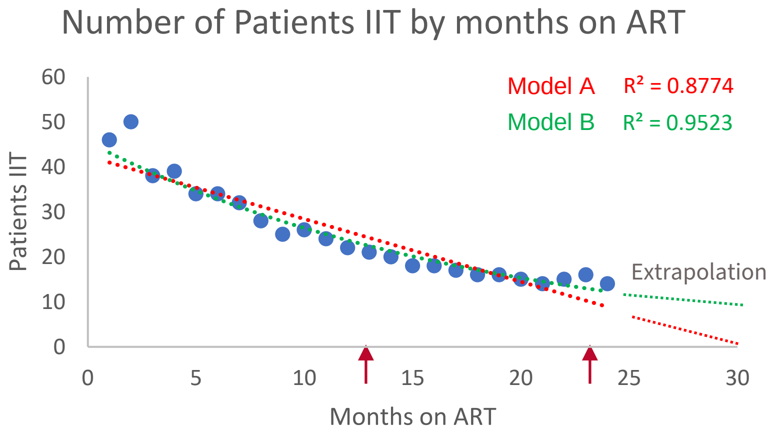
- Descriptive analytics and high frequency reporting (HFR)
  - Creates better communication
  - Daily situation rooms – forum to discuss daily/weekly with different stakeholders
  - Alarm IPs on the problem – highlights first level of danger around performance (e.g. 7% Yield last week vs 3% Yield this week).
  - Can be automated by one person
    - Reduce error
    - Reduce staffing
    - Static or interactive (dashboard) reports

Predictive/Advanced Analytics: What is it?

- Predictive analytics makes use of statistical models
- The model on the right has one feature (Months on ART) which can be plotted in a 2D plot. However, models can have a virtually infinite number of features that may not be as easy to depict by conventional means.

- How well does a model fit the data
  - Model selection (try different models and pick the best fit – generally) impacts the ability to accurately predict an outcome
    - For time series, extrapolation becomes necessary (can be heavily impacted by model choice)
  - There are other considerations regarding modelling, many of which can be domain specific.
    - For example, 6 MMD is prescribed to VL suppressed patients. Therefore, 6 MMD may correlate well with a suppressed VL, but it is not independent of outcome (correlation does not imply causation)

Example of a predictive model with one dependent (# Patients IIT) and one independent (Months on ART) variable.



A model can result in overestimation, underestimation of values, or false prediction of outcomes.

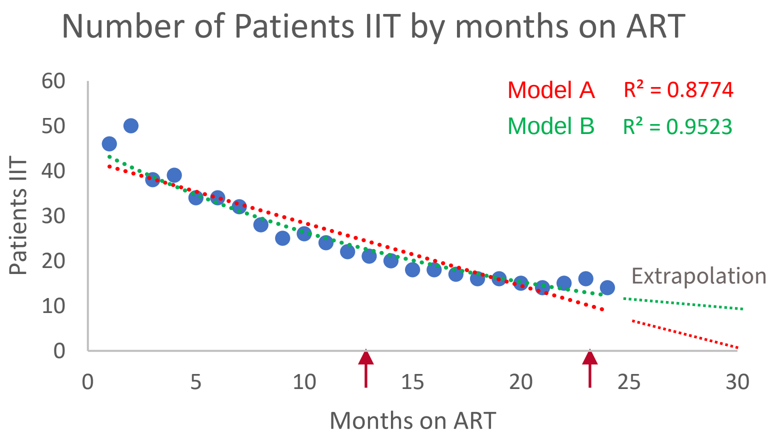
Question: Select the correct answer

A) Model A underestimates at 13 months on ART and overestimates at 23 months on ART

B) Model A overestimates at 13 months on ART and underestimates at 23 months on ART

