

# PROCESS MINING OBSTETRICAL CARE CLAIMS DATA

## Summary

The Social Insurance Bank of Curaçao (SVB) reimburses healthcare providers for delivering obstetrical care (childbirth). These reimbursements are processed as claims data and meet all the requirements of an event log for process mining. This case study provides the findings of a process mining initiative applied to obstetrical care claims data in Curaçao, covering three years from 2018 until 2020.

## Background

Obstetrical care in Curaçao is provided by two types of healthcare providers: (1) midwives and (2) gynecologists.

In theory, healthy pregnant women should receive obstetrical care from midwives, while at-risk pregnant women should be directed to a gynecologist. In practice, the volume of deliveries at the gynecologist is much higher compared to the midwives.

One particular claim by the midwife clinic is that whenever a midwife sends her client to a gynecologist for a single check-up in the form of a pregnancy ultrasound, the chance is great that this client never returns to the midwife clinic. The claim perpetuated by the midwife

- Process mining case study based on insurance claims data
- Goal was to test the validity of a concrete hypothesis
- Key success factor was the segmentation of the data to create baseline for analysis

clinic is that there is a high level of undue retention of pregnant women amongst gynecologists, especially of clients that initially started their process at the midwife clinic.

Another relevant consideration is that some women may prefer to be treated by a gynecologist rather than a midwife. In case of an emergency during labor, the woman needs to be transported to the gynecologist at the hospital in a rushed fashion. The midwife clinic is at least a ten-minute ambulance drive from the hospital, not including response time.

These claims and considerations merit deeper analysis to understand the patient journey of pregnant women. Because the obstetrical process has a clear beginning and end, process mining is perfect for analyzing this case.

## Research questions

We formulated the following research questions to guide the process mining project.

1. What does the overall obstetrical process look like?
2. How is the interaction of women flowing between the gynecologist and the midwife clinic?

Remember that the SVB is not an active player in this process but merely a passive purchaser of the services. The focus of this process mining project is not to improve operations. Instead, we want to test the validity of the claim that gynecologists “steal away” patients from the midwife clinic.

## Data pre-processing of gynecologist claims data

The activities in the event log of the gynecologist are based on ‘fee-for-service’ claims. As a result, each activity is recorded on a fairly detailed level with its corresponding timestamp. The timestamp contains the date but no hours or minutes.

Our analysis of the gynecologist claims dataset required us to understand what distinguishes a gynecological process from an obstetrical process. All gynecologists in Curaçao are OB-GYN doctors, which means that they deliver both obstetrical (childbirth, or OB) and gynecological (female reproductive system, or GYN) care.

Most care delivered by OB-GYN doctors in Curaçao is GYN-related, not OB. Including all GYN-related cases in the analysis results in a process map that downplays the OB process because the associated frequencies of OB care are smaller. However, without any medical domain knowledge, it can be challenging to discern which activities are OB and which are GYN.

When we import the complete OB-GYN event log in Disco, we see a “spaghetti map” [1]. The spaghetti is heavily concentrated around the activity ‘Follow up consultation’ (see Figure 1).

