

Between Capacity and Care

Understanding and Predicting the Risk of Long Length of Stay Associated with Hospital Overflow Bed Exposure

T. Bentvelsen

TU Delft Delft University of Technology

Master of Science thesis

Engineering and Policy Analysis

November 2026

Between Capacity and Care

Understanding and Predicting the Risk of Long Length of Stay Associated with Hospital Overflow Bed Exposure

By

T. Bentvelsen

In partial fulfilment of the requirements for the degree of:

Master of Science
In Engineering and Policy Analysis

at the Delft University of Technology,
to be defended publicly on Monday, November 9, 2026 at 16:00



Thesis committee:	Dr. P.J. Marang-van de Mheen ,	1st Supervisor & Chair, TU Delft
	Dr. N. Mouter,	2nd Supervisor, TU Delft
	Dr. B. Filtjens,	Advisor, TU Delft

An electronic version of this thesis is available at http://repository.tudelft.nl/.

TU Delft

Preface

Throughout your studies, you hear words like complex systems, systemic view, problem framing and implications. But it was only during the thesis that I discovered what a programme like Engineering and Policy Analysis (or studying in that sense) is really about, as I believe you only truly learn it once you have to work with it. In the end, for me it is about becoming aware of a complex system and finding the place within it where you can make a contribution, genuinely trying to add something that others can act on. After all, you can only make a meaningful contribution once you understand the system you are part of.

That was also the biggest challenge for me: navigating a thesis, with all the rabbit holes you could go down, while still delivering a focused thesis that answers real-world questions. However creative or interesting a method may be, it is only useful if it can be implemented and clearly communicated. Working on it in practice enabled me to learn a lot about this, and I can safely say that the thesis has been the most valuable learning experience of my studies at the TU Delft.

I would like to thank Perla. At the start of this project I was still searching: I wanted to do too much in too little time and found it difficult to narrow down the scope. You helped me keep my thesis focused and not lose myself in solutions in search of a problem. Only now do I fully realize that a thesis is much more about the process than about the content, because without the right narrative, focus and communication, the content cannot come across. You often left me with a head full of ideas to reflect on and continually challenged me. Thank you for all the time you made for my thesis. I would also like to thank Niek and Benjamin for the time and effort they put into my thesis. Your feedback often helped me look at my work from a new perspective.

I would also like to thank my external supervisor at the hospital, [redacted]. Thank you for your substantive feedback, for connecting my work to daily operations, and for reading every first (and often unstructured) draft. You regularly reminded me to think about how to communicate my work and how it could be used in practice. And, perhaps even more, thank you for taking me into the world of the hospital for six months. I am very grateful for this experience, and it would not have been the same without [redacted] and [redacted] and the other staff of the hospital. Thank you for the warm welcome and the time I could be part of the [redacted]. Ping-pong victories, trips to the local cake shop and surviving heat waves together made sure my thesis never became a grind and made me very much enjoy the time spent at the hospital.

Finally, I would like to thank my family, friends and (former) roommates. Thank you for the draft reading, welcome distractions and for your support! But mostly, for putting things into perspective (sometimes without even knowing it). I tend to lose myself in a project, and it is at (or after) those moments that I realise how lucky I am to have the right people around me.

With my graduation from EPA, my time as a student comes to an end. I have enjoyed these years immensely, filled with friendships and memories. I will look back on this period with great gratitude. At the same time, I am looking forward to the next chapter, with no idea yet where it will take me.

Tim Bentvelsen,
Delft, November 2026

Summary

Dutch hospitals face a growing demand for care, while resources are limited. Because beds and staff are costly, hospitals cannot base the capacity of each ward on peak demand. When the ward that matches a patient's diagnosis is full, the patient can as an alternative be placed on an overflow bed: a bed on a ward whose staff are not specialised in the patient's condition. But this can reduce the quality of care, for example because staff are less familiar with the condition and physicians review these patients less often and later in the day. Whether overflow beds are an acceptable way to absorb peaks depends on how large the risks associated with overflow placement are.

This thesis is the third study on overflow beds in the same Dutch hospital, after Mans (2025), who reconstructed how placement decisions are made, and Vloet (2025), who found increased odds of mortality for patients placed on overflow beds. As death is a rare outcome, most of the harm that overflow placement may cause does not show in mortality. This thesis therefore adds a second one: a long length of stay (L-LOS). For the patient, a long stay can signal that care did not go as planned, for example after a complication. For the hospital, it keeps a bed occupied that the next patient needs, which adds to the very shortage that overflow beds are meant to relieve. The thesis had two objectives: to gain a more complete understanding of the risks associated with overflow placement, and to explore how placement decisions can be supported with data. The main research question was:

To what extent is overflow bed exposure associated with long length of stay, and how can this risk be predicted alongside in-hospital mortality?

The study analysed admissions between 2022 and 2025. An admission was counted as long when it was among the longest quarter of stays for patients with a similar diagnosis, and all comparisons were adjusted for differences case mix. Phase 1 estimated the risk of L-LOS along four dimensions of exposure: whether a patient was exposed, for how long, at which moment and on which ward, the latter evaluated against the hospital's flowcharts, which indicate the most suitable alternative ward when the appropriate ward is full. Phase 2 examined whether mortality distorts this association, and Phase 3 tested whether more flexible machine learning models, which can capture complex relationships between patient characteristics, predict both outcomes at admission better than logistic regression.

Overflow exposure was associated with a clearly higher risk of a long stay: after adjustment for case mix, about one in three exposed admissions had a long stay, against one in four comparable admissions without exposure. This association held under a stricter definition and was hardly affected by mortality. The increase, however, differed between forms of exposure. Exposure at the start of an admission, the most common form, carried a modest increase. The largest increase lay in brief exposures later in long admissions, most likely after a change of specialty or a need for care on another ward, while a longer exposure did not go together with a larger increase. Placements on the preferred alternative wards of the flowcharts showed the smallest increase, and the ICU as a last resort a clearly larger one. One in six exposed admissions fell outside the flowcharts, and these mostly did not result from a full ward.

For prediction, the limit lies in the data rather than in the model. At admission, the patients at high risk of a long stay were largely the same as those at high risk of dying. None of the machine learning models clearly outperformed logistic regression, and mortality was predicted considerably better than L-LOS. Admission data describe how ill a patient is on arrival, whereas whether a stay becomes long depends

largely on what happens afterwards. Better prediction of L-LOS therefore requires information on the course of the admission rather than a more complex model.

For the hospital under study, these findings broaden the picture of the risks associated with overflow placement beyond mortality. If part of the association reflects an effect of the placement itself, longer stays may cancel out part of the capacity gained by using overflow beds, and exposures that did not result from a shortage of beds should be left out when overflow is evaluated as a capacity problem. The flowcharts hold up in the data, but exposure later in the admission falls outside the current measures; giving these patients priority in the daily Bed Availability Consultation may limit the part of the risk that stems from the overflow bed. For decision support, a single mortality model, such as that of Vloet (2025), already identifies the patients at highest risk of both outcomes, provided it is recalibrated over time. Its output then signals more than the risk of death alone: a likely complicated course and a bed that is likely to remain occupied for longer.

As demand for care grows, overflow beds are likely to remain part of how hospitals absorb peaks. By showing where the association with a long stay arises and which decisions can influence it, this thesis helps the hospital to base these decisions on data alongside clinical judgement. A potential next step is to learn why the increased risks arise, to see which mechanisms actually occur in the hospital under study

Keywords: *Overflow Bed, Outlier Patient, Long Length of Stay, Hospital Capacity Management, Patient Safety, In-Hospital Mortality, Clinical Prediction Model*

Content

Preface ... 3
Summary ... 4
1 Introduction ... 8
1.1 Problem statement ... 8
1.2 Research objective ... 9
1.3 Relevance to EPA ... 10
1.4 Overview of the thesis ... 10
2 Scientific Relevance and Research Gaps ... 11
2.1 Search method ... 11
2.2 Overflow bed practice in general and in hospital under study ... 12
2.3 Mechanisms how overflow bed placement may affect quality of care ... 14
2.4 Overflow exposure and length of stay ... 15
2.5 Interaction of LOS & Mortality ... 20
2.6 Prediction models for L-LOS and mortality ... 22
2.7 Research gaps ... 24
3 Study Design ... 25
3.1 Research questions ... 25
3.2 Research approach ... 27
3.3 Data preparation ... 27
3.4 A.I. Disclosure ... 34
4 Research Methodology ... 35
4.1 Length of stay and overflow exposure in the study population ... 35
4.2 Phase 1 ... 36
4.3 Phase 2 ... 40
4.4 Phase 3 ... 41
4.5 Sensitivity Analysis ... 43
5 Results ... 44
5.1 Length of stay and overflow exposure in the study population ... 44
5.2 Phase 1 ... 47
5.3 Phase 2 ... 54
5.4 Phase 3 ... 57

5.5 Sensitivity analysis 59
6 Discussion 60
6.1 Key findings 61
6.2 Limitations 61
6.3 Interpretation & addition to literature 64
6.4 Implications for hospital under study 66
6.5 Recommendations for future research 67
7 Conclusion 69
References 71
A Description of dimensions 75
B Charlson's comorbidity groups 76
C Metrics for model performance and evaluation 77
D Stratified L-LOS models 79
E Model hyperparameter settings 80
F Data preparation sensitivity analysis L-LOS95 81
G Recoding case-mix covariates 83
H Sensitivity, FPR & Precision of LR predictions 85

1 Introduction

1.1 Problem statement

There is growing pressure on the Dutch healthcare system, driven by an ageing population. The number of people aged 65 and over is expected to rise from about 3,6 million to 4,8 million by 2050, and with it the number of people living with one or more chronic conditions (RIVM, 2025). At the same time, the availability of staff is not keeping pace. While about one in seven workers is currently employed in healthcare, meeting the projected demand would require one in three within forty years (Wetenschappelijke Raad voor het Regeringsbeleid, 2021). As a result, not all care that is needed may remain deliverable (RIVM, 2025). These constraints, combined with growing demand, mean that available capacity must be managed as efficiently as possible.

