

---

# Nouvelles approches de mesure de la rétention dans les files actives VIH

---

Grégoire Lurton

## 1 Contexte et Justification

Ce projet s'inscrit dans le cadre d'un projet de recherche doctorale sur des méthodes innovantes d'analyse des données de systèmes d'information sanitaire dans les pays à faible ressources. Plus précisément, nous étudions comment adapter la définition des indicateurs et les méthodes de collecte et d'analyse de données pour les adapter à des situations locales ne correspondant pas aux conditions des systèmes statistiques des pays riches.

La mesure de la rétention des patients sous ARV est une mesure clé de la performance des programmes VIH. La chronicité du VIH rend en effet nécessaire une mesure du succès sur le long terme, et la persistance dans le soin des patients est un élément essentiel de cette évaluation. Néanmoins, la mesure de la persistance des patients dans le système de soin est loin d'être triviale, et est largement impactée par les conditions de production des services de soin.

A l'initiation des programmes de traitement du VIH, les professionnels du VIH ont choisi d'utiliser des indicateurs déjà définis et utilisés dans le cadre d'études cliniques, et ont choisi de mesurer la rétention des patients dans le soin par la proportion de patients qui cessent de venir à leurs rendez-vous de suivi[1, 2, 3]. . La définition de la catégorie Perdu de Vue (PDV) provient donc de programmes pouvant s'appuyer sur des systèmes de collecte de données très performants et mettant en œuvre un suivi des patients renforcé. Dans le cadre de programmes de santé publique, cette définition est néanmoins problématique, car elle repose sur une hypothèse de complétude et de qualité des données patients qui est rarement respectée. En définissant les PDV par une absence de données plus que par un élément positif, les différentes définitions des PDV utilisées mettent en effet dans une même catégorie des patients ayant interrompu leur suivi, et des patients dont les visites n'ont pas été correctement enregistrées.

De plus, en basant l'évaluation des programmes sur la catégorisation de certains patients, la mesure d'évaluation a modifié les conditions de soin et les relations entre soignants et soignés [4].

De plus, au fur et à mesure que les programmes VIH étaient mis en œuvre à travers le monde, il est apparu que cette catégorie était peu spécifique et regroupait des patients dans des situations diverses[5, 6, 7]. Cette faible spécificité pose des problèmes importants pour l'évaluation des programmes de prise en charge du VIH, et plusieurs auteurs ont proposé des méthodes pour corriger la mesure de la rétention. Ces méthodes ont pu se baser sur la collecte de données additionnelles [8, 9, 10], ou sur la mise en œuvre de modèles de correction [11, 12, 13, 14, 15]. Plus récemment, d'autres auteurs ont exploré comment la définition des PDV utilisée par différents programmes peut avoir un impact important sur la mesure de la rétention[16, 17, 18, 19, 20, 21, 22].

Une hypothèse sous-tendant ces définitions de la rétention est que, si un patient se rend à un rendez-vous, cette visite sera enregistrée correctement dans le système de données correspondant. Il y a donc une équivalence entre la mesure de la rétention et la mesure de la qualité des données dans les centres de santé. Cependant, l'évaluation de cette hypothèse est rarement faite, et son impact sur la mesure du nombre de PDV dans un centre est difficile.

Avec le développement et la mise en œuvre croissante de systèmes de dossiers patients informatisés dans les centres de santé, nous disposons d'une opportunité pour améliorer cette situation. La plupart des systèmes informatisés de collecte de données enregistrent en effet des métadonnées sur la saisie des données, comme leur date d'entrée ou le rôle de la personne saisissant les données. Ces métadonnées peuvent être mobilisées pour mesurer la qualité des données, et pour comprendre l'impact de la qualité des données sur la mesure de la rétention.

## 2 Objectif de l'étude

Le projet décrit ici cherche à simuler et à décrire l'impact de la qualité des données collectées dans les centres de santé sur une mesure de la rétention axée sur la catégorie PDV. Cet objectif sera atteint à travers trois objectifs spécifiques :

**Impact de la qualité des données sur la mesure de la rétention** Nous modéliserons le chemin par lequel la qualité des données affecte la mesure de la rétention, et nous simulerons la mesure des résultats d'une cohorte de prise en charge du VIH en fonction de la qualité de ses données.