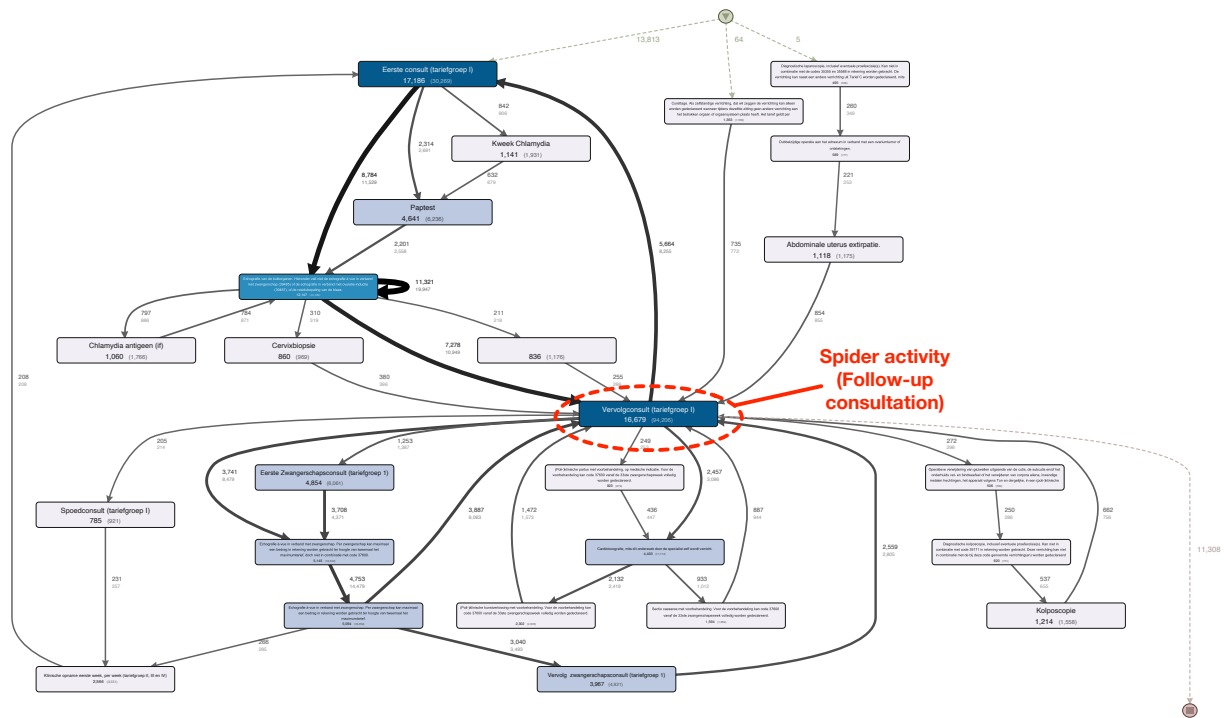


Figure 1. The raw process map for OB-GYN doctors

The follow-up consultation is the most frequent activity. It is performed in nearly all stages of the OB-GYN process. Thus, in terms of process mining, it can be considered to be a spider activity [2]. This means that almost all activities on the process map point towards or from it. It is a helpful practice in process mining to remove spider activities from the map.

Upon removing the spider activity, yet another spider activity presents itself, namely the 'First consultation'. After removing these two spider activities, a much more logical process map appears (see Figure 2).

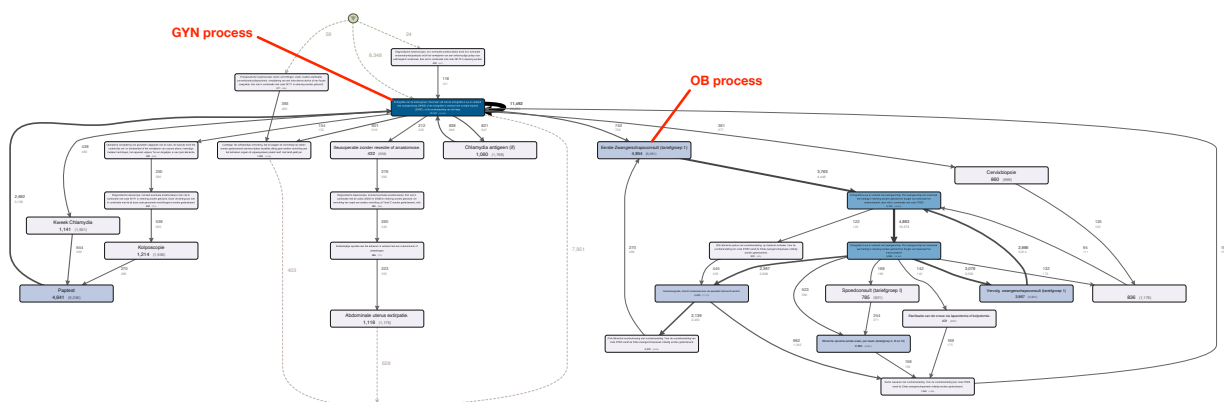


Figure 2. The process map for OB-GYN doctors after filtering out two spider activities

From the process map in Figure 2, one can distinguish two different processes. On the left side, we see several GYN-procedures. On the right side, we see a series of activities that culminate in childbirth (delivery or caesarian section). Thus, a process mining analyst can now recognize which activities are sequentially related to OB without any domain knowledge. This includes activities that are less recognizable for a layperson, such as a CTG scan. We use this process discovery to identify which medical activities in the OB-GYN event log are related to OB and which ones are related to GYN.