Hospitals must balance the availability of beds and care professionals against the inflow of patients, so that, ideally, a suitable bed is ready whenever one is needed. A suitable bed is not simply any free bed. It is a bed on a ward (a nursing unit within the hospital) whose staff have the expertise and equipment to care for the patient's condition. Hospital capacity is therefore specialty specific. For elective care (or planned care), this balance can largely be controlled. For emergency care, whose inflow is unplanned and fluctuates, it is much harder. Some peaks can be anticipated to an extent, such as a seasonal influenza epidemic or holiday periods. Others are unexpected, such as an extreme heatwave or when patients are diverted from neighbouring hospitals that have reached their capacity.

When no bed is available on a patient's appropriate ward at the moment they need to be admitted, they may instead be placed on an overflow bed: a bed on a ward that, based on their diagnosis, does not match their care needs. An example is a patient with cardiac problems who is placed on a neurology ward. Although such a patient remains under the medical responsibility of physicians from their own specialty, they are physically located on a clinically inappropriate ward and receive their day-to-day nursing care from the staff of that ward. This thesis studies overflow placement in a Dutch hospital, hereafter referred to as the *hospital under study*. In this hospital, roughly one in twenty admissions between 2022 and 2025 involved at least one overflow bed episode.

Hospitals manage capacity across several planning horizons. These range from long-term strategic decisions about how beds are allocated between specialties, to shorter-term tactical decisions, down to the day-to-day operational placement of patients (Mans, 2025). Overflow beds are used to absorb peak demand when structural capacity planning cannot prevent a shortage (La Regina et al., 2019).

Yet, overflow practices can come with a certain risk. Interviews with hospital staff identified five mechanisms through which overflow placement can compromise patient safety (Goulding et al., 2012): competing and different demands on staff time, poor communication between the responsible specialty and the overflow ward, a lack of knowledge about the patient's condition among overflow ward staff, an unsuitable ward environment, and the assumption that overflow patients are stable because they were selected to be moved. Section 2.3 discusses these mechanisms in more detail.

The alternative to overflow bed practices is to give wards more structural capacity, so that fewer patients need to be placed on overflow beds. This alternative can be costly: it requires resources and staff that, given the shortages described above, are often not available. Whereas a limited amount of flexible capacity can absorb most peaks (Bekker et al., 2017). Whether overflow beds are an acceptable way of absorbing peaks, or whether structural capacity is worth these costs, therefore depends on how large the risks of overflow placement are for patients.

However, it is unclear how large these risks are: the quantitative evidence is inconsistent. A systematic review concluded that the literature cannot yet draw firm conclusions about the quality and safety of care for patients on overflow beds (La Regina et al., 2019). Next to that, the consequences of overflow placement depend on how a hospital organises its wards and its overflow practice, so estimates from one hospital cannot simply be generalised to another. The hospital under study has, for example, introduced flowcharts to determine the most appropriate alternative ward for a patient in order to mitigate risks that come with overflow practices.

Within the hospital under study, the risks of overflow placement have so far only been quantified for in-hospital mortality (Vloet, 2025). A more complete picture of these risks, including outcomes beyond mortality, could provide relevant information for both decisions described above: whether overflow beds are an acceptable way of absorbing peaks compared with expanding structural capacity, and where an individual patient can best be placed when their appropriate ward is full. With this information, the hospital can make both choices in a better informed and more deliberate way.

1.2 Research objective

This thesis has been conducted in the same hospital as the work of Vloet (2025) and Mans (2025). It has two objectives, which correspond to the two problems identified above.

The first objective is to gain a more complete understanding of the risks associated with placing patients on overflow beds in the hospital under study. Findings from the literature are inconsistent and cannot easily be generalised to other hospitals, as they depend on local overflow practice. Within the hospital under study, Vloet (2025) examined these risks for in-hospital mortality. Mortality captures the most severe harm a patient can suffer and is an established indicator of hospital quality. It is, however, a relatively rare outcome, so most of the harm that overflow placement may cause does not end in death.

This thesis therefore adds a second outcome: long length of stay (L-LOS), a stay that is long compared with other patients with the same diagnosis. For the hospital, a long stay keeps a bed occupied that the next patient needs, which links this outcome directly to the capacity problem that overflow beds are meant to solve. For the patient, it can capture harm that prolongs an admission without ending it. Patients who receive poorer care often stay longer (Thomas et al., 1997), and patients in Dutch hospitals who experience an adverse event stay on average five days longer (Hoogervorst-Schilp et al., 2015). Staff at the hospital under study name a longer stay as a direct consequence of overflow

placement: because overflow patients are reviewed later in the day, tests and treatment shift to the next day (Mans, 2025). Because death ends an admission, the two outcomes cannot be interpreted independently of each other: a patient who dies early in the admission cannot have a long stay. This thesis therefore also examines to what extent mortality influences the association between overflow bed exposure and L-LOS.

The second objective is to explore how decisions on overflow bed placement can best be supported with data. When a patient cannot be placed on the appropriate ward, hospital staff use ward-specific flowcharts to determine the most suitable alternative. As Mans (2025) found, these flowcharts are based on the clinical judgement of care providers rather than on data. A risk estimate available at the moment of placement could complement them. Vloet (2025) took a first step by developing logistic regression models that predict the risk of in-hospital mortality for patients placed on overflow beds. This thesis extends that work in two ways: by investigating L-LOS as a metric alongside mortality, and by examining whether more complex machine learning models predict these outcomes better than logistic regression based on the data available.

1.3 Relevance to EPA

Hospital care under growing capacity pressure is the kind of grand societal challenge the Engineering and Policy Analysis (EPA) programme concerns itself with. An ageing population and staff shortages force hospitals to keep care both safe and accessible. Overflow bed placement is a good example of this tension. It allows a hospital to admit a patient when the appropriate ward is full, but it may do so at a cost to that patient. The surrounding system is a complex multi-actor, multi-level system. The Patient Flow Coordinator who authorises a placement, the nurses and physicians on the receiving ward, the responsible specialty and the patient each hold a different stake. And decisions are made at strategic, tactical and operational levels.

The underlying trade-off, between operational efficiency, patient safety and the availability of care, is currently navigated with flowcharts based on clinical judgement rather than data (Mans, 2025). This thesis quantifies the risks in a data-driven manner. Doing so means working with complex, routinely collected hospital data that were not recorded for research, and translating the resulting insights into a form that can support operational placement decisions. Interpreting the results therefore requires an understanding of the system in which overflow placement takes place.

1.4 Overview of the thesis

This thesis consists of seven chapters. Chapter 2 reviews the literature this thesis builds on and ends with the research gaps this thesis addresses. Chapter 3 turns these gaps into the main research question and uses subquestions to structure the answer. Next it outlines the research approach and describes the data preparation needed for the analyses. Chapter 4 describes the method of each of the three phases of the analysis (each answering one subquestion) and Chapter 5 presents the results in the same order. Chapter 6 discusses the key findings and the limitations of the research, how the findings address the research gaps, recommendations for future research, and the implications for the hospital under study. Chapter 7 concludes the thesis by answering the main research question. The appendices contain supporting material.

2

Scientific Relevance and Research Gaps

This chapter reviews the existing knowledge relevant to the objectives of this thesis. Figure 2.1 gives an overview of its structure. Section 2.1 describes how the literature was searched. Sections 2.2 and 2.3 focus on overflow in practice: Section 2.2 describes how overflow placement is organised, both in general and in the hospital under study, and Section 2.3 discusses the mechanisms through which overflow placement may affect the quality of care. Sections 2.4 and 2.5 focus on quantifying risk: Section 2.4 reviews the quantitative evidence on overflow exposure and length of stay, including why studies reach different conclusions, and Section 2.5 explains why length of stay cannot be interpreted without taking mortality into account. Section 2.6 turns from quantifying risk to predicting it, comparing logistic regression with machine learning models. Section 2.7 brings these findings together into the research gaps that this thesis addresses.