**Maturité des données** La capacité des données dans une base de données à mesurer la rétention de patients peut être caractérisée par le temps s'écoulant entre un événement et l'enregistrement de cet événement. Nous définirons une mesure de *maturité des données*, pour décrire la capacité d'une base de données à mesurer la rétention dans une cohorte. Cette mesure sera testée et validée sur la cohorte simulée.

**Mesures robustes de la rétention** A partir des connaissances obtenues avec les objectifs précédents, nous définirons, testerons et validerons des mesures de la rétention dans les cohortes VIH qui seront robustes à la qualité des données. Ces mesures seront évaluées

sur la capacité à bien mesurer la rétention des patients sur notre cohorte simulée, et sur leur robustesse aux variations de qualité des données.

### 3 Méthode

#### 3.1 Cadre théorique

##### 3.1.1 Processus de suivi du VIH

Les patients VIH se rendent à des rendez-vous réguliers dans leur centre de traitement. A chaque visite, un rendez-vous est fixé pour leur prochaine visite. La date de la  $\nu^{\text{ème}}$  visite du patient  $i$  est notée  $V_\nu^i$ , et la date du rendez-vous qui était fixé pour cette visite est notée  $A_\nu^i$ . Si  $V_\nu^i < A_\nu^i$ , le patient est venu en avance à son rendez-vous. Si  $V_\nu^i > A_\nu^i$ , le patient est venu en retard. On note  $l_\nu^i = A_\nu^i - V_\nu^i$ , le temps écoulé entre le rendez-vous fixé et la visite effective. La Figure 3.1 montre la distribution de  $l_\nu^i$  dans nos données.

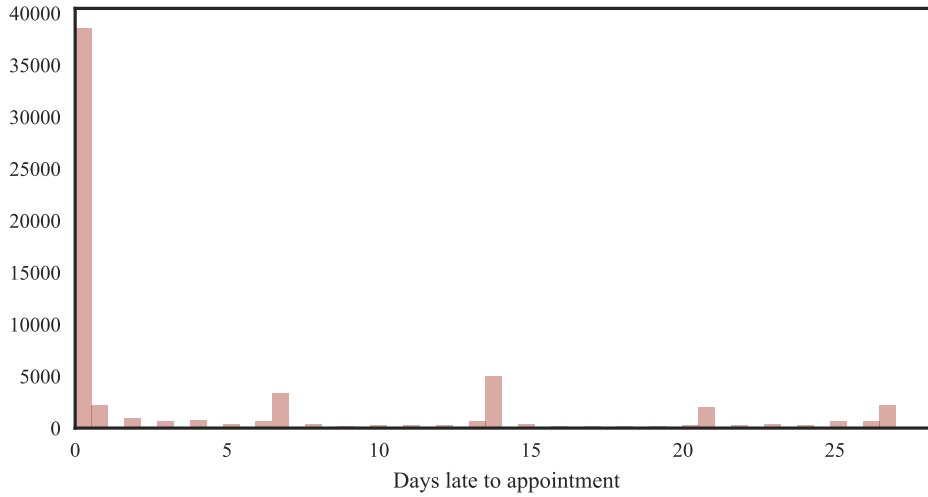


Figure 3.1: Distribution de  $l_\nu^i$  dans nos données.

Le délai entre une visite et la date de prochain rendez-vous donné est défini par les normes et directives nationales sur le suivi des patients sous ARV. Les patients suivis depuis plus longtemps auront en moyenne des temps entre visites plus longs que les patients enrôlés plus récemment. On note  $\bar{f}_\nu^i$  la fréquence de visite du patient  $i$  à la visite  $\nu$ . Cette unité est souvent un multiple de 28 jours, du fait que les patients ont souvent des jours privilégiés pour planifier leurs visites médicales. La Figure 3.2 montre la distribution de  $\bar{f}_\nu^i$  dans nos données.