As a next step, we now only filter the OB-related activities. The GYN codes cover about 85% of all the OB-GYN event data, whereas the OB-activities only cover 15%. For the next part of this project, we will only include these 15% of the OB-GYN doctors' activities to compare them with the activities at the midwife clinic.

## Data pre-processing of the midwife clinic's data

Some of the codes for the midwife clinic are 'fee-for-service', but others are 'bundled payments'. 'Bundled payment' means that multiple activities are billed together. As a result, the level of granularity for activities in the midwife clinic's data is more diverse than the claims data generated by OB-GYN doctors.

For example, there are multiple activities covering different stages of prenatal care. One activity covers the first 14 weeks of pregnancy, another covers care between 15-29 weeks, and another covers prenatal care beyond 29 weeks. These activities are bundled payments and typically represent more than one physical consultation over a longer period of time. Moreover, these stage-based activities do not follow each other as a process. Instead, they indicate that the pregnancy was only partially treated by the midwife clinic and later referred to an OB-GYN doctor or terminated. Thus, a case with a bundled payment claim for prenatal care for the first 14 weeks is unlikely to have a separate claim for the activity beyond 29 weeks. Patients that undergo the whole OB process at the midwife clinic are recorded with a separate code: 'Complete natal care'.

Although such bundled payment events typically cover multiple weeks or even months, the timestamp in the event log merely records the last day of treatment of that bundled payment. For example, the activity 'Complete natal care' will only have one timestamp reflecting the date of birth. In reality, however, it represents multiple months of work by the midwife clinic (up to nine months). There is nothing we can do about this limitation, but we need to keep this data property in mind when we interpret the process maps later.

Furthermore, there are many different bundled payment descriptions for similar activities. For example, the data contains the descriptions 'Maternity care 1 day', 'Maternity care 2 days', and 'Maternity care 3 days' (see Table 1). All three codes belong to maternity care (care delivered at home for a few days after childbirth). However, without further pre-processing, these bundled payments would show up as three separate activities in the process map.

Original activity	Newly created activity (higher level categorization)
Maternity care 1 day	Maternity care
Maternity care 2 days	Maternity care
Maternity care 3 days	Maternity care

Table 1. Example of bundled payments re-arranged to higher-level categorization

To avoid a process map with many different but similar activities, we have grouped several bundled payments into higher-level categories. For example, the three descriptions in Table 1 were assigned to the category 'Maternity care'. Similarly, we have grouped multiple types of prenatal care into a higher-level activity 'Prenatal care'.

## Combined data set

The extracted 15% of OB-activities of the OB-GYN doctors are vertically appended [3] to the dataset with the claims data from the midwife clinic (after applying higher-level categorizations to some events as explained above). By definition, all event log data generated by the midwife clinic is OB-related.

A distinct count created in a pivot table in Excel shows a degree of overlap between the two entities (see Table 2). This is expected because the midwife clinic refers complicated cases to the OB-GYN doctors and the OB-GYN doctors refer uncomplicated cases to the midwife clinic. So, patients flow between these entities. Therefore, the Total is lower than the sum of the OB-GYN Doctor and the Midwife Clinic counts because many patients are treated by both types of providers.

Year	OB-GYN Doctor	Midwife Clinic	Total
2018	2,132	1,265	2,452
2019	1,950	1,282	2,341
2020	1,641	1,146	2,042

Table 2. Total volume of clients (distinct count)

The event log containing both the 15% extracted OB-data and the midwife clinic's event data is imported into Disco. You can find a sample of the event log in Table 3.

The unique case identifier and timestamp are labeled according to conventional process mining logic [4]. However, the activities are labeled in a slightly different way. Because we want to know who is who in the process map, the activities are concatenated with the type of provider. Thus, a consultation by the midwife clinic (MC) will appear in the process map as 'MC-consultation', while a consultation by an OB-GYN doctor (OB) will appear in the process map as 'OB-

consultation'. This concatenation can be done in Disco by simply labeling both the activity and the 'Type of provider' column as activities [5].