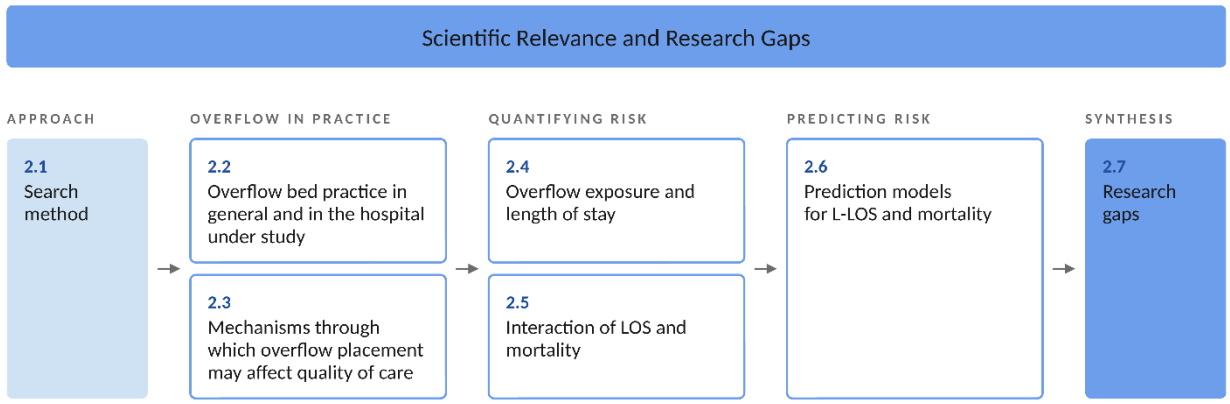

Figure 2.1: Overview of the structure of Chapter 2.

2.1 Search method

For the search of quantitative studies, the work of La Regina et al. (2019) and Vloet (2025) formed the starting point for the additional literature search on length of stay conducted for this thesis. The search combined terms for overflow bed exposure (*"bedspacing", "medical outlier", "outlier patient", "outlying*

patient", "overflow bed") with terms for quality and safety outcomes *("safety", "quality", "patient outcome", "adverse event", "risk")*, extended with length-of-stay-specific terms *("LOS", "Length of Stay", "Hospital stay").* This search was run in Scopus and PubMed. This resulted in 19 papers from PubMed and 29 from Scopus.

Titles and abstracts were then screened. Some papers used "outlier" to describe a patient who stood out on length of stay or another quality indicator for reasons unrelated to ward placement. Some examined a specific, narrowly defined length of stay (e.g. number of days following a particular operation) that was not comparable to the whole-admission LOS measure used here. A number of studies focused on very specific populations or crisis periods, such as the COVID-19 pandemic, or on specific ward-to-ward transfers (for example, between two ICUs) where different rules applied than standard overflow bed placement. Finally, some papers studied the probability of a patient being placed on an overflow bed rather than the consequences of overflow bed exposure for length of stay.

After applying these exclusion criteria, 7 documents from Scopus and 9 from PubMed remained for full-text analysis. Of these, 4 were identified in both databases, leaving 12 unique articles that form the overview of length of stay.

The remaining sections draw citation searching from the included studies and the review of La Regina et al. (2019), targeted searches on (amongst other things) LOS as a quality indicator, the competing risk of death, clinical prediction, documents from DHD and CBS and on the literature identified by Vloet (2025) and Mans (2025).

2.2 Overflow bed practice in general and in hospital under study

Across hospitals, the process of assigning a patient to an overflow bed generally follows a comparable structure. First, a clinical decision is made that the patient requires admission. Second, the availability on the patient's own appropriate ward is checked. If no bed is available there, an alternative ward is identified. Third, a decision-maker authorizes and coordinates the transfer. Finally, because overflow placement is intended to be temporary, patients are expected to be moved back to their home ward once a bed becomes available there. At the hospital under study, Mans (2025) reconstructed the daily decision-making process for overflow bed placement in an MSc thesis, based on interviews and internal documentation. The remainder of this section draws on that work.

The flow of patients in the hospital

In the hospital under study, acute patients enter the inpatient clinic through two routes: mainly via the Emergency Department (ED), and less often via the outpatient clinic when a patient turns out to need more extensive care. Planned admissions are usually placed directly on the appropriate ward and can be rescheduled when capacity is under pressure. Overflow placement is therefore mainly an issue of acute care.

Most acute patients admitted through the ED first go to the Acute Medical Unit (AMU), a ward between the ED and the specialty wards. Exceptions are made when a patient's condition requires the care of the appropriate ward directly. The AMU specialises in the diagnostic phase of an acute admission, and for the first 48 hours of an admission it is therefore considered the most appropriate place for the patient. Patients there receive care from the medical team in line with their diagnosis, even though they have not yet been admitted to a specialty ward. It thereby also relieves the specialty wards of the extra workload that the first hours of an admission bring. Patients stay on the AMU for a maximum of 48 hours, after which they are either discharged or transferred to their appropriate ward. This creates an

additional decision moment. Comparable units are described in the literature as a way to relieve capacity problems such as overflow placement (Santamaria et al., 2014; Serafini et al., 2015).

Decision making around overflow beds

Decision-making around bed capacity (and therefore overflow placement) occurs at different organizational levels. At the strategic level, hospital-wide capacity policies are set. These include determining the bed count for each ward and negotiating maximum care agreements with health insurers. At the tactical level, every 2 weeks, a Tactical Planning Meeting reviews bed capacity, staffing levels and discusses potential solutions to issues. These meetings alternate between evaluating the use of capacity in the month before and looking forward to how capacity is used in the upcoming month. The daily, individual placement decisions take place at the operational level. On this level, the Patient Flow Coordinator (PFC) functions as the hospital's central decision-maker for overflow bed placement using a ward-specific flowchart to guide these decisions.

Mans (2025) described the decision making as a trade-off between operational efficiency (ensuring every patient has a bed) and clinical suitability (placing the patient in the appropriate specialty ward). The PFC becomes particularly important when hospital occupancy is high. Their role includes maintaining an overview of bed availability across wards and working to optimize placements.

Mans (2025) described 4 situations in which patients can be placed on overflow beds. The most common cause is simply a lack of space when the appropriate ward is at full capacity. Next to that, a shortage of specific single rooms (such as those needed for isolation or palliative care) can create a necessity for placement on another ward. It also occurs that a patient’s primary medical team changes during their stay, but the patient physically remains on their initial ward. Finally, beds strictly reserved for planned care can limit the flexibility and capacity for unexpected emergency admissions.

The main tool supporting overflow placement is a set of ward-specific flowcharts. For each specialty, the flowchart indicates the preferred ward and a defined route to the best available alternative, based on similarities in the care delivered between wards. The flowcharts also list the diagnoses that exclude a patient from being placed as an overflow patient on a given ward. They are complemented by a bed policy document, which prescribes, among other things, that overflow patients are clustered by room, ward or floor where possible, and that they are transferred back to their appropriate ward as soon as possible (Mans, 2025). Mans highlighted that decisions regarding overflow bed placement were primarily guided by the flowcharts developed through the clinical judgment of care providers.

Beyond individual placement decisions, the hospital also holds a daily Bed Availability Consultation (BAC), chaired by the PFC and attended by representatives of the hospital wards. This meeting is used to review current occupancy, discharges that can be anticipated, and to raise clinical concerns about patients on overflow beds. Including the quality of care they receive, so that their transfer back to their home ward can be prioritized.

Mitigations to the risks of overflow patients

When overflow placement cannot be avoided, the hospital takes several measures to limit its risks (Mans, 2025). The first concerns which patients are placed on overflow beds: patients with complex care needs are kept off them as much as possible, through the exclusion criteria in the flowcharts and the judgement of the PFC. The second concerns how long exposure lasts. Overflow patients are moved back to their own ward as soon as a bed becomes available there, unless their discharge is expected within one or two days, in which case an additional transfer is considered more burdensome than the

remaining stay on the overflow bed. The third addresses the lack of specialty knowledge on the receiving ward. Physicians support the nurses caring for overflow patients with explicit, step-by-step care plans. And lastly, in some cases a dedicated resident is assigned to the overflow patients of a specialty.

2.3 Mechanisms how overflow bed placement may affect quality of care

This section examines why placement on an overflow bed might affect the quality of care at all. Qualitative studies among staff, patients and nurses identify several mechanisms, which help to interpret the quantitative findings that follow.