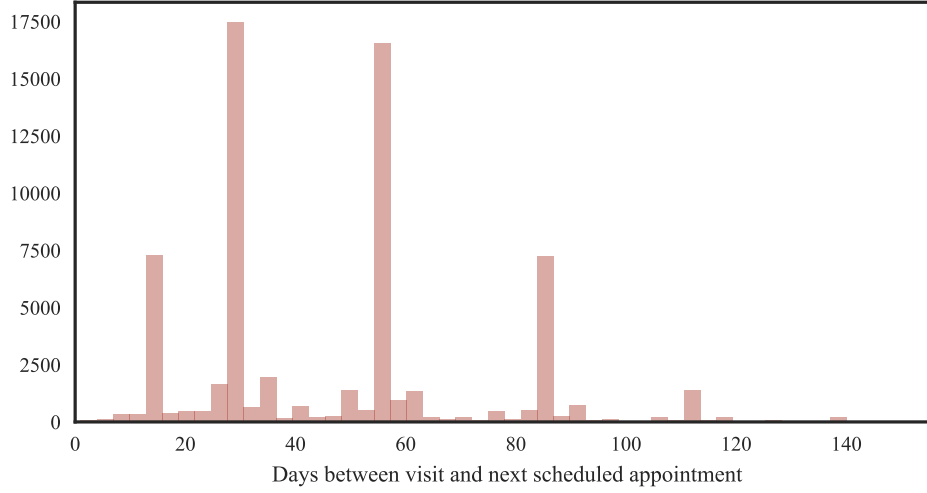


Figure 3.2: Distribution of  $\tilde{f}_v^i$  in our data.

Finalement, le temps entre deux visites peut être décrit comme :

$$V_{v+1}^i - V_v^i = \tilde{f}_v^i + l_v^i$$

### 3.1.2 Processus d'entrée des données

La date à laquelle une visite est enregistrée dans une base de dossiers informatisés est notée  $R(V_v^i)$ . Par définition,  $R(V_v^i) \geq V_v^i$ , et le délai pour l'enregistrement des données est noté:

$$R(V_v^i) - V_v^i = \delta_v^i \geq 0$$

Dans certains cas, une visite ne sera jamais enregistrée, nous noterons cette situation  $\delta_v^i \rightarrow \infty$ .  $\delta_v^i$  peut varier dans un centre de santé, en fonction de la charge de travail, des ressources humaines ou d'autres facteurs. La Figure 3.3 montre la distribution de  $\delta_v^i$  dans nos données.

Finalement, l'enregistrement des données est clos à la date  $T_{close}$  avant que les données ne soient analysées. Pour simplifier, la date de cloture de la base de données sera assimilée à la date d'analyse, quitte à relaxer cette hypothèse lors de notre travail sur la maturité des données.

La Figure 3.4 montre comment les différents paramètres que nous venons de définir s'agencent pour un suivi de patient donné. Ce patient imaginaire a eu une première visite le 1<sup>er</sup> janvier, et

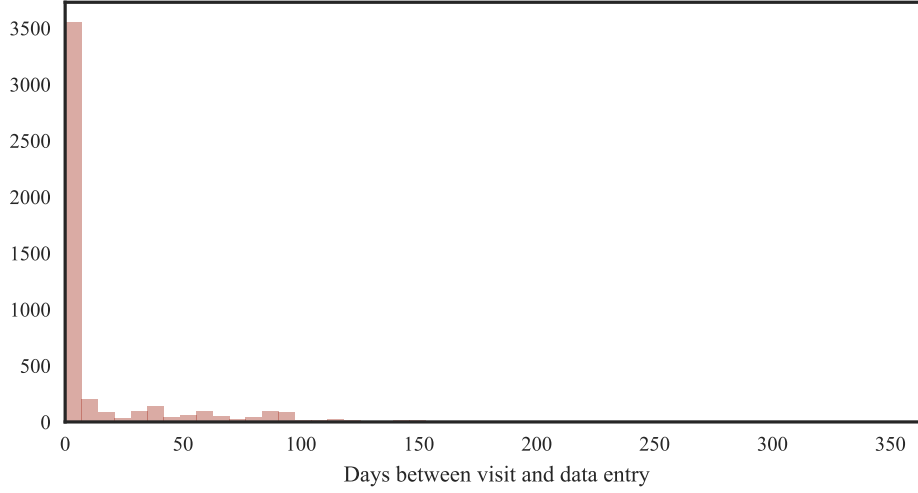


Figure 3.3: Distribution of  $\delta_v^i$  in our data.