Process mining input	Case	Time Stamp	Resource (concatenated with activity)	Activity
Column name in event log	Patient ID	Date	Type of provider	Description
Example event 1	123	February 19, 2020	MC	Ultrasound
Example event 2	532	March 4, 2020	OB	Caesarian Section

Table 3. Sample of the event log

## Analysis results

We have created two different process maps based on this event log.

The first process map describes the entire process, including both OB doctor and midwife clinic cases. This process map is called 'Total obstetrical care' and covers all the SVB population's obstetrical care. Activities performed by the OB-GYN specialist are labeled as OB in red and midwife clinic activities are labeled as MC in orange (see Figure 3). Keep in mind that the patient journey process for obstetrical care is not linear. There are several beginnings and endpoints possible.

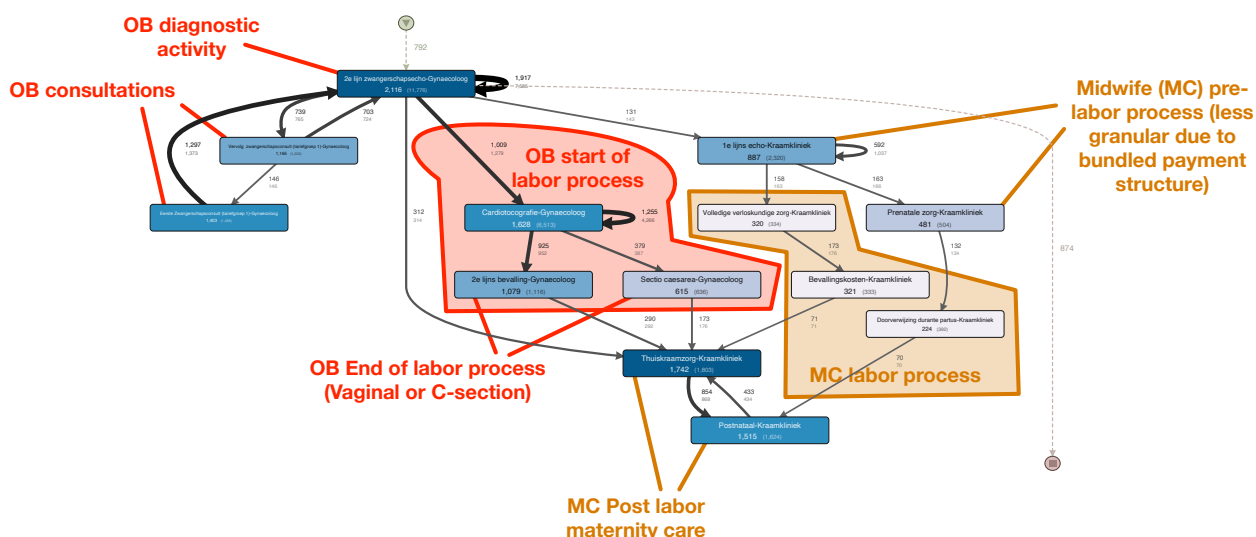


Figure 3. Total obstetrical care (Primary metric: Case count, Secondary metric: Frequency count)

When we look at the process map for the total obstetrical care in Figure 3, we see that the diagnostic activity at the very top that almost all cases appear to undergo is a pregnancy

ultrasound by the OB-GYN specialist ('2e lijns zwangerschapsecho'). It is important to note that, at around 20 weeks of pregnancy, all patients are expected to undergo at least a single ultrasound at the OB doctor to scan for any serious defects. For many cases this also appears to be the start of the process.<sup>1</sup> The second most common activity is postnatal maternity care delivered by the midwife clinic. This is also the endpoint for many cases.

The left side of the process map depicts the OB doctor's process, whereas the right side of the process map depicts the midwife clinic's process. In the OB doctor's process map, we can distinguish diagnostics and consultations on the one hand and the actual labor process on the other hand. The same is true for the midwife clinic. Both processes converge towards postnatal maternity care delivered by the midwife clinic.