In a study to investigate the mechanisms, Goulding et al. (2012) interviewed care workers and identified five factors for reduced patient safety on inappropriate wards. The *first* was competing demands on staff time: nursing and medical staff at the same time must care for overflow patients as well as other patients in that ward. They describe being pulled between the two groups with broadly different needs. Ward rounds were also typically structured so that overflow patients were reviewed last, creating a possible delay through the rounds for investigation, diagnosis and treatment. *Second*, poor communication between the correct specialty ward and the overflow ward. Because overflow patients are not looked after by a team of nurses and doctors who know each other and work together often, informal channels of communication broke down. *Third*, staff reported a lack of knowledge or expertise on the overflow wards, with nursing and junior medical staff sometimes unable to recognize warning signs of the worsening of conditions. *Fourth*, an unsuitable ward environment as a mechanism. It involves both the physical distance between overflow patients and the responsible medical team, and the unavailability of specific equipment and medication. *Fifth* and finally, staff described how the perceived characteristics of overflow patients could compromise care. Patients were often assumed to be medically fit and stable simply by the fact of them being moved, so that they were not prioritised even when unwell.

The mechanisms described by Goulding et al. (2012) are also recognized by staff in the hospital under study. In interviews by Mans (2025), nurses described less familiarity with the protocols and medication of other specialties. Physicians reported reviewing overflow patients less often, usually once a day, and later in the day, after the round on their own ward. Staff named some resulting risks: minor mistakes in care, delays in care, and a longer length of stay. They linked the last partly to the late ward rounds: when the treatment plan is determined late in the day, tests and treatment shift to the next day, and the patient stays longer than necessary (Mans, 2025).

The main measure of the hospital under study to mitigate these risks is the set of ward-specific flowcharts described in Section 2.2. They mainly address two of the mechanisms described by Goulding et al. (2012). Because the order of alternative wards is based on similarity of care, patients are placed on wards where staff are more familiar with their condition. In addition, the exclusion criteria keep patients whose diagnosis requires specialised equipment or expertise off overflow beds (Mans, 2025). The flowcharts do not address later and less frequent review, communication between teams or distance to the responsible team, all of which staff still reported (Mans, 2025). Song et al. (2020) found that distance, rather than clinical similarity, drove the longer length of stay of overflow patients, which suggests that placement on a preferred ward may still carry an increased risk.

Cheung et al. (2020) in their study tried to unpack the experience of nurses in more detail. They found that specialty-trained nurses experienced a decline of their clinical competence when caring for

patients on overflow beds. Nurses described becoming less able to align their care routines with the schedules of medical teams of the overflow patients, less familiar with disease equipment and procedures, less able to prioritize competing care demands, less able to anticipate complications or plan proactively, and struggling more to develop working relationships with the medical team responsible for the patient's care. The authors explain this using the metaphor of a "care ladder," in which a nurse's specialty-based competencies become broken or entirely absent when applied to overflow patients. This means that it takes nurses more effort to reach a baseline standard of care, and patient outcomes may be compromised as a result.

Goulding et al. (2015) turned to the patient perspective in an English hospital by interviewing patients who had been placed on overflow beds. Most patients accepted their placement on an overflow ward; a minority expressed clear dissatisfaction or a strong preference for the correct specialty ward. They described feeling forgotten or not prioritized, with gaps of multiple days without medical review. Patients were moved between wards with little explanation, and discharge planning was frequently disrupted because it was unclear which team held responsibility for their aftercare, showing the same problems staff had described in earlier research. Multiple participants also reported that nursing staff on the inappropriate ward were unfamiliar with their specific condition, and that equipment and medication were sometimes unavailable, requiring staff to look for items from elsewhere or delaying treatment.

2.4 Overflow exposure and length of stay

This thesis aims to add length of stay as a metric for risks associated with overflow stays. Several of the mechanisms described in Section 2.3, such as overflow patients being reviewed last on ward rounds and nursing staff being less familiar with their condition, give reason to expect that overflow placement may delay diagnosis and treatment. This section reviews what the quantitative literature has found on this relationship. First, however, it is worth considering what length of stay (LOS) actually measures.

Length of stay as a metric

LOS carries a dual meaning within healthcare analysis. It functions as a proxy for hospital efficiency, yet it can also be used as a metric for patient safety. From an efficiency standpoint, hospitals with long average stays are considered relatively inefficient in their resource use, while shorter stays are interpreted as a sign of efficient resource allocation (Thomas et al., 1997). From a safety standpoint, Thomas et al. found that patients who received poorer quality care had longer risk-adjusted hospital stays, thereby confirming that complications can extend a hospitalization.

A relation can also be found between adverse events and length of stay. Several studies found that adverse events, defined as *"an unintended injury that results in temporary or permanent disability, death or prolonged hospital stay, and is caused by healthcare management rather than by the patient's underlying disease process"* (Hoogervorst-Schilp et al., 2015) are associated with a longer LOS (Ackroyd-Stolarz et al., 2011; Sari et al., 2007). Hoogervorst-Schilp et al.'s study on adverse events in twenty Dutch hospitals revealed that patients who experienced an adverse event stayed an average of 5,11 days longer than those who did not.

Two things follow from this that matter for how the studies below should be read. First, a shorter LOS is not automatically the better outcome: Bueno et al. (2010) and Kaboli et al. (2012) found that discharging patients earlier than clinically expected was accompanied by higher 30-day readmission

rates. Second, sicker patients naturally require longer stays, an effect that can never be entirely corrected for through statistical adjustment (Han et al., 2022). Any difference in LOS between two groups of patients can therefore reflect the case mix of those groups as well as the care they received. In this thesis, a long stay is therefore not read as poor care in itself, but as a marker that an admission deviated from the usual course for its diagnosis, for example through complications, delays in diagnosis and treatment, or waiting for follow-up care.

Overview of quantitative studies

Table 2.1 provides an overview of the studies (identified through the search strategy above) that examined the relationship between overflow bed exposure and length of stay.

Study	Scope	Length of Stay	Methodology	Case-mix adjustment	Determination of Exposure
Alameda & Suárez (2009)	Single-centre study, Internal Medicine admissions restricted to heart failure	Longer LOS for overflow patients (mean 11,8 vs 9,2 days)	Multiple Linear Regression	Restricted to single diagnosis, adjusted for age, sex, diabetes, multiple secondary diagnoses, previous stay, laboratory values, weekend/week admission	Patients whose initial ward of admission was different from the Internal Medicine ward
Perimal-Lewis et al. (2013)	Single-centre study, General Medicine admissions	LOS reduced by approximately 23% for overflow patients	Poisson regression analysis	Adjusted for age, Charlson comorbidity index, gender, and time waiting for a bed in the Emergency Department	Patients spending ≥70% of admission outside their home ward
Stowell et al. (2013)	Single-centre study, mixed Medical/Surgical admissions	Longer LOS for overflow patients (8 vs 7 days)	Prospective matched-pair cohort study	1:1 matching on age (within 5 years), sex and reason for admission	Patients admitted to a ward outside their admitting specialty
Liu et al. (2014)	Single-centre study, General Internal Medicine admissions	No significant difference in LOS (4,9 vs 6,0 days)	Matched-cohort comparison	Direct 1:1 matching pairing each overflow patient with a patient admitted on the same call shift, under the same GIM attending physician, with the exact same GIM diagnosis	Patients admitted to a non-GIM inpatient ward
Serafini et al. (2015)	Single-centre study, Medical and Geriatric admissions	No significant difference in LOS	Descriptive comparison (no adjustment reported for LOS)	No adjustment for LOS. Mean LOS was compared directly between the groups, separately for Medicine and Geriatrics.	Patients admitted outside Medicine or Geriatrics
Perimal-Lewis et al. (2016)	Single-centre study, dementia/delirium admissions	Marginally shorter LOS for overflow patients	Logistic regression	Restricted to patients with a primary or secondary diagnosis of dementia. Adjusted for age, LOS, ED wait time, working hours, acute/non-acute, and Charlson Comorbidity Index	Patients spending ≥70% of admission outside their home ward
Stylianou et al. (2017)	Single-centre study, Medical admissions	Longer LOS for patients exposed, but effect size small after adjustment	Zero-truncated negative binomial regression	Adjusted for age, sex, primary diagnosis, emergency admission, number of procedures, number of secondary diagnoses, weekend admission, weekend discharge and socio-economic status	Patients with at least one spell of care at a non-medical ward
Bai et al. (2018)	Single-centre study, General Internal Medicine only	Shorter median LOS for overflow patients (5,31 vs 5,97 days)	Cox proportional hazards and competing risk models	Adjusted for age, sex, admission month, weekend admission, capacity ratio, comorbidities, case-mix group and resource intensity weight.	Defined by the last room location temporally closest to the end of the hospital stay (discharge or death)
Gómez-Rosado et al. (2019)	Single-centre study, General and Digestive Surgery admissions	No significant difference in LOS (4,33 vs 4,65 days)	Matched-cohort comparison (matched by diagnosis and admission date)	Matching (1:1) based strictly on surgical diagnosis (DRG) and exact admission date	Patients admitted to a ward outside General and Digestive Surgery
Bogler et al. (2021)	Multi-centre study, General Internal Medicine admissions	Shorter LOS for overflow patients (4,1 vs 6,2 days)	Generalized Estimating Equations	Adjusted for age, gender, comorbidity score, site, and discharge destination	Patients discharged from a non-GIM inpatient ward
Kohn et al. (2021)	Multi-centre study, General/Family/Geriatric Medicine admissions	Longer LOS for overflow patients across all LOS percentiles	Quantile regression (adjusted for patient-related confounders)	Adjusted for age, gender, race, ethnicity, insurance type, Elixhauser comorbidity score, ICU stay, severity risk adjustment, admission diagnosis category, daily mean medicine service patients, hospital admission source, season, and weekend admission	Patients placed outside GIM/family/geriatric specialty wards for their entire stay
Zannella et al. (2022)	Multi-centre study, General Internal Medicine admissions	Slightly shorter LOS for overflow patients	Multivariate quantile regression	Adjusted for age, sex, discharge diagnosis, Charlson comorbidity index, Physiology Score (LAPS), capacity ratio, weekend admission, night-time admission, admission year, hospital and time spent in the ED.	Patients assigned to a non-GIM ward directly after ED