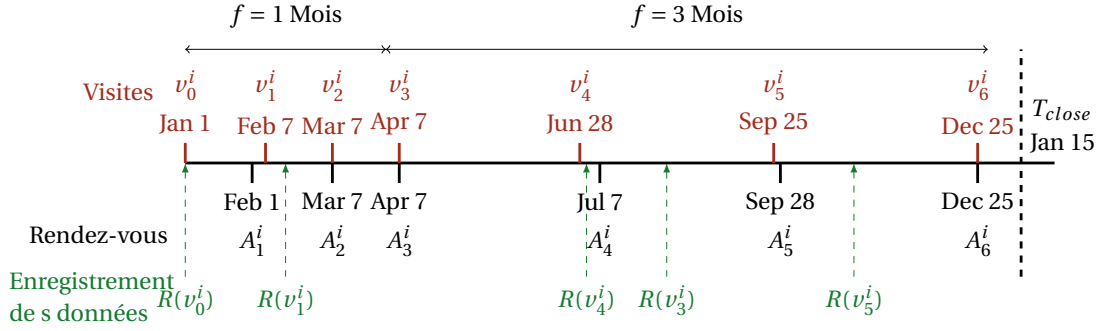


Figure 3.4: Individual follow-up and data entry process

a eu un rendez-vous fixé pour le 1<sup>er</sup> février. Il est venu à ce rendez-vous avec 6 jours de retard. Après trois mois de suivi mensuel, il est passé à un suivi trimestriel. Il est venu en avance à son rendez-vous du mois de juillet, et est venu à tous ses rendez-vous jusqu'à la fin de l'année. Les données ont été enregistrées très rapidement en début d'année, mais  $V_2^i$  n'a jamais été enregistrée.  $V_4^i$  a été enregistrée avant  $V_3^i$ , et  $V_6^i$  n'a pas pu être enregistrée avant la cloture de la base de données pour analyse, le 15 janvier de l'année suivante. Cet exemple donne une démonstration rapide des différentes situations qui peuvent être rencontrées lors de l'analyse de données de suivi patient.

### 3.1.3 Définition des Perdus de Vue

Une pièce centrale de la définition des PDV est la *période de grace* au cours de laquelle un patient, même si il a manqué un rendez-vous fixé, est toujours considéré suivi. Cette *période*

de grace est notée  $G_0$ .

Un patient  $i$  est considéré activement suivi si il n'est pas en retard à son dernier rendez-vous  $v^*$  de plus de  $G_0$  jours.

$$l_v^{*i} \leq G_0$$

En considérant cette situation, on voit qu'elle regroupe trois situations différentes :

1. Le patient est PDV et ne reviendra jamais dans le centre de santé
2. Le patient est en retard à son rendez-vous mais reviendra plus tard et  $l_v^{*i} > G_0$
3. Le patient est venu pour son rendez-vous  $v^* + 1$  mais  $\delta_{v^*+1}^i > G_0$ .

En utilisant cette définition, on peu exprimer la probabilité qu'une patient soit identifié comme PDV à partir des données disponibles. On défini  $X = 1$  le fait qu'un patient soit activement suivi, et  $X = 0$  un patient PDV. On a donc  $p(X = 0 | l_v^{*i} \leq G_0)$

$$p(X = 0 | l_v^{*i} \leq G_0) = 1 - p(X = 1 \cap l_v^{*i} \leq G_0) - p(\delta_{v^*+1}^i > G_0)$$

Using this definition, we can express the probability that a LTFU patient is identified based on the data at hand. Indeed, definin gthe event that a patient is actively in care, and  $X = 0$  if the patient is LTFU. We can thus get  $p(X = 0 | l_v^{*i} \leq G_0)$  as the combination of elements we can measure :

$$p(X = 0 | l_v^{*i} \leq G_0) = 1 - p(X = 1 \cap l_v^{*i} \leq G_0) - p(\delta_{v^*+1}^i > G_0)$$

On voit que  $X = 1 \cap l_v^{*i} \leq G_0$  peut être vu comme l'expression de la qualité des données, et  $\delta_{v^*+1}^i > G_0$  est la myopie intrinsèque du système d'information. Parvenir à faire la différence entre ces deux termes est essentiels pour comprendre la rétention des patients dans la file active et son incertitude.