The second process map only describes the cases that had at least one interaction with the midwife clinic (see Figure 4). We filtered on the resource 'Type of healthcare provider' and specified to only include cases that go through a specific activity at the midwife clinic (in Disco, this filter is called Attribute > Filter by: Activity > Select 'Ultrasound by midwife clinic'. Filtering mode: 'Mandatory' [6]). As a result of applying this filter, we get a more detailed view of the process flow for the clients of the midwife clinic.

Note that we have not just filtered for any mandatory midwife clinic activity. This is because many cases end with maternity care by the midwife clinic, even if the midwife clinic was not involved throughout the pregnancy. We are specifically interested in patients that at some point before labor had an interaction with the midwife clinic. The activity 'Ultrasound by midwife clinic' (in Dutch: '1e lijns echo') is a good filter activity to identify cases that, at least initially during the early stages of pregnancy, were deemed suitable to be handled by the midwife clinic.

The start for most cases is the ultrasound activity at the midwife clinic (see 'MC ultrasound' in Figure 4). In reality, the process does not start with this activity. The midwife clinic will have conducted some consultations already before this activity. Those consultations are reflected in one of the bundled payment packages of which the timestamp does not reflect the beginning but rather the end. Nevertheless, the timestamp for the ultrasound echo by the midwife clinic is a single activity and does approximate the early stages of pregnancy. Like the prior process map, the process ends with postnatal maternity care.

An important observation in this process map is the distinction between "no rush" OB doctor care and "rushed" OB doctor care. "No rush" typically means that the OB doctor has seen the patient before labor in consultations and diagnostic tests during pregnancy. "Rushed" implies that the patient is transferred to the OB doctor during labor. In such a case, the OB doctor may see that patient for the first time when she is in labor, or at the very least has only seen the patient earlier for a one-time ultrasound, lacking any regular prior consultations.

---

<sup>1</sup> This is actually not completely true: The bundled payments by the midwife clinics skew the timestamps and confuse the process map as they do not represent the start date but rather the end date. Furthermore, the OB doctor will often do consultation and ultrasound in the same sitting. These activities will then have the same timestamp, further confusing the process map unless these activities are explicitly sorted [7]. Nevertheless, the ultrasound echo by the OB doctor still overshadows the OB doctor consultations in terms of sheer volume.