Table 2.1: Overview of studies researching Length of Stay amongst patients exposed to overflow beds

What stands out from the table is that the studies do not agree on the direction of the effect. Bogler et al. (2021), Perimal-Lewis et al. (2013), and Zannella et al. (2022) all found that overflow patients had a shorter LOS than patients kept on their home ward after adjustment for case mix, and Perimal-Lewis et al. (2016) report the same direction but with a small effect size. Kohn et al. (2021), Stowell et al. (2013), and Stylianou et al. (2017) found the opposite: a significant, but generally small, effect of a longer LOS for overflow patients. Gómez-Rosado et al. (2019), Liu et al. (2014), and Serafini et al. (2015) found no significant LOS difference between groups at all.

Three explanations for this inconsistency can be drawn from the table itself: the studies differ in how they define overflow bed exposure, in how they deal with the case mix of the two groups, and in the setting and patient population they cover.

Definition of overflow bed exposure

In the literature, various definitions are used for an overflow bed. For example, Zannella et al. (2022) define overflow beds based on the patient's first ward assignment immediately following admission from the emergency department (ED). Another method exclusively considers the final ward of a patient (Bai et al., 2018; Bogler et al., 2021). Other studies use time thresholds to measure cumulative exposure. Perimal-Lewis et al. (2013, 2016), for example, classified a patient as an overflow patient when they spent at least 70% of their length of stay outside their appropriate ward, and as a non-overflow patient when they spent at least 70% inside it. Finally, some studies exclude patients who change status during their stay to maintain a 'pure' comparison across the entire admission (Kohn et al., 2021), or classify patients as overflow patients if they experienced at least one off-service ward placement during their care pathway (Stylianou et al., 2017). In their systematic review, La Regina et al. (2019) name these differences as an explanation for different findings across studies.

These definitions do not capture the same patients. A patient who spends one night on an overflow bed and is then moved back counts as exposed under the definition of Stylianou et al. (2017), but not under that of Perimal-Lewis et al. (2013, 2016), and only counts under the definition of Bogler et al. (2021) if the move happens to be the last one of the admission. Vloet (2025), at the hospital under study, addressed this by not choosing a single definition but by distinguishing three dimensions of exposure: whether a patient was exposed at all, where the overflow bed was located, and how long the exposure lasted.

Setting and population

A second difference is the setting. While newer multicentre studies, such as Zannella et al. (2022) across six hospitals and Kohn et al. (2021) across three institutions, can improve generalisability, most of the literature remains single-centre. Hospitals differ in how they manage overflow beds (as mentioned in section 2.2), but also in how their wards are organised, and this determines what counts as an overflow bed in the first place. In Stylianou et al. (2017), for example, medical patients only counted as overflow patients when placed on a non-medical ward, such as a surgical ward. In a hospital whose wards are organised by specialty, as is largely the case in the hospital under study, placement on the ward of another specialty already counts as overflow, even when the two specialties are clinically close.

The studies also differ in which patients they include. While several restrict themselves to a single specialty like General Internal Medicine, others span multiple departments. For instance, Serafini et al.

(2015) investigate both medical and geriatric wards, whereas Stowell et al. (2013) expand their cohorts to include both medical and surgical admissions.

The differences in setting, population and definition of exposure can also explain the difference in the number of patients flagged as overflow patients across studies. This ranges from 9,9% of medical admissions when only placements on non-medical wards are counted (Stylianou et al., 2017) and 15,7% when patients had to spend at least 70% of their stay outside their home ward (Perimal-Lewis et al., 2016), to around a third of General Internal Medicine patients (Bai et al., 2018; Kohn et al., 2021).

Case mix differences

The third difference involves the case mix of patients admitted to overflow beds compared to those admitted elsewhere. Some studies present a picture of a generally unhealthier patient population being placed in overflow beds. For instance, within the population studied by Santamaria et al. (2014), overflow patients were older, had a higher baseline mortality, experienced more complications, and presented with a higher average severity of illness. The population studied by Stylianou et al. (2017) showed overflow patients with a higher number of secondary diagnoses who had to undergo more procedures, and Perimal-Lewis et al. (2016) reported a higher Charlson index for overflow patients. Other studies report different population mixes. Bogler et al. (2021) observed that although the recorded comorbidity scores appeared similar on paper between both groups, overflow patients were more likely to be discharged directly home, whereas patients on their designated home ward more frequently required complex aftercare in a nursing or rehabilitation institute. Similarly, Zannella et al. (2022) found a younger population in overflow beds with less complex diagnoses.

At the hospital under study, Vloet (2025) showed that patients exposed to overflow beds have significant structural case-mix differences compared to those admitted to clinically appropriate wards. Analyzing the hospital's 2022 admission data, Vloet found that patients exposed to overflow beds were significantly older (mean age 69,7 vs. 64,9 years) and substantially more likely to be acute, unplanned admissions (91,1% vs. 66,7%) compared to non-exposed patients. Furthermore, the overflow population had more often an severe diagnosis and multiple comorbidities. Origin location prior to admission also differed significantly, with overflow patients being over three times more likely to originate from nursing homes or other care institutions instead of coming directly from home. Hospital staff give two explanations for this difference. First, the hospital has a limited number of single rooms. Older, frail patients with an infection or delirium need one, and when none is available on their own ward, especially in winter, they are placed in a single room elsewhere. Second, it is also a matter of acute admissions. An admission for a younger patient can sometimes be postponed or delayed over the weekend while older patients are often presented acutely and need immediate admission.

These findings confirm that in the context of the hospital under study, exposure to overflow happens systematically more to a sicker, older, and more frail cohort of patients. These differences matter for interpreting the effect of overflow bed exposure on LOS. When overflow patients are sicker, they may be expected to have a higher risk of a longer LOS based on case mix alone. Adjustment is therefore necessary to isolate any additional effect of the ward location itself. Similarly, in settings where overflow patients are the less sick patients, a more favourable outcome among overflow patients may be the result of patient selection instead of an effect of overflow bed exposure.

Methodological approaches to case mix differences

The studies in Table 2.1 face the same problem: patients who end up in an overflow bed are not the same patients as those who do not. There are different approaches to dealing with this case mix difference.

The first approach is no adjustment at all. Santamaria et al. (2014) and Serafini et al. (2015) compare LOS between groups directly. Serafini et al. (2015) find no significant difference in LOS between the groups. The second approach is restriction. Alameda and Suárez (2009) include only patients with a specific diagnosis: heart failure and cardiac arrhythmia with major complications, which removes diagnosis as a source of difference by design. Within this group, they find a longer LOS for overflow patients. The authors note the trade-off themselves: internal validity improves, but the findings cannot be generalised to patients from other departments. Perimal-Lewis et al. (2016) also use restriction, to patients with dementia, but find a marginally shorter LOS for overflow patients.

The third approach is matching at the point of sampling. Rather than correcting for case mix afterwards in a model, these studies make patients comparable beforehand by pairing each overflow patient with a non-overflow patient who shares the same values on a set of chosen variables. Gómez-Rosado et al. (2019) match on diagnosis and admission date, Stowell et al. (2013) on age, sex and pathology, and Liu et al. (2014) on admitting diagnosis, call shift and attending physician. Two of the three find no significant difference in LOS (Gómez-Rosado et al., 2019; Liu et al., 2014), while Stowell et al. (2013) find a longer LOS for overflow patients.

The fourth approach is regression, used by the majority of studies. These differ considerably in how much they adjust for. Perimal-Lewis et al. (2013) adjust for age, Charlson index, sex and time waiting in the emergency department and find a shorter LOS for overflow patients, while Kohn et al. (2021) additionally adjust for insurance, severity category, admission diagnosis category, ward census, season and hospital and find a longer LOS for overflow patients. Stylianou et al. (2017) show how much adjusting for case mix matters: their unadjusted comparison shows overflow patients with more than double the LOS, which shrinks to an adjusted difference of 0,17 days after correcting for age, diagnosis, procedure count and comorbidities. Overall, the regression studies do not agree on the direction of the effect. Bai et al. (2018), Bogler et al. (2021), Perimal-Lewis et al. (2013) and Zannella et al. (2022) find a shorter LOS for overflow patients, while Kohn et al. (2021) and Stylianou et al. (2017) find a longer LOS. How much a study adjusts for does not explain this split: Zannella et al. (2022) adjust for a set of variables about as extensive as Kohn et al. (2021), but find a shorter LOS.