## 4 Données

Dans l'état actuel du projet, nous disposons de données issues d'une base de dossiers patients informatisés obtenus auprès du programme VIH du Kenya dans le cadre de l'étude ABCE à l'Institute for Health Metrics and Evaluation (IHME). Dans ce centre de santé, 4833 patients ont été enregistrés pour leur suivi VIH, de 2005 à juin 2012, et 69591 visites ont été enregistrées sur cette période. Le délai d'entrée des données est disponible pour au moins 4853 de ces visites.

Nous souhaitons ajouter d'autres bases de données de dossiers patients informatisés à notre étude pour varier à la fois les délais de rendez-vous, et les conditions d'entrée des données. En Haïti, les données du système ISanté présentent sont à la fois une opportunité en terme de nombre de patients suivis, et de durée de suivi, qui nous permettent de bien évaluer la variance de nos différents paramètres par sites et dans le temps.

Ces données sont analysées anonymement et en agrégat. Pour chaque patient, nous n'utilisons que les dates de visites planifiées et effectives. En cas d'absence des dates de rendez-vous, nous imputons des dates de rendez-vous en fonction des pratiques apparentes dans les données. Nous utilisons aussi les métadonnées enregistrées par les systèmes de collecte de données. En particulier, comme expliqué plus haut, nous utilisons les dates d'enregistrement des données pour produire une mesure de la qualité des données disponibles.

Nous ne rapportons pas de données précises sur les résultats du programme de prise en charge, et toutes les données disponibles sont utilisées pour informer les paramètres de simulation de la cohorte VIH.

## 5 Méthodes

Notre travail sera effectué en trois étapes principales. Dans un premier temps, nous estimerons les paramètres nécessaires à notre modèle, à partir des données disponibles. Dans un second temps, nous simulerons une cohorte ainsi que le processus de collecte de données correspondant. Finalement, nous utiliserons ces simulations pour répondre les questions qui nous intéressent.

**Modélisation** - Les différents paramètres présentés dans la section 3.1 seront estimés à partir des cohortes pour lesquelles nous disposerons de dossiers informatisés. Nous utiliserons une approche bayésienne, et estimerons les distributions postérieures de nos paramètres d'intérêt sur lesquelles nous simulerons les réalisations des paramètres à inclure dans la simulation.

**Simulation** - A partir des paramètres estimés à l'étape précédente, nous simulerons une cohorte de suivi de patient VIH sur plusieurs années. Cette simulation sera faite en utilisant le cadre du Cost Effectiveness Analysis Microsimulation (CEAM), développé à IHME pour simuler des files actives.

Dans un second temps, nous simulerons le processus d'enregistrement des données, pour chaque mois, en simulant un  $\delta$  pour chaque visite ayant été simulée à l'étape précédente. Nous obtiendrons ainsi une date de saisie pour chacune de ces visites. Nous disposerons donc finalement de toute l'information nécessaire pour estimer la rétention et mesurer la rétention d'une cohorte donnée. En renouvelant l'exercice en variant les paramètres, nous pourrions estimer les paramètres d'intérêt pour les objectifs de notre étude.

**Paramètre d'intérêt** - A partir de ces données simulées, nous serons donc en mesure de mesurer les éléments suivants.

1. Impact de la qualité des données Nous simulerons et mesurerons  $p(X = 0|\theta_\delta)$  avec différentes valeurs du paramètre  $\theta_\delta$ . Plusieurs scénarios différents seront estimés pour la qualité des données, en variant à la fois la moyenne et la variance de  $\delta$ . Une qualité de donnée parfaite sera comparée à des situations avec des longs délais de saisie de données, et des situations avec d'importantes pertes de données (grande variance de