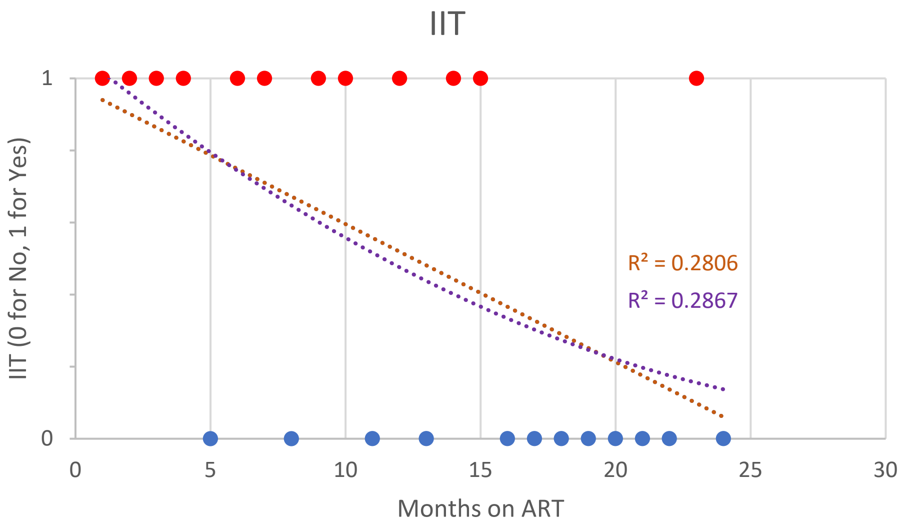
C) Model B is not as good as model A

D) Model A is better than model B for extrapolation



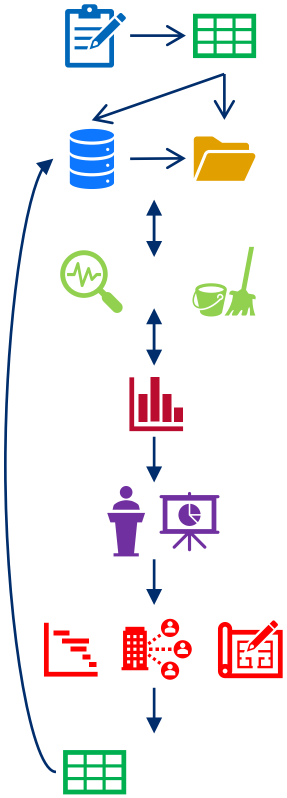
Classification model

- For a classification model (binary outcome – Yes/No)
- Numeric values are assigned
  - 0 for No
  - 1 for Yes
- The threshold for prediction will impact how many we predict correctly/incorrectly



Questions and Answers Session

Data – Process Overview and Considerations



Data Process (Stage)	Considerations
Collection / Capture	Paper records are scanned and accessible for data quality assurance (DQA) and verification by users.
Storage	Raw data is unmodified. Save modifications as a processed dataset. [Folders: Data (Raw; Processed), Analyses, Output (Figures; Reports)]
Data check and cleaning	Logically guided. Justifiable. (e.g., VL test before ART start: test done before ART, TX_NEW > HTS_POS: test done elsewhere)
Analytics (Descriptive, Predictive)	Skills and time. A dedicated analyst my be needed for predictive analytics and implementation thereof
Interpretation and dissemination	Data interpretation skills needed for program managers and implementors
Planning and implementation	Deployment of resources or tools to maximise impact.
Validation (Prescriptive)	Monitoring the effect of interventions. Data needs to be captured, stored, and analysed to validate impact.

What is Required for Advanced Analytics, and what are the Benefits?

Requirements:

- Skilled people – to do statistical analytics, descriptive and predictive analytics, principal component analysis, dimensionality reduction
- Capacitation on interpretation of analyses
- Patient level data
- Larger dataset
  - More observations i.e., patient records and
  - More explanatory variables (e.g. demographics, location, education, etc.)

Benefits: Data for action frameworks (what is the action to be taken)

- Get insights into a key problem
  - Learn more about why is a problem occurring – to plan cost effective intervention
- Leverage analytics into action for impacting programs
- Tailored (to individuals) targeted intervention to address the problems
- Impact evaluation – short term and long-term impact (refine and improve intervention)
- Standardise effective interventions across all programs
- Improve patient outcomes

Advanced Data Analytics for Action

Can be applied on any part of the HIV cascade/program

- HIV screening (when resources are limited for mass testing)
- IIT (predict clients at risk of defaulting for pre-emptive intervention)
- VL suppression (to identify clients that are likely not adhering well to treatment in the absence of routinely updated VL test results)
- Any other outcome that is monitored

Considerations:

- Some statistical models will perform better than others, additional effort may be required. Though, lesser models may be good enough to predict individuals with very high risk.
- A better model may not be needed if resources are insufficient to, for example offer EAC to all those at risk of IIT
    - Logistic regression – possibly the most widely used model

Examples

General intervention

- Wanted to know if 6 months MMD could improve patient adherence

Multimonth dispensing of up to 6 months of antiretroviral therapy in Malawi and Zambia (INTERVAL): a cluster-randomised, non-blinded, non-inferiority trial

- Influenced expansion of 6 months MMD in Malawi and Zambia

Hoffman, R. M., Moyo, C., Balakasi, K. T., Siwale, Z., Hubbard, J., Bardon, A., ... & Rosen, S. (2021). Multimonth dispensing of up to 6 months of antiretroviral therapy in Malawi and Zambia (INTERVAL): a cluster-randomised, non-blinded, non-inferiority trial. *The Lancet Global Health*, *9*(5), e628-e638.