2.5 Interaction of LOS & Mortality

The interpretation of length of stay as a measure of how long a patient's recovery took holds only for patients who leave the hospital alive. For a patient who dies during the admission, the admission still ends and a length of stay is still recorded, but what that number measures is the time until death rather than the time until recovery. Analyzing length of stay therefore requires analyzing mortality alongside it.

Why the two outcomes are related

Hofstede et al. (2018) and Lingsma et al. (2018) found that the two indicators are related at the hospital level: hospitals with relatively many patients who have a long length of stay, compared with the average for their diagnosis, also tend to have higher mortality rates. This partly reflects a clinical mechanism, as complications and case-mix characteristics that lengthen an admission also increase the risk of death.

At the level of the individual patient the relationship depends on the condition and can even reverse. Stroke patients with long length of stay (L-LOS), defined as a stay in the upper quartile of length of stay within their diagnosis group, have 58% lower odds of dying in hospital than patients with a short stay (Hofstede et al., 2018). This can be attributed to the fact that stroke patients who die tend to do so quickly after admission, while survivors often stay longer for rehabilitation. The same study shows that both patterns can hold at the same time. Individual stroke patients with L-LOS had lower mortality odds, while hospitals treating more L-LOS stroke patients still showed higher average mortality, most likely because hospitals with a generally sicker case mix have both more early deaths and longer stays among survivors. A correlation at an aggregated level therefore does not necessarily hold at the individual level and may even point in the opposite direction.

Underlying this is a mechanical problem: death functions as a competing event for LOS. A patient is either discharged alive or dies, but not both, so death gets in the way of the outcome that LOS is meant to measure. For a patient who dies in hospital, the discharge date used to calculate LOS is the date of death. LOS then no longer reflects the duration of an admission but how long it took for the competing event to occur. Because the timing of death varies by condition, this can push the observed relationship between LOS and mortality in different directions across conditions without any of this reflecting a difference in the quality of care (Lingsma et al., 2018).

How the reviewed studies handle this

Most of the studies in Table 2.1 report both outcomes but present them as two separate results of the same study, without the interpretation of one being dependent on the other. Only a small number account for the fact that death interferes with the measurement of length of stay. Two studies model discharge and death as competing endpoints (Bai et al., 2018; Kohn et al., 2021). Perimal-Lewis et al. (2013) take a lighter approach, re-running their analysis after excluding patients who died in hospital and reporting that the shorter-LOS finding was unaltered. The resulting estimates are not given. This removes the interference, but at the cost of excluding precisely the patients in whom the two outcomes interact.

The initial analysis of Perimal-Lewis et al. (2013) found that overflow patients had both a shorter LOS and a 41% higher mortality risk, with half of the deaths among overflow patients occurring within 48 hours of admission. This raises the possibility that their shorter-LOS finding partly reflects faster deaths instead of faster recoveries. Their 2016 study again found a higher risk of dying within 48 hours for patients on overflow beds, alongside a higher chance of being transferred to another hospital.

Across the remaining studies that report both outcomes, the two rarely move together in a consistent way. Bogler et al. (2021) found both a shorter LOS and a lower mortality risk among overflow patients, a pairing they attribute to the selective placement of lower-risk patients in overflow beds instead of a genuine effect of overflow placement. Zannella et al. (2022) reported a slightly shorter LOS but no significant difference in mortality. Kohn et al. (2021), Stowell et al. (2013) and Stylianou et al. (2017) each found a longer LOS with no accompanying change in mortality risk.

Mortality risk over the course of the admission

A recurring pattern is that the mortality risk of overflow patients is not constant over time. Bai et al. (2018) and Perimal-Lewis et al. (2013) both show that the risk is highest around the time of admission and declines thereafter. In Bai et al., the hazard ratio falls from 3,26 for days 0–7 to 1,16 for days 21–28, while Perimal-Lewis et al. (2013) found that 50% of deaths among patients exposed to overflow beds

occurred within 48 hours of admission. Both studies explain this through the hypothesis that the first days of an admission are clinically the most intensive and that reduced supervision on an overflow bed is therefore most risky during precisely this period.

This pattern matters for the interpretation of LOS as well. If the excess mortality associated with overflow beds is concentrated in the first days of admission, then the admissions that are cut short by death are disproportionately overflow admissions, and any comparison of LOS between the two groups that does not account for death will be affected most strongly in exactly the overflow group.

2.6 Prediction models for L-LOS and mortality

The studies on overflow exposure discussed ask whether exposure is associated with worse outcomes, a question about groups of patients on average. A placement decision needs something else: an estimate of the risk for an individual patient, based on what is known at that moment. Explaining and predicting are distinct tasks. Where an explanatory model is built to capture the underlying relationship between variables, a predictive model is built and judged on how accurately it predicts outcomes for new observations, and the two are not optimised in the same way (Shmueli, 2010). Vloet (2025) developed logistic regression (LR) models for in-hospital mortality of acute admissions, with overflow exposure as one of the predictors, to show for which patients exposure substantially increases the risk.

LR is often the standard approach for binary outcomes in clinical prediction, partly because it is interpretable: every coefficient translates into an odds ratio. The case-mix models behind the Dutch HSMR are also logistic regressions (Statistics Netherlands (CBS), 2025). But alternative machine learning (ML) methods such as decision trees, random forests and gradient boosting are more flexible, but whether this leads to better predictions is not self-evident. This section explains how these models differ from LR, under which conditions they can be expected to predict better, and what is known about predicting length of stay compared with mortality.

Considering ML models

LR assumes that each predictor adds a fixed amount to the log-odds of the outcome, regardless of the other predictors. Non-linear relationships and interactions are only included when the modeller specifies them. Tree-based models do not make such assumptions. Such models ask a sequence of yes-or-no questions about the patient, for example whether the admission is acute or if the patient is older than 80. Each question splits the admissions into two groups that differ as much as possible in the outcome. The predicted risk is the proportion of the outcome in the final group. Because a question further down applies only within its own branch, the effect of a predictor can differ between subgroups (an interaction). Because each question uses a threshold, the effect does not have to be the same across the predictor's range (non-linearity).

A single tree is easy to read, as its structure resembles the flowcharts that staff already use. It is, however, unstable. Because each split builds on the ones above it, a small change in the data can change an early split and with it the entire tree below it (James et al., 2021). This can be corrected for by combining many trees (ensembling). A random forest averages a large number of trees, each grown on a random sample of the admissions and predictors. Gradient boosting builds trees one after another, with each tree correcting the errors of the previous ones. XGBoost is a widely used implementation of this approach (Chen & Guestrin, 2016). Contributions of predictors to the outcome have to be derived

after creation of the model, for example with SHAP values, which split each individual prediction into the parts contributed by each predictor (Lundberg & Lee, 2017).

When do they outperform logistic regression?

Whether tree-based models predict better than LR depends on the data. Austin et al. (2021) showed this in simulations in which they varied the process that generated the data. When the relationships were largely additive, LR performed best. When they contained interactions or non-linear relationships, tree-based models could capture what LR missed. No method was superior, but LR and boosted trees generally performed well. The amount of data also matters. Van der Ploeg et al. (2014) found that random forests and other modern techniques needed many times more events per predictor than LR to reach stable performance. With fewer events, these models tended to overfit.

The type of predictors also matters. A binary predictor takes only two values and cannot have a non-linear effect, so for such predictors a tree-based model can only add interactions. Across a large benchmark of datasets, random forests outperformed LR in about two-thirds of the cases, but the size of the difference depended on the characteristics of the dataset (Couronné et al., 2018).

ML models in clinical settings

In the clinical literature these conditions are often not met, and the evidence for better performance of ML over LR is limited. Christodoulou et al. (2019) systematically reviewed 71 studies that compared LR with ML methods for binary outcomes. Each comparison between LR and ML was assessed for risk of bias, based on whether both approaches had been given an equal chance. This was often not the case, for example when tuning or variable selection was performed on all data instead of only the training data, when the two approaches used different predictors, or when continuous variables were handled differently. For the 145 comparisons at low risk of bias, the difference in AUC between LR and ML was 0,00, whereas for the 137 comparisons at high risk of bias, ML came out 0,34 higher. The apparent advantage of ML thus mainly appeared in methodologically weaker studies. The review also found that calibration was not addressed in 79% of the studies, even though calibration is essential when predicted probabilities are used for clinical decisions (Van Calster et al., 2019).