- l). Les variations de  $p$  qui seront alors observées seront décrites comme l'impact de la qualité des données sur la mesure de la rétention.
2. Maturité des données Au fur et à mesure que les données sont saisies, ou que des patients ayant manqué des visites réintègrent leur suivi, les données d'une certaine période sont complétées. Au fur et à mesure que la maturité des données augmente, la variance due à la qualité des données diminue. Faire varier  $T_c lose$ , peut donc avoir un impact sur la mesure de la rétention d'un patient à une date donnée. Nous mesurerons la rétention de notre cohorte simulée en utilisant plusieurs dates de cloture de la base de données. Ces mesures nous permettront de définir et tester une mesure de la maturité des données, à partir d'une combinaison de  $\tilde{f}$ ,  $l$  et  $\delta$  qui nous permettra d'identifier une date minimale optimale d'analyse pour estimer la rétention dans un programme, et une période de grace optimale  $G_0$  à utiliser pour différents niveaux de maturité.
  3. Mesures robustes de la rétention Finalement, nous explorerons des mesures de la rétention plus robustes que le taux de PDV utilisé. Ces mesures pourront inclure:
    - Le ratio du nombre mensuel moyen de visites, corrigé pour la qualité des données, divisé par le nombre de visites attendues.
    - Le ratio du nombre de nouveau au nombre d'anciens patients.
    - La probabilité que le taux de PDV est plus haut qu'un palier standard fixé.

Pour chacune de ces mesures, nous évaluerons leur capacité à mesurer la rétention dans la file active, en les comparant à la méthode de référence avec données parfaites. Nous évaluerons aussi la sensibilité de ces mesures à la qualité et à la maturité des données.

## 6 Agenda Prévisionnel

## 7 Références

### References

- [1] J. Ioannidis, T. Taha, N. Kumwenda, R. Broadhead, L. Mtimavalye, P. Miotti, F. Yellin, D. Contopoulos-Ioannidis, and R. Biggar, "Predictors and impact of losses to follow up in an HIV1 perinatal transmission cohort in Malawi.pdf," *International Journal of Epidemiology*, vol. 28, pp. 769–775, 1999.
- [2] B. Lebouché, Y. Yazdanpanah, Y. Gerard, D. Sissoko, F. Ajana, I. Alcaraz, P. Boitte, B. Cadoré, and Y. Mouton, "Incidence rate and risk factors for loss to follow-up in a French clinical cohort of HIV-infected patients from January 1985 to January 1998," *HIV medicine*, vol. 7, no. 3, pp. 140–145, 2006.
- [3] R. Moh, C. Danel, E. Messou, T. Ouassa, D. Gabillard, A. Anzian, Y. Abo, R. Salamon, E. Bissagnene, C. Seyler, S. Eholie, and X. Anglaret, "Incidence and determinants of mortality and morbidity following early antiretroviral therapy initiation in HIV-infected adults in West Africa," *AIDS*, vol. 21, no. 18, p. 2483, 2007.