Targeted intervention

- APACE program wanted to predict those that would miss appointment or be virally non-suppressed

Applying machine learning and predictive modeling to retention and viral suppression in South African HIV treatment cohorts

- Predictive models to anticipate patients that would become IIT or virally non-suppressed
- Paved way for pre-emptive HIV programming

Maskew, M., Sharpey-Schafer, K., De Voux, L., Crompton, T., Bor, J., Rennick, M., ... & Pisa, P. (2022). Applying machine learning and predictive modeling to retention and viral suppression in South African HIV treatment cohorts. *Scientific reports*, *12*(1), 1-10.

- Cost effective ways of running a program
- More insight leads to more action and better patient outcomes
- Targeted programming and intervention
- Creates accountability for money spent
- Evaluate impact of intervention
- Up and down scaling
- Is it worth to continue a program?

Questions and Answers Session

Case study:

Problem?

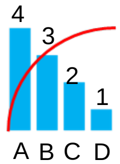
- Chief of party flags IIT

- How do we approach this challenge?
  - Conduct descriptive analysis to investigate the occurrence of IIT
  - Set up a dashboard
  - Identifying groups that are most likely to IIT for pre-emptive intervention (i.e., counselling before IIT happens)

Dashboard / Descriptive Analytics

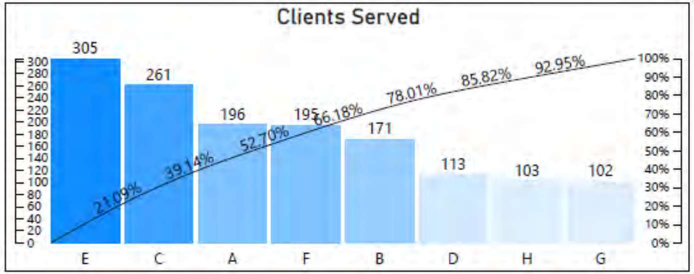
Power BI

Off the shelf and free to use



- Distribution of IIT
- Prioritisation of effort (Pareto principle)
  - Assumes that "vital few" causes (~ 20%) account for 80% of an outcome (e.g. 20% of facilities account for 80% of IIT)
  - Useful in guiding targeted intervention to optimise best use of resources
  - Identify and target groups (facilities, demographic groups, etc) for intervention
- Strategies can be formulated based on the observations
  - Guide program implementation

Priority Facilities

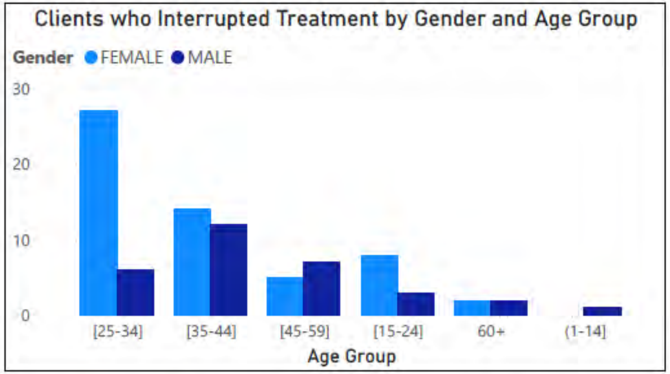


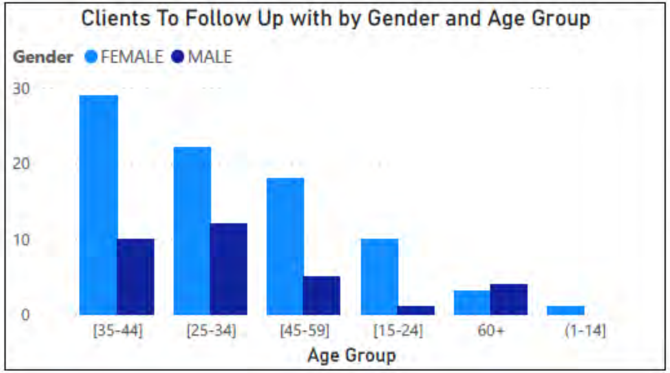
Resources are sometimes limited. While we can guide all facilities to follow up better on some particular clients (age groups, gender, etc), it may be necessary to prioritise certain facilities to maximise impact while minimising effort. Pareto's 80:20 rule (guideline).