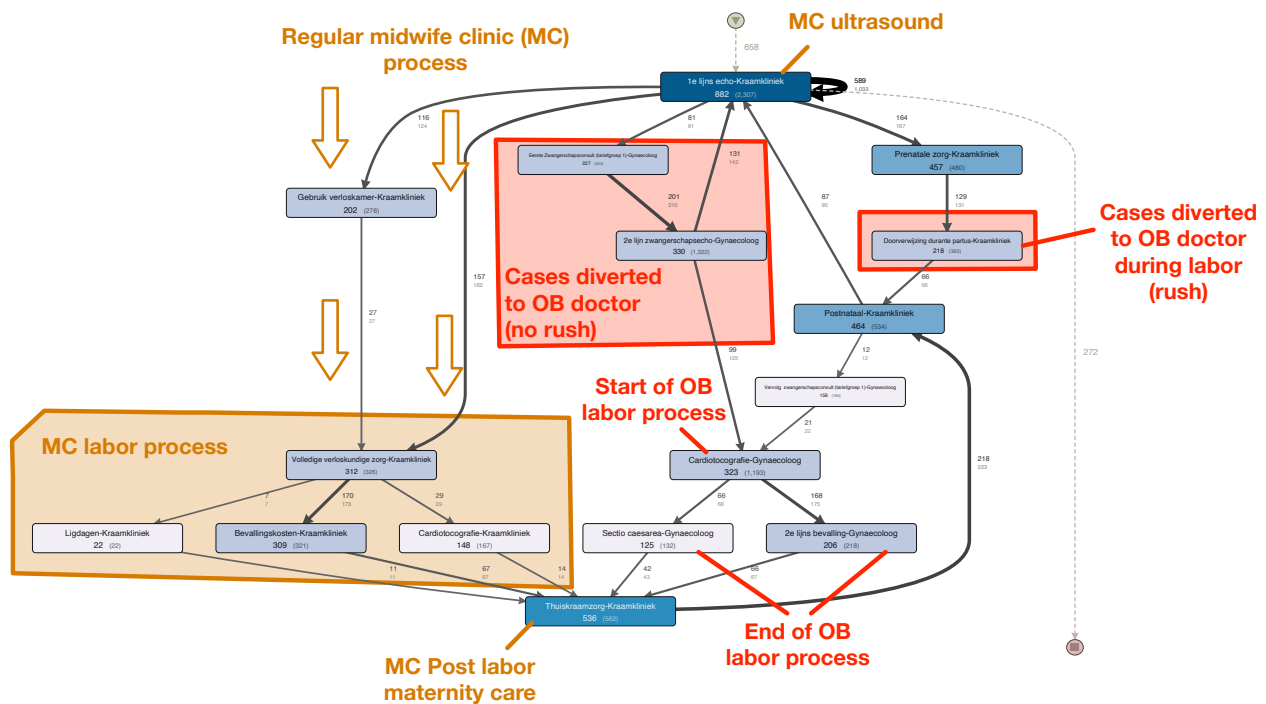


Figure 4. Obstetrical care with involvement of the midwife clinic

The CTG scan activity ('Cardiotocografie') indicates the start of the labor process under the supervision of the OB doctor. From this activity, we can discern that about a third of cases are "fed" to the OB doctor from the "no rush" process, while two thirds are coming to the OB doctor from the "rushed" process.

In the process map in Figure 4, the actual deliveries are split roughly 50/50 between the midwife clinic and the OB doctors. This means that there is a 50% chance of delivering at the OB doctor for any patient who starts at the midwife clinic. Of these OB doctor deliveries, two thirds are actually initiated by the midwife clinic themselves as part of the 'rushed diversion during labor'. Only about one third of cases that undergo the actual delivery at the OB doctor appear to "drift" towards the OB doctor under non-rushed circumstances. The OB doctor could have persuaded these patients to stay with them, but it can also be the case that they were labeled as high risk and, therefore, were transferred to the OB doctor weeks before the actual delivery (thus, "no rush").

When we look at the entire data set again, the OB doctors performed 1,694 deliveries during the research period. The midwife clinic provided ultrasound services for merely 887 patients. Around 330 of them were diverted to the OB Doctor (of which about 67% at the initiative of the midwife clinic itself during labor) and 320 delivered at the midwife clinic. The remaining 26% of these 887 cases delivered outside the research period.

## Conclusion

Our findings suggest that the OB doctors are not actively trying to “steal away patients” from the midwife clinic. In fact, most deliveries by the OB doctor for patients originating from the midwife clinic (around 67%) appear to be last-minute rushed transfers initiated by the midwife clinic (diversion during labor). From the 33% of the cases that are diverted to the OB doctor under non-rushed conditions, most likely at least a portion of these cases will be legitimate transfers, such as pregnancies that have been classified as high-risk weeks before the delivery date.

The claim that OB doctors “steal patients” from the midwife clinic cannot be substantiated by the data. Most patients never set foot in the midwife clinic and can thus not be “stolen” by the OB doctors. The midwife clinic is called upon only after delivery for postnatal maternity care. On the other hand, there seems to be little evidence that OB doctors refer uncomplicated cases that start their process at the OB doctor to the midwife clinic. So, the challenge for the midwife clinic is not necessarily retaining patients in their system but rather acquiring them in the first place.

There may be some room for OB doctors to refer patients to the midwife clinic. Currently, this does not seem to be the case: Patients rarely start at the OB doctor and flow to the midwife clinic. The other way around is much more common.

## References

- [1] Simplify Complex Process Maps: <https://processminingbook.com/read/simplification/>
- [2] Removing “Spider” Activities: <https://processminingbook.com/read/simplification/#strategy-8-removing-spider-activities>
- [3] Combine Data Sets of the Same Shape: <https://fluxicon.com/blog/2018/11/process-mining-transformations-part-3-combining-data-sets-of-the-same-shape/>
- [4] Import Configuration Settings: <https://processminingbook.com/read/import/#import-configuration-settings>
- [5] Combined Activity: <https://processminingbook.com/read/perspectives/#combined-activity>
- [6] Attribute Filter: <https://processminingbook.com/read/filtering/#attribute-filter>
- [7] Same Timestamp Activities: <https://fluxicon.com/book/read/dataquality/#same-timestamp-activities>