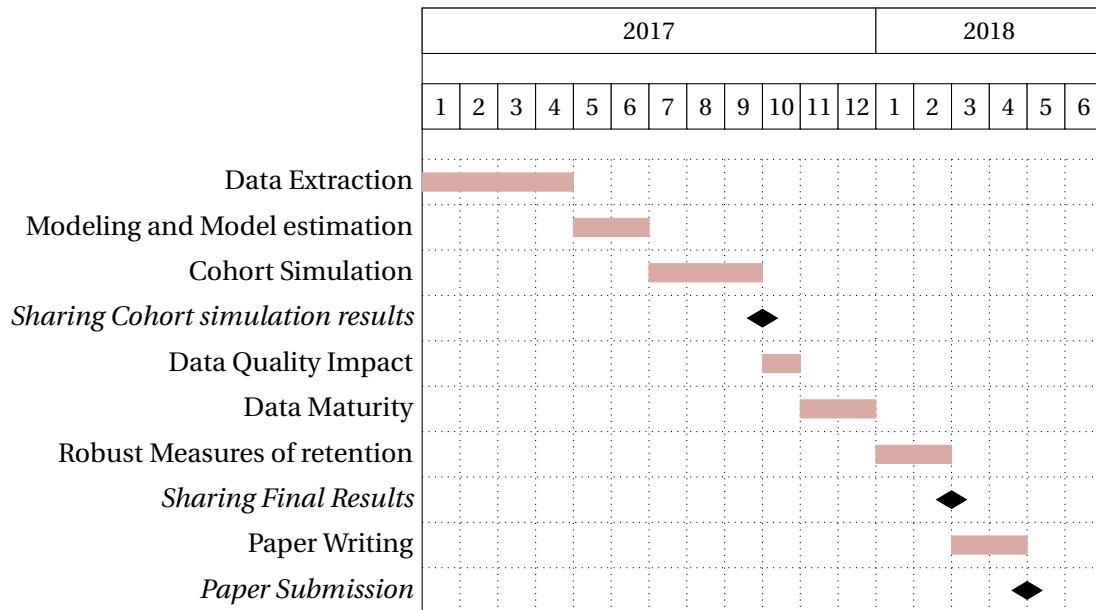


Figure 6.1: Gantt Chart for Aim 1

- [4] S. Carillon, "Les ruptures de suivi médical des personnes vivant avec le VIH à Kayes (Mali). Approche anthropologique," *Sciences Sociales et Santé*, vol. 29, no. 2, 2011.
- [5] J. Kwong-Leung Yu, S. Chen, K.-Y. Wang, C.-S. Chang, S. Makombe, E. Schouten, and A. Harries, "True outcomes for patients on antiretroviral therapy who are lost to follow-up in Malawi," *Bulletin of the World Health Organization*, vol. 85, no. 7, pp. 550–554, 2007.
- [6] R. Dalal, C. MacPhail, M. Mqhayi, J. Wing, C. Feldman, M. Chersich, and W. Venter, "Characteristics and outcomes of adult patients lost to follow-up at an antiretroviral treatment clinic in Johannesburg, South Africa," *Journal of Acquired Immune Deficiency Syndromes*, vol. 47, no. 1, p. 101, 2008.
- [7] M. McGuire, T. Munyenembe, E. Szumilin, A. Heinzelmänn, M. Le Paih, N. Bouithy, and M. Pujades-Rodríguez, "Vital status of pre-ART and ART patients defaulting from care in rural Malawi," *Tropical Medicine & International Health*, vol. 15, pp. 55–62, 2010.
- [8] C. T. Yiannoutsos, M.-W. An, C. E. Frangakis, B. S. Musick, P. Braitstein, K. Wooll-Kaloustian, D. Ochieng, J. N. Martin, M. C. Bacon, V. Ochieng, and S. Kimaiyo, "Sampling-Based Approaches to Improve Estimation of Mortality among Patient Dropouts: Experience from a Large PEPFAR-Funded Program in Western Kenya," *PLoS ONE*, vol. 3, p. e3843, Dec. 2008.
- [9] E. Geng, D. Glidden, N. Emenyonu, N. Musinguzi, M. Bwana, T. Neilands, W. Muyindike, C. Yiannoutsos, S. Deeks, D. Bangsberg, *et al.*, "Tracking a sample of patients lost to follow-up has a major impact on understanding determinants of survival in HIV-infected

patients on antiretroviral therapy in Africa,” *Tropical Medicine & International Health*, vol. 15, pp. 63–69, 2010.

