# Instructie verstrekken Sapphire Celldyn data door UPOD datamanagers

Inhoud

[Instructie verstrekken Sapphire Celldyn data door UPOD datamanagers 4](#_Toc89357316)

[Doel 4](#_Toc89357317)

[Doelgroepen 4](#_Toc89357318)

[1) Opbouw UPOD Sapphire bestand 4](#_Toc89357319)

[a) Bronnen 5](#_Toc89357320)

[b) Dataflow 6](#_Toc89357321)

[c) Kwaliteits controle 6](#_Toc89357322)

[2) Vraagstelling 6](#_Toc89357323)

[a) Biologische vraagstelling 6](#_Toc89357324)

[b) Logisitieke en procesvraagstelling 7](#_Toc89357325)

[c) Uitlevering 7](#_Toc89357326)

[3) Parameter types 7](#_Toc89357327)

-----------------------

# Doel

De interpretatie van de Sapphire data is complex. Doel van dit document is om UPOD datamanagers meer achtergrond informatie te geven om de juiste Sapphire Celldyn bestanden voor onderzoek te extraheren. Om dit doel te bereiken dient het juiste gesprek met de onderzoeker over de benodigde data gevoerd te worden waarbij de datamanager begrijpt welke zaken bepalend zijn voor het maken van de juiste dataset. Deze instructie bevat daarom achtergrondinformatie over de sapphire en aandachtspunten voor extractie die besproken moeten worden.

# Doelgroepen

Hfdst 1: uitvoering voor datamanagers, achtergrondinfo voor data scientists in het belang van de duiding van data.

Hdst 2: van belang voor de data selectie en duiding van data en resultaten. Wanneer dit bij de intake duidelijk is, kan de datamanager deze aspecten meenemen. Wanneer dit gaandeweg duidelijk wordt, geeft dit de data scientist richting in de data processing.

## Opbouw UPOD Sapphire bestand

De Abbott Celldyn Sapphire is een bloedcell analyzer die van 2005 tot 2020 in het UMCU in gebruik is (geweest) op het Centraal lab, de bloedprik poli, en voor een korte periode in het Prinses Maxima Centrum. In de hoogtij dagen van de Sapphire waren tot 6 analyzers tegelijk actief. Enkele zijn in de loop der jaren vervangen door nieuwe met een ander serienummer.

Sinds 11 februari 2020 zijn op het Centraal lab Abbott Alinity HQ analyzers in gebruik en wordt er nog 1 Sapphire (42710AZ96) op het lab gebruikt voor speciale patienten en vergelijkende studies (Stand november 2021).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Analyser name | aantal | eerste\_sample | laatste\_sample | (locatie voor zover bekend) |
| 00203ab | 83824 | 06JAN2005 | 20DEC2005 |  |
| 273ab | 60674 | 01JAN2005 | 05DEC2005 |  |
| 32037AB | 33382 | 01JAN2005 | 05JAN2006 |  |
| 32122AB | 83582 | 03JAN2005 | 01DEC2009 |  |
| 42192AZ | 528357 | 22DEC2005 | 08FEB2020 | Centraal lab 1 |
| 42193AZ | 669341 | 13DEC2005 | 13AUG2014 | Centraal lab 2 |
| 42194AZ | 1081359 | 02DEC2005 | 29JAN2020 | Centraal lab 3 (voor WKZ monsters) |
| 42710AZ96 | 740017 | 19MAR2013 |  | Poli UMC, centraal CD61 |
| 42783az | 186690 | 02DEC2009 | 27MAY2020 | Poli UMC, tijdelijk PMC |
| 45146AZ | 74722 | 16MAR2018 | 10JUN2020 | Tijdelijk PMC |
| AF31933 | 2524 | 15FEB2006 | 26FEB2006 |  |
| S-005 | 39412 | 11MAY2006 | 24DEC2015 | research |

### Bronnen

De ruwe data van de sapphire wordt opgeslagen in FCS files, waarbij van elk event (bloedcel of partikel) door het ‘Algoritme’ bepaald wordt welk type bloedcellen het sample bevat. De resultaten van deze classificatie komen in het DataLogExtract ofwel de DLE file terecht.