Predicting length of stay and mortality

Mortality is generally predicted better than length of stay. Using the health records of two academic hospitals, Rajkomar et al. (2018) predicted in-hospital mortality with an AUC of 0,93 to 0,94, against 0,85 to 0,86 for a prolonged length of stay. However, this study used all available information; models restricted to information available at admission perform worse. Verburg et al. (2014) found that models for intensive care admissions in the Netherlands explained only 9% to 20% of the variation in length of stay. They concluded that it is hard to predict from patient characteristics available at admission. In the development of the renewed LBZ indicator, CBS also found that adding patient characteristics such as socio-economic status and Charlson comorbidities did not improve the prediction of length of stay (DHD, 2026b).

Studies point to several factors outside the patient's condition at admission that determine length of stay. Stone et al. (2022) found in their review that length of stay depends on complications, discharge planning and social circumstances, besides patient characteristics. In Dutch hospitals, patients who experienced an adverse event during their admission stayed on average five days longer than patients who did not (Hoogervorst-Schilp et al., 2015). A systematic review by Rojas-García et al. (2018) showed

that admissions can be prolonged because patients wait for a nursing home, rehabilitation or social care, and that these delays are related to placement capacity and coordination between services.

2.7 Research gaps

Bringing together the literature reviewed in this chapter, some research gaps stand out.

First, it is unclear whether, and in what form, overflow bed exposure carries a risk of a long length of stay. Studies disagree on the direction of the effect of placing patients on overflow beds and on the size of the effect when one is found. Differences in definitions, populations and case-mix adjustment explain part of this disagreement. Yet, differences remain even among studies that adjust for case mix through regression. This suggests that results also depend on how a hospital organises its wards and manages overflow placement. None of the reviewed studies focussing on L-LOS was conducted in a Dutch hospital. Most focus on narrowly defined patient groups such as General Internal Medicine. The hospital under study also already applies risk-mitigating measures that differ from those in the reviewed settings. It is also unclear which aspect of exposure matters, Vloet (2025) showed that mortality risk differs by whether, where and how long a patient was exposed. It is therefore unknown whether, and through which forms of exposure, overflow placement is associated with L-LOS in a Dutch hospital that already mitigates its risks

Second, it remains unclear to what extent mortality influences the association between overflow exposure and length of stay. Most studies report both outcomes side by side, and only few account for death when analysing length of stay, for example by excluding patients who died (Perimal-Lewis et al., 2013). Because death acts as a competing event, length of stay does not measure the same thing in exposed and unexposed patients. Especially if the excess mortality of overflow patients occurs early in the admission. It is also unknown whether the two outcomes identify the same patients as being at risk. If they do, considering both adds little to considering one. If they do not, together they give a more complete picture of the risks of overflow placement. This matters for the hospital, as mortality captures the most severe form of harm, while length of stay could also reflect harm that does not end in death and directly affects bed capacity.

Third, staff currently rely on ward-specific flowcharts based on clinical judgement (Mans, 2025); data-driven steps so far consist of the logistic regression models for in-hospital mortality developed by Vloet (2025). Some gaps can be identified. The first is whether L-LOS can serve as a metric alongside mortality. Mortality captures the most severe form of harm, but not the harm that prolongs an admission without ending in death or the bed capacity an admission needs. To research whether such a model adds to a mortality model, two questions arise: whether L-LOS can be predicted well from the information available at admission, and whether it identifies other patients as being at risk than mortality does. The second is how models other than logistic regression perform in predicting these outcomes. Machine learning methods other than LR can capture non-linear relationships and interactions without specifying them in advance. Because logistic regression is easier to interpret, a more flexible model is only worth adopting if it clearly predicts better (Austin et al., 2021). Whether that is the case for mortality and L-LOS based on the information available at admission has not been examined.

3 Study Design

This chapter translates the research gaps identified into the design of this study. Figure 3.1 gives an overview of its structure. Section 3.1 formulates the main research question and the three subquestions that follow from these gaps. Section 3.2 describes the research approach in which each subquestion is answered in a separate phase. Section 3.3 describes how the data were prepared for these analyses: how the two data sources were linked, how overflow bed exposure and L-LOS were defined, and how an expected probability of L-LOS was calculated to adjust for case mix. Section 3.4 sets out how generative AI was used in this thesis. The methods of the three phases follow in Chapter 4.

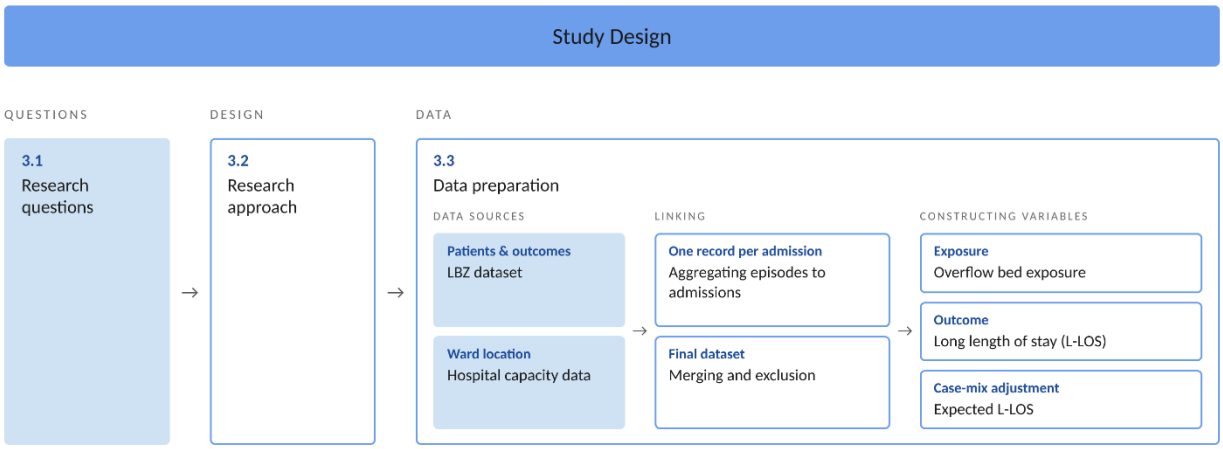

Figure 3.1: Overview of the structure of Chapter 3.

3.1 Research questions

Based on the research gaps identified in Section 2.7, the main research question of this thesis is:

> *RQ - To what extent is overflow bed exposure associated with long length of stay, and how can this risk be predicted alongside in-hospital mortality?*

To help answer the main research question in a structured manner, subquestions have been formulated.

Subquestions

The first subquestion addresses the first research gap and answers the first part of the main research question.

SQ1 - What is the risk of L-LOS associated with overflow bed exposure, after adjusting for case-mix?

The subquestion asks whether in the hospital under study patients placed on an overflow bed are more likely to have an L-LOS than comparable patients who are not. Patients placed on overflow beds are older, more often admitted acutely and clinically more complex than other patients (Vloet, 2025), so they would be expected to stay longer regardless of where they are placed. The subquestion therefore is about the risk that remains after adjusting for these differences in case mix.

Because exposure can be defined in different ways, this risk is examined for four dimensions, or four ways of how exposure can be defined. *General exposure* captures whether a patient was placed on an overflow bed at all. *Duration of exposure* captures how long the exposure lasted. *Moment of exposure* captures when during the admission it occurred. *Place of exposure* captures on which type of ward the patient was placed, and where that ward is positioned in the hospital's flowcharts. Each dimension can expose different mechanisms of how exposure influences the quality of care discussed in *section 2.3*. The answer to this subquestion therefore shows not only whether overflow bed exposure carries a risk of L-LOS, but also which forms of exposure do, and thereby where in the placement process the risk arises.

The second subquestion addresses the second research gap and, together with subquestion 1, answers the first part of the main research question.

SQ2 - To what extent is the relationship between overflow bed exposure and L-LOS distorted by mortality?

For a patient who dies during the admission, length of stay measures the time until death, and a patient who dies early cannot reach the L-LOS threshold at all. Vloet (2025) found a higher mortality risk among exposed patients at this hospital, highest in short exposure fragments. Part of any difference in L-LOS between exposed and unexposed admissions could therefore reflect how quickly patients died rather than how long they took to recover. This subquestion therefore asks whether the association examined under subquestion 1 still holds once mortality is taken into account.

The subquestion also asks whether the patients at high risk of an L-LOS are the same patients who are at high risk of dying. This links to the second part of the main research question, which concerns predicting L-LOS alongside mortality. If both outcomes identify the same patients, predicting both tells a decision-maker little more than predicting one. If they do not, the two outcomes together give a more complete picture of a patient's risk. The answer also determines how the two outcomes are modelled under subquestion 3: separately or combined.

SQ3 - Which model best predicts L-LOS and in-hospital mortality?

Placement decisions in the hospital under study are currently supported by flowcharts based on clinical judgement (Mans, 2025). A risk estimate available at the moment of placement could complement these flowcharts. Building on Vloet (2025), logistic regression serves as the reference model, as it is interpretable for the staff who would use its output. It is compared with machine learning models that can capture non-linear relationships and interactions without these having to be specified in advance.