Facility: All
Gender: All

- Knowing which facilities have the most clients may guide deployment of limited resources
- Knowing which facilities have the largest number of clients IIT:
  - Facilities C and E
    - Can identify facilities where enhanced adherence counselling (EAC) may be needed
  - To follow up with (facility E)
    - Can identify which facilities could use a call centre or a client-reminder SMS platform
- Does this explain what are the characteristics (age, sex, years on treatment, etc.) of clients that are associated with IIT?
  - No. You may have some idea but it should be investigated
    - That is – conduct descriptive data analytics to understand why or what may contribute to high IIT/TFO

Target Groups and Strategic Considerations





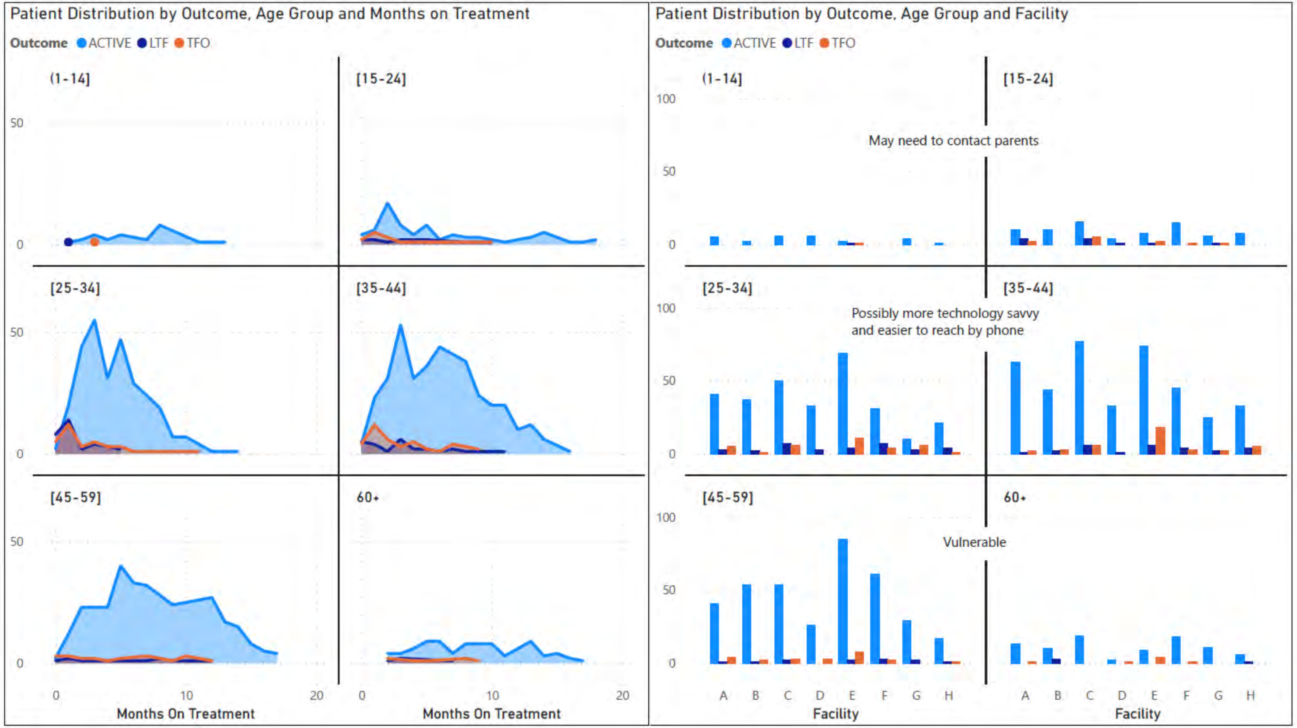
- Let's look at a few more variables
  - Age group and sex
  - Now we can see that most IIT/TFO clients are females aged 25-44
  - We can recall that facilities C and E did have the highest number of IIT/TFO
  - And that IIT/TFO occur most in clients under 1 year of treatment

Question: What happens if someone is aged 45-59?

A) Assume that they have the same IIT profile as other groups
B) Assume a low probability of IIT as identified in the descriptive analysis for the age group
C) Further analyse across other variables to investigate IIT for the age group

- If we only have 4 explanatory variables (sex, age, facility and years on treatment), then we can look at the other 3.