De Sapphire genereert daarnaast voor de kliniek een PDF rapport met een lijstje gemeten parameters en enkele scatter-plots. Deze 3 bestandstypen worden in UPOD opgeslagen. Via het AbbottLink protocol worden alleen geselecteerde parameters direct naar GLIMS verstuurd. Dit protocol heeft een eigen syntax met eigen foutcodes in een soort RS-232 format. In GLIMS wordt extra informatie over het sample, verdunning en rejection van de (gedeeltelijke) sapphire run en eventuele re-run advies vastgelegd, nadat de data van de analyzer verzonden is. Deze beslissingen worden gedeeltelijk bepaald door regels in GLIMS. Dit genereert meldingen en waarschuwingen waar analysten op reageren waarbij zij dienen te handelen volgens de SOP. Deze informatie kan soms worden toegevoegd aan de dat

a indien dit in GLIMS inzichtelijk is. De data op de analyzer wordt niet meer aangepast wat het terugkoppelen van GLIMS data ingewikkeld maakt.

### Dataflow

De verwerking van de data is grafisch weergegeven in het stroomdiagram (..).

### Kwaliteits controle

Kwaliteitscontrole vindt op een aantal niveau’s plaats.

* + 1. Metingen worden door de Sapphire volledig mislukt of ongeldig verklaard. Denk aan “te weinig geaspireerd sample”, “Klonten in aspiratienaald”, “Technische fouten”. Dit betreft een 9 tal technical failures die leiden tot cancellation van de hele meting. UPOD markeert die samples als logical\_delete=1.
    2. Glims vlaggen obv groupalerts?

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| SasName | UPOD\_Name | label | Parameter\_type |  |
| wbcAlrt | c\_Alrt\_wbc | Alert for WBC parameter | group alert | WBC |
| rbcAlrt | c\_Alrt\_rbc | Alert on RBC count | group alert | RBC |
| pltAlrt | c\_Alrt\_plt | Alert on PLT count | group alert | PLT |
| hgbAlrt | c\_Alrt\_hb | Alert on HGB count | group alert | HGB |
| rtcAlrt | c\_Alrt\_retc | Alert for the presence of a flag on the reticulocyte count | group alert | RETC |
| '61Alrt'n | c\_Alrt\_61 | Alert for the presence of a CD61 flag | group alert | PLT CD61 |

* + 1. Suspect vlaggen (c\_s\_…) in de celldyn data. Dit zijn parameter gebonden vlaggen met de volgene betekenis (mail Ludi van Dun , Abbott)

Code 1 betekent inderdaad dat het resultaat goed is en er geen gerelateerd alarm of flagging is die het resultaat in twijfel doen trekken.

Code 2 betekent dat het resultaat “invalid” is, er staat een \* achter het resultaat en dit houdt in dat het resultaat niet zondermeer gerapporteerd mag worden. Bij Code 2 hoort derhalve altijd een alarm op basis waarvan het algorithme van mening is dat het resultaat onbetrouwbaar is. Voorbeeld is bijvoorbeeld wanneer het alarm “unable to find clear separation between WBC subpopulations” geplaatst wordt en er een \* achter het resultaat van de Neu en Eo staat. Analyzer geeft hiermee aan dat de scheiding tussen de Neu en Eo niet optimaal is en trekt het resultaat in twijfel.

Code 3 houdt in dat een resultaat niet gerapporteerd kan worden omdat het te ver buiten het meetbereik van de analyzer ligt. Het resultaat voor de betreffende parameter wordt dan weergegeven als >>>>. Code 3 wordt **niet** weergegeven wanneer een resultaat buiten het analytisch meetbereik van de analyzer ligt. In dit geval wordt er een Code 2 gegeven op basis van het alarm “outside analytical measurement range”.

Code 4 wordt geplaatst wanneer er geen resultaat beschikbaar is voor een bepaalde parameter. Het resultaat ontbreekt en op het scherm of de printout zie je een leeg veld. Meestal zie je de Code 4 voorkomen bij een Background waarbij alleen de WBC, RBC, HGB en PLT gerapporteerd worden. Alle andere parameters zoals bv WBC differentiatie, MCV, RDW krijgen dan een Code 4. Sporadisch zal het voorkomen dat een resultaat om een technische reden niet gerapporteerd wordt.