- [10] J.-M. Tassie, K. Malateste, M. Pujades-Rodríguez, E. Poulet, D. Bennett, A. Harries, M. Mahy, M. Schechter, Y. Souteyrand, F. Dabis, and IeDEA, “Evaluation of Three Sampling Methods to Monitor Outcomes of Antiretroviral Treatment Programmes in Low- and Middle-Income Countries,” *PLoS ONE*, vol. 5, p. e13899, Nov. 2010.
- [11] M. Brinkhof, B. Spycher, C. Yiannoutsos, R. Weigel, R. Wood, E. Messou, A. Boulle, M. Egger, J. Sterne, and IeDEA, “Adjusting Mortality for Loss to Follow-Up: Analysis of Five ART Programmes in Sub-Saharan Africa,” *PLoS ONE*, vol. 5, p. e14149, Nov. 2010.
- [12] M. Egger, B. Spycher, J. Sidle, R. Weigel, E. Geng, M. Fox, P. MacPhail, G. van Cutsem, E. Messou, R. Wood, D. Nash, M. Pascoe, D. Dickinson, J.-F. Etard, J. McIntyre, M. Brinkhof, and IeDEA, “Correcting Mortality for Loss to Follow-Up: A Nomogram Applied to Antiretroviral Treatment Programmes in Sub-Saharan Africa,” *PLoS Medicine*, vol. 8, p. e1000390, Jan. 2011.
- [13] G. Van Cutsem, N. Ford, K. Hildebrand, E. Goemaere, S. Mathee, M. Abrahams, D. Coetzee, and A. Boulle, “Correcting for Mortality Among Patients Lost to Follow Up on Antiretroviral Therapy in South Africa: A Cohort Analysis,” *PLoS ONE*, vol. 6, p. e14684, Feb. 2011.
- [14] J. Henriques, M. Pujades-Rodríguez, M. McGuire, E. Szumilin, J. Iwaz, J.-F. Etard, and R. Ecochard, “Comparison of Methods to Correct Survival Estimates and Survival Regression Analysis on a Large HIV African Cohort,” *PLoS ONE*, vol. 7, p. e31706, Feb. 2012.
- [15] S. Verguet, S. S. Lim, C. J. L. Murray, E. Gakidou, and J. A. Salomon, “Incorporating loss to follow-up in estimates of survival among HIV-infected individuals in sub-saharan Africa enrolled in antiretroviral therapy programs,” *Journal of Infectious Diseases*, vol. 207, no. 1, 2013.
- [16] B. Chi, R. Cantrell, A. Mwango, A. Westfall, W. Mutale, M. Limbada, L. Mulenga, S. Vermund, and J. Stringer, “An Empirical Approach to Defining Loss to Follow-up Among Patients Enrolled in Antiretroviral Treatment Programs,” *American Journal of Epidemiology*, vol. 171, pp. 924–931, Jan. 2010.
- [17] B. Chi, C. Yiannoutsos, A. Westfall, J. Newman, J. Zhou, C. Cesar, M. Brinkhof, A. Mwango, E. Balestre, G. Carriquiry, T. Sirisanthana, H. Mukumbi, J. Martin, A. Grimsrud, M. Bacon, R. Thiebaut, and IeDEA, “Universal Definition of Loss to Follow-Up in HIV Treatment Programs: A Statistical Analysis of 111 Facilities in Africa, Asia, and Latin America,” *PLoS Medicine*, vol. 9, p. e1001111, Oct. 2011.
- [18] M. P. Fox, B. Larson, and S. Rosen, “Defining retention and attrition in pre-antiretroviral HIV care: proposals based on experience in Africa: Editorial,” *Tropical Medicine & International Health*, vol. 17, pp. 1235–1244, Oct. 2012.

- [19] M. Mugavero, A. Westfall, A. Zinski, J. Davila, M.-L. Drainoni, L. I. Gardner, J. Keruly, F. Malitz, G. Marks, L. Metsch, T. Wilson, and T. Giordano, "Measuring retention in HIV care: the elusive gold standard.," *Journal of Acquired Immune Deficiency Syndromes*, vol. 61, pp. 574–80, Dec. 2012.
- [20] B. R. Yehia, J. A. Fleishman, J. P. Metlay, P. T. Korthuis, A. L. Agwu, S. A. Berry, R. D. Moore, and K. A. Gebo, "Comparing different measures of retention in outpatient HIV care," *AIDS*, vol. 26, pp. 1131–1139, June 2012.
- [21] A. T. Grimsrud, M. Cornell, M. Egger, A. Boulle, and L. Myer, "Impact of definitions of loss to follow-up (LTFU) in antiretroviral therapy program evaluation: variation in the definition can have an appreciable impact on estimated proportions of LTFU.," *Journal of clinical epidemiology*, vol. 66, pp. 1006–13, Sept. 2013.
- [22] B. E. Shepherd, M. Blevins, L. M. E. Vaz, T. D. Moon, A. M. Kipp, E. José, F. G. Ferreira, and S. H. Vermund, "Impact of definitions of loss to follow-up on estimates of retention, disease progression, and mortality: application to an HIV program in Mozambique.," *American journal of epidemiology*, vol. 178, pp. 819–28, Sept. 2013.