Some Deep-Dive Analytics



- Deep dive reveals that the duration on ART does not affect IIT for the 45-59 age group, rather facility E seems to be somehow causative of IIT of 45-59.
- But how do we keep track of all those variables (and others)?

Odds and odds ratios

- Applies to logistic regressions
- Odds = probability an event occurs / probability an event does not occur
- Odds = (IIT/Total) / (Active/Total) = IIT/Active
- Odds ratio = Odds of a group / Odds of reference group
- For example:

Location	Active	IIT	Odds	Odds Ratio
X	4	1	1/4 = 0.25	0.25/0.166 = 1.5
Y (ref)	6	1	1/6 = 0.166	0.166/0.166 = 1 (ref)

Questions and Answers Session

Odds and odds ratios

Outcome		Active	IIT	OR (univariable)	OR (multivariable)
Facility	E	246 (93.5)	17 (6.5)	-	-
	A	170 (92.4)	14 (7.6)	1.19 (0.56-2.48, p=0.640)	1.48 (0.68-3.22, p=0.319)
	B	150 (94.9)	8 (5.1)	0.77 (0.31-1.78, p=0.557)	0.99 (0.38-2.41, p=0.982)
	C	213 (91.0)	21 (9.0)	1.43 (0.73-2.81, p=0.295)	2.14 (1.06-4.40, p=0.035)
	D	92 (91.1)	9 (8.9)	1.42 (0.58-3.22, p=0.419)	1.80 (0.70-4.37, p=0.205)
	F	170 (95.0)	9 (5.0)	0.77 (0.32-1.72, p=0.530)	0.98 (0.39-2.31, p=0.963)
	G	83 (90.2)	9 (9.8)	1.57 (0.65-3.58, p=0.296)	2.43 (0.95-5.91, p=0.055)
	H	86 (87.8)	12 (12.2)	2.02 (0.91-4.37, p=0.077)	2.38 (1.01-5.45, p=0.042)
Gender	FEMALE	835 (93.0)	63 (7.0)	-	-
	MALE	375 (91.2)	36 (8.8)	1.27 (0.82-1.94, p=0.269)	1.45 (0.90-2.31, p=0.123)
AgeGroup	35-44	416 (92.7)	33 (7.3)	-	-
	1-14	26 (96.3)	1 (3.7)	0.48 (0.03-2.40, p=0.484)	0.38 (0.02-2.06, p=0.366)
	15-24	81 (89.0)	10 (11.0)	____ (0.70-3.18, p=0.245)	1.17 (0.51-2.49, p=0.700)
	25-34	260 (87.2)	38 (12.8)	1.84 (1.13-3.02, p=0.015)	1.37 (0.80-2.35, p=0.247)
	45-59	345 (96.6)	12 (3.4)	0.44 (0.21-0.84, p=0.017)	0.48 (0.23-0.94, p=0.040)
	60+	82 (94.3)	5 (5.7)	0.77 (0.26-1.86, p=0.595)	0.97 (0.32-2.45, p=0.956)
YearsOnTreatment	5+	731 (96.4)	27 (3.6)	-	-
	[0-1)	29 (63.0)	17 (37.0)	15.87 (7.73-32.30, p<0.001)	13.96 (6.52-29.71, p<0.001)
	[1-2)	78 (79.6)	20 (20.4)	6.94 (3.69-12.91, p<0.001)	6.20 (3.19-11.96, p<0.001)
	[2-3)	145 (90.6)	15 (9.4)	2.80 (1.42-5.33, p=0.002)	2.41 (1.20-4.69, p=0.011)
	[3-5)	227 (91.9)	20 (8.1)	2.39 (1.30-4.32, p=0.004)	2.07 (1.11-3.81, p=0.020)