Code 5 wordt geplaatst wanneer er aan het resultaat een “suspect population” gekoppeld is. Dit wil dus niet automatisch zeggen dat het resultaat onbetrouwbaar is. Voorbeeld is bijvoorbeeld wanneer er een VarLym flagging geplaatst wordt en de absolute en procentuele LYM een code 5 krijgen.

Het houdt echter ook niet in dat een resultaat altijd betrouwbaar te rapporteren is. Voorbeeld hiervan kan de BLAST flag zijn. Analyzer geeft hierbij aan dat er mogelijk blasten aanwezig zijn. Wanneer er uit de manuele differentiatie blijkt dat er inderdaad blasten aanwezig zijn dan houdt dit ook automatisch in dat de door de Sapphire gerapporteerde 5-part differentiatie niet klopt en waarschijnlijk de Mono of LYM populatie te hoog ingeschat wordt (blasten worden meest waarschijnlijk bij een van beiden opgeteld).

## Vraagstelling

De klinische vraagstelling bepaalt in grote mate hoe de uit te leveren bloedcel data er uit moet zien en welke data nodig zijn. De volledige sapphire dataset in UPOD telt ongeveer 300 parameters die niet allemaal even zinvol zijn in elke context. Ook de mate van quality control (QC) en de selectie van zowel parameters als bepalingen hangt deels van de klinische vraagstelling af. Vraagstellingen kunnen biologisch, medisch van aard zijn waarbij het gaat om biologische effecten die onderzocht worden met de data, of meer gericht op de doorlooptijden, klinisch handelen of beslissen in het zorgproces. Bij de intake is het belangrijk de toepasbaarheid van volgende situaties te bespreken. Ter illustratie staan het eind van dit onderdeel een drietal ‘use cases’ beschreven.

### Biologische vraagstelling

Bij deze vraagstellingen wordt gekeken of er bij bepaalde ziekten veranderingen in het bloed te zien zijn. Afhankelijk van de vraagstelling kunnen variaties in de celldyn data voorspellend, verklarend of juist de uitkomst zijn bij de conditie van de patient. Dus als confounder, modifier of outcome. Daarnaast wordt er afhankelijk van de vraagstelling op één vastgesteld moment bij de patient naar data van een sapphire bepaling gekeken of naar het verloop van de bloedparameters over de tijd in verschillende bepalingen bij dezelfde patient op diverse tijdstippen. Ook de variabiliteit van specifieke parameters, zowel in de tijd als intra- en inter-individueel zijn belangrijk.

#### Algemene patientpopulaties

Als het om algemene patientenpopulaties gaat heeft het niet altijd zin om hele zeldzame, specifieke celldyn uitkomsten te rapporteren of uit te leveren. Zeker als enkele parameters bestudeerd worden als confounder of outcome, dan zijn parameters die standaard gerapporteerd worden in HIX meestal genoeg. Daarnaast hebben outliers in deze data waarbij suspect vlaggen optreden meestal geen klinische waarde. Dus valt te overwegen om het datapunt van een ‘suspecte’, ‘invalid’ of ‘out of range’ waarde weg te laten omdat dit dan vaak metingen betreft die geen valide uitkomst bevatten (use case 1).

#### Specifieke populatie

In sommige patientpopulaties valt te verwachten dat het bloedbeeld afwijkend is en daardoor tegen de grenzen van het meetbereik van de sapphires ligt. Denk daarbij aan patienten die veel bloed verloren hebben, een anemie (=tekort van een soort bloedcellen) met lage aantallen cellen of juist leukemie waarbij woekeringen van leukocyten optreden en hoge waarden gemeten worden in andere verhoudingen. De kans bestaat dat het algoritme in de sapphire dan alarmen en vlaggen geeft voor uitzonderlijke (zeldzame) waarden terwijl die in deze specifieke populatie vaak wel reëel zijn (use case 2).

### Logisitieke en procesvraagstelling

Bij deze vragen gaat het niet alleen om de gemeten bloedwaarden, maar ook om de tijden waarop en de aantal metingen in de diagnostiek. Andere parameters zijn dan relevant en ook de her-metingen bij geheel of mislukte metingen van de Sapphire.

#### Doorlooptijd

Voorbeelden hiervan zijn doelmatigheidsstudies in het lab en in de kliniek ter verbetering van de zorg of het proces hier om heen (use case3).

#### Foutmeldingen

Het optreden van foutmeldingen kan in sommige studies ook voorspellende waarde hebben. Anomalieën in het bloedbeeld bij patienten met een bepaalde conditie waarbij atypische cellen of extreme verhoudingen tussen cellen optreden kunnen tot foutmeldingen van de Sapphire leiden. Oorzaak kan dan zijn dat het algoritme tegen zijn betrouwbaarheidsgrenzen aanloopt en er een handdiff gedaan moet worden. Dat gegeven kan in sommige vraagstellingen informatie zijn.

## Uitlevering

### Parameter types

### Use cases

*Use case 1a: Voorspeller van cardiovasculair event bij patiënten van afdeling X en poli Y.*

Er zijn een aantal cardiovasculaire risicofactoren bekend die in de kliniek gebruikt worden, zoals leeftijd, geslacht, roken, bloeddruk en labwaarden van cholesterol (totaal en HDL). Uitgesproken en bewezen bloedcelkarakteristieken horen **niet** tot het standaard lijstje. De a priori kans op extreme celldyn uitslagen of bijzondere vlaggen is **klein**. Er zou volstaan kunnen worden met een “standaardset” van celldyn uitslagen. Ter controle worden het aandeel aan vlaggen bijgeleverd en kan een nadere keuze worden gemaakt. Runs met vlaggen die technical failures aangeven (technisch mankement of verstoring tijdens de meting van het monster; logical\_delete = 1) zouden verwijderd moeten worden omdat de uitslagen onbetrouwbaar zijn.

*Use case 1b: Voorspeller van cardiovasculair event bij patiënten van afdeling X en poli Y met afwijkende hematologie-eigenschappen.*

Er zijn een aantal cardiovasculaire risicofactoren bekend die in de kliniek gebruikt worden, zoals leeftijd, geslacht, roken, bloeddruk en labwaarden van cholesterol (totaal en HDL). Uitgesproken en bewezen bloedcelkarakteristieken horen **niet** tot het standaard lijstje. De a priori kans op extreme celldyn uitslagen of bijzondere vlaggen is **groot**. Bij monsters van de patientengroep zou een bepaald type suspectvlag bijvoorbeeld vaker kunnen voorkomen, bijvoorbeeld thromboklontjes (platelet clumps Pltclmp) of zelfs verstoppende naalden (een van de technical failures) door snel stollend bloed. Runs met deze suspect of technical faliure vlaggen kunnen dan juist wel waardevol zijn, waarbij de vlag zelf onderdeel kan worden van het model.

*Use case 1c: Andere voorspeller van cardiovasculair event bij patiënten van afdeling X en poli Y.*

Er zijn een aantal cardiovasculaire risicofactoren bekend die in de kliniek gebruikt worden, zoals leeftijd, geslacht, roken, bloeddruk en labwaarden van cholesterol (totaal en HDL). Uitgesproken en bewezen bloedcelkarakteristieken horen **wel** tot het standaard lijstje. De performance van het model zou afgezet kunnen worden tegen de afkapwaarde van hematologie-uitslagen die de arts normaliter gerapporteerd krijgt. Die uitslag kan verkregen zijn uit de celldyn, uit de handdif of manueel gecorrigeerd zijn. Het model zou een andere afkapwaarde kunnen voorstellen, al dan niet in combinatie met andere kenmerken zoals vlaggen.

*Use case 2: voorkomen van blasten in hematologie/hemato-oncologie populatie*

Een blast is een leukocyt in het jongste stadium. Deze komen normaliter niet voor in het bloed, maar wel bij aandoeningen als leukemie en bij uitzaaiingen. Deze patienten zouden zich overal kunnen presenteren en de monsters kunnen dus van alle aanvragers kunnen zijn, maar het overgrote deel zal de aanvrager hematoloog/hematolo-oncoloog of oncoloog hebben.

De celldyn vlagt voor een verdenking op de aanwezigheid van blasten in het bloed. Er wordt dan microscopisch onderzoek (Handdiff) uitgevoerd op het monster en de uitslagen van de Handdiff worden gerapporteerd naar het EPD en overrulen dus de celldyn uitslagen. Een blast kan gemeten worden door de Celldyn in de latere versies (?) als research-only parameter, maar deze uitslag wordt ook niet gerapporteerd aan de aanvrager in HiX.