Odds ratio for facility A
= (14/170)/(17/246) = 1.19

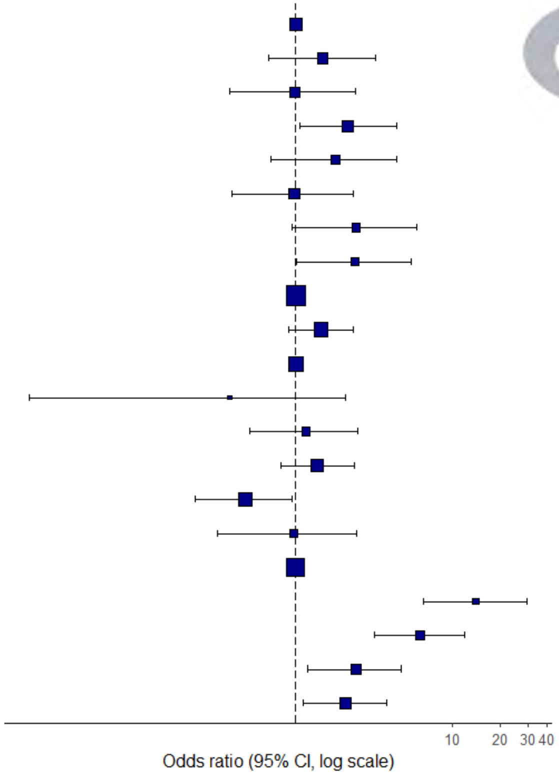
Question: What is the unadjusted / univariable odds ratio for the 15-24 age group (using 35-44 as the reference group)?

A) 0.98
B) 1.32
C) 1.56
D) 2.43

Data Analytics Findings - Factors associated with IIT

Can accommodate a virtually infinite number of explanatory variables and observations (number of patients) granted sufficient computational resources.

		Outcome: OR (95% CI, p-value)
Facility	E	-
	A	1.48 (0.68-3.22, p=0.319)
	B	0.99 (0.38-2.41, p=0.982)
	C	2.14 (1.06-4.40, p=0.035)
	D	1.80 (0.70-4.37, p=0.205)
	F	0.98 (0.39-2.31, p=0.963)
	G	2.43 (0.95-5.91, p=0.055)
	H	2.38 (1.01-5.45, p=0.042)
Gender	FEMALE	-
	MALE	1.45 (0.90-2.31, p=0.123)
AgeGroup	35-44	-
	1-14	0.38 (0.02-2.06, p=0.366)
	15-24	1.17 (0.51-2.49, p=0.700)
	25-34	1.37 (0.80-2.35, p=0.247)
	45-59	0.48 (0.23-0.94, p=0.040)
	60+	0.97 (0.32-2.45, p=0.956)
YearsOnTreatment	5+	-
	[0-1)	13.96 (6.52-29.71, p<0.001)
	[1-2)	6.20 (3.19-11.96, p<0.001)
	[2-3)	2.41 (1.20-4.69, p=0.011)
	[3-5)	2.07 (1.11-3.81, p=0.020)



Programming language

Odds ratios make a model practically useful without intensive IT infrastructure. In most cases, the counsellors are the limitation, not the lack of clients we can send for EAC. Therefore, why invest in the effort to detect all of them when we can easily focus on clients very likely to IIT.

Turning findings into action

Let's say our solution to IIT is EAC.

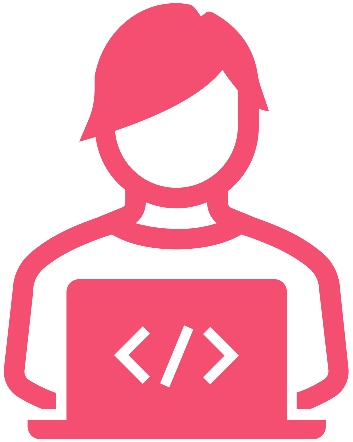
Score	0	1	2	3
Gender	Female	Male		
Age Group	1-14, 35-44, 45+	15-34		
Years on Treatment	5+	2+	1-2	0-1

Aggregated Score (IIT Risk)		
0-2	3-4	5

Question: You can only counsel one person, who would it be?

- A: Female , 45+, 1-2 years on treatment?
- B: Female, 15-34, 0-1 years on treatment?
- C: Male , 15-34, 0-1 years on treatment

Demonstration Time!



Conclusions and recommendations

- The demo here provides a very basic approach to multivariable analysis that could be a rough starting point
- Modelling (machine learning) will take time to investigate properly and produce the best model
- This tutorial aimed to provide some abstraction for the benefit of applying advanced analytics for data-driven decision making and action
- To make the most of advanced analytics, dedicated data analyst(s) / scientist(s) would be needed

ASAP II

ACCELERATING SUPPORT TO ADVANCED LOCAL PARTNERS II

Questions?

This publication is made possible by the support of the American people through the United States Agency for International Development (USAID) and the President's Emergency Plan for AIDS Relief (PEPFAR). The contents are the sole responsibility of IntraHealth International and do not necessarily reflect the views of USAID or the United States Government.

PEPFAR

USAID FROM THE AMERICAN PEOPLE