Als de waarneming van blasten als gouden standaard meegenomen worden, dan moeten dus ook de blast-uitslagen in Glims of het EPD betrokken worden. In tegenstelling tot use case 1 moeten dus nog andere bronnen voor bloedcelkarakteristieken geraadpleegd worden.

Omdat van sommige monsters geen microscopisch onderzoek op blasten wordt uitgevoerd (efficiency en beperkte meerwaarde van de bevinding), kan niet gesteld worden dat de afwezigheid van een blastentelling gelijkt staat aan “geen blasten”. Een blastenuitslag van 0 (dus wel uitgevoerd in plaats van ‘leeg’) betekent dat er geen blasten waargenomen werden.

*Use case 3: De doorlooptijd van monsters op de SEH tot het lab en tot aan rapportage.* Dit is dus onderzoek waarbij de labuitslagen niet het primaire doel zijn, maar de timestamps. Om vast te stellen wanneer een monster daadwerkelijk gemeten wordt, is de eerste run van belang, ongeacht het bestaan van vlaggen. In tegenstelling tot use case 1 zijn vlaggen dus niet relevant voor het primaire doel. Wanneer de vraag is welke monsters meer of minder vertraging hebben tot uiteindelijke rapportage, kan het juist wel interessant zijn om te kijken naar multiple runs en de bijbehorende vlaggen.

Losse aantekeningen.....

Voor je bepalingen zijn de zogenaamde suspect vlaggen van de cell-dyn Sapphire toegevoegd. De betekenis is als volgt:

|  |  |
| --- | --- |
| Code | Betekenis |
| 1 | Goed |
| 2 | \*Invalid data (in Abbott pdf report) |
| 3 | Komt blijkbaar niet voor |
| 4 | \*No FDA reportable data in report - out of range (low?) |
| 5 | s flag in report - out of range (high) |

\*--------------------------------------------------------

Kolommen die beginnen met c\_s\_ bevatten zogenaamde suspect vlaggen voor bijbehorende c\_b\_ kolommen. Deze hebben waarden tussen 1 en 5.

1=uitslag is OK

2=ínvalid data

4=no (FDA) reportable data

5=out of range (suspect flag in reporting)

Bijgaand in excel een datadictionary voor de 140 variabelen in deze file.

\*--------------------------------------------------------

Bijgaand de files zoals maandag besproken. Het zijn 3 bestanden met:

-Patienten

-Bloedtransfusies

-Celldynmetingen.

Daarnaast een excel bestand met eenvoudige beschrijving van de inhoud van de bestanden. Het betreft dus NICU opnames in 2017 waarbij kind geboren is in UMCU (in diezelfde opname). De patientenfile bevat de episode per afdeling van deze opname. Dus indien patient verhuist tijdens de opname ontstaat er een extra regel in de patientenfile.

De celldyn file bevat alle celldyn metingen-momenten voor deze patienten. Ik heb de rode bloedcel gerelateerde parameters geselecteerd inclusief de erythroblasten. Van een aantal parameters heb ik naast de waarde ook de zogenaamde 'suspect vlag' meegenomen. Voor bijvoorbeeld Erythroblasten =C\_b\_NRBC  is dat parameter c\_s\_NRBC. Deze kan volgende waarden hebben:

1=uitslag is OK

2=ínvalid data

4=no (FDA) reportable data

5=out of range (suspect flag in reporting)

Omdat we toch de grenzen opzoeken lijkt het me goed om, voor dat we iets willen publiceren, met deze vlaggen rekening houden.

Concept opbouw instructiedocument

1. Herhaling van de klinische vraagstelling
   1. Biologie
      1. Generieke populatie
      2. Specifieke populatie met algemeen bekende bijzondere kenmerken in bloedbeeld
   2. Workflow/logistiek
      1. Doorlooptijd
      2. # foutmeldingen
2. Opbouw bestand en bronnen, verwijzen naar dataflow, korte beschrijving
3. Parameter types
   1. Betekenis van parameter types
   2. Selectiekeuzes en impact
   3. Beslisboom
      1. biologievraag in generieke populatie
      2. Specifieke populaties, en workflow/logistiek: op maat, met domeinexpert(s)