The subquestion asks not only which model predicts best, but also whether the difference is large enough to matter. A more complex model is only worth adopting if it clearly outperforms logistic regression. If L-LOS and mortality are modelled separately, the comparison also shows how well each outcome can be predicted, and thus which offers the more reliable basis for decision support.

3.2 Research approach

This thesis is a quantitative, observational single-centre study of admissions to the hospital under study between 2022 and 2025, based on routinely collected data. This research consists of a data preparation step, three research phases and a sensitivity analysis, each building on the previous one. The data preparation step merged the national hospital discharge register (LBZ) with the hospital's capacity data into one dataset at admission level, in which overflow bed exposure, L-LOS and an expected probability of L-LOS were constructed for each admission (Section 3.3). This forms the basis for all three phases.

The first phase, answering SQ1, examined whether overflow bed exposure is associated with a higher risk of L-LOS after adjustment for case mix, and for which forms of exposure. It produced case-mix adjusted estimates of risk for each of the four dimensions of exposure. The second phase, answering SQ2, examined whether mortality distorts this association. Phase 1 analysis was repeated among survivors, and expected L-LOS and expected mortality were compared to establish whether both outcomes identify the same patients as being at risk. This result determined whether L-LOS and mortality were modelled separately or combined in the next phase. The third phase, answering SQ3, developed prediction models for L-LOS and in-hospital mortality. Logistic regression, as used by Vloet (2025), served as the reference and was compared with three machine learning models. The models were trained on earlier admissions and tested on admissions from the most recent year, to assess how they would perform on future patients.

Because the definition of L-LOS determines which admissions count as long, the data preparation and all three phases were repeated in a sensitivity analysis with a stricter threshold at the 95th percentile. Together, the phases answer the main research question: Phases 1 and 2 address the association between overflow bed exposure and L-LOS, and Phase 3 addresses how this risk can be predicted alongside in-hospital mortality.

All data preparation and analyses were performed in Python, using Jupyter notebooks in Visual Studio Code.

3.3 Data preparation

Data sources

This thesis combines two datasets. The first is a record of all admissions to the hospital, with the patient characteristics needed to correct for case mix and the length of each admission. The second is a record of where each patient was physically located during the admission, since whether a patient occupied an overflow bed is not registered nationally. The first is provided by the LBZ dataset, the second by the hospital capacity dataset.

LBZ dataset

The LBZ is a national database managed by Dutch Hospital Data (DHD) that captures medical and administrative details of patient care across Dutch hospitals. It is used to evaluate hospital

performance, to monitor quality indicators such as the standardised mortality ratio (HSMR) and unexpectedly long admissions, and for national benchmarking (DHD, 2026a).

The data provided for this study cover admissions from 2022 to 2025 and forms the reference frame for the analysis. It records patient characteristics (sex, age and socio-economic class), the primary diagnosis and any secondary diagnoses, and admission characteristics including the admitting and discharging specialty, the type of care, and where the patient came from and was discharged to. Each patient has one primary diagnosis per admission and may have several secondary diagnoses, which together describe the comorbidity burden. Diagnoses are classified in ICD-10 codes and are additionally grouped into a diagnosis group and a main cluster, both derived from the primary diagnosis. The dataset also contains the quality indicators used in this study: whether the patient died during the admission, the expected mortality, and the length of stay.

Hospital capacity data

The hospital capacity dataset is maintained by the hospital itself and is primarily used for capacity management: to measure how busy the hospital is overall, to track workload for specific specialties and wards, and to identify patterns such as seasonal trends. During the Tactical Planning Meeting described in section 2.2, capacity advisors use this historical data to evaluate ward occupancy and anticipate future demand.

For this study the dataset is needed because it records the ward on which a patient was located, which the LBZ does not, and therefore makes it possible to determine whether a patient occupied an overflow bed. The dataset describes the episodes that make up an admission. An episode lasts a maximum of eight hours and covers a day, evening or night period, which corresponds to the shifts along which staff are deployed. Each episode records the ward, the admission number, the responsible and executing specialty, the primary diagnosis, the date of service, the duration in minutes, and whether the patient had a delayed discharge, meaning the patient was medically ready to leave.

Aggregating episodes to admissions

Each row in the capacity dataset corresponds to a single registered episode rather than to the admission as a whole, so that one admission can span many episode-level records, for example when a patient moves between wards. Because the LBZ records one row per admission, the capacity data has to be aggregated to the same level before the two can be linked.

Episodes belonging to the same admission number were ordered chronologically by date, period of day and ward, and then aggregated into a single record. The ward of the first episode was taken as the admission ward and the ward of the last episode as the discharge ward. The time-based variables (total admission time, total time spent in an overflow bed and total time spent on the intended ward) were obtained by summing the corresponding values across all episodes of the admission.

Merging and exclusion

Figure 3.1 shows the flow from the two source datasets to the analysis dataset.

The capacity dataset contains nearly 1,5 million episodes, which aggregate to 112.457 distinct admissions. The LBZ extracts for 2022 through 2025 were combined into a single file containing

92.364 admissions. Linking the two on the shared admission number yields 84.327 admissions that are present in both.

A total of 28.130 admissions appear in the capacity data but not in the LBZ, mostly admissions shorter than one day, which the LBZ does not register. In the other direction, 8.034 admissions appear in the LBZ but not in the capacity data, mostly planned observational intakes. Since both a record of the admission and a record of the ward are required, only the admissions present in both datasets are retained.

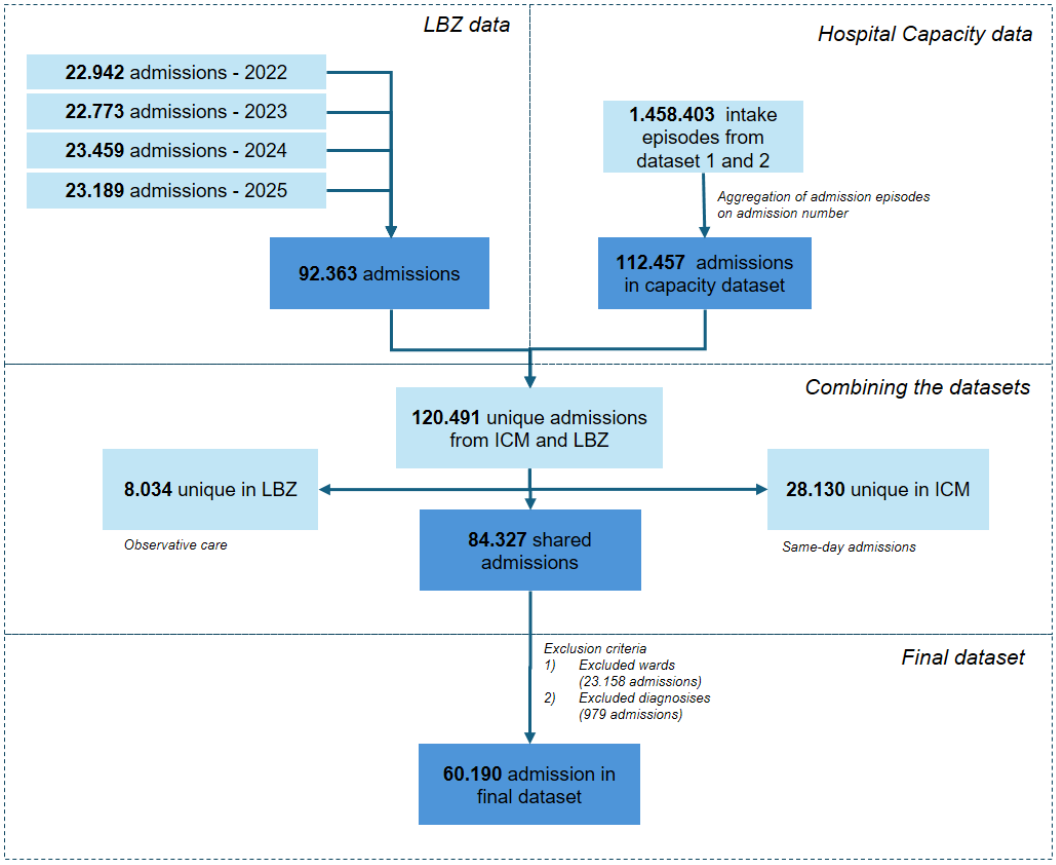

Figure 3.1. Visualisation of the merging of datasets and composition of the final dataset

Exclusion criteria

Two sets of exclusions were applied to the linked dataset. First, wards that fall outside the scope of this study were excluded: Obstetrics, the Maternity ward, Paediatrics and Neonatology, together accounting for 23.158 admissions. In principle, these wards do not exchange overflow patients: they serve specific populations of mothers, infants and children who need highly specialized care, which leaves less room to influence their capacity planning. Unlike the wards retained, they do not attend the daily Bed Availability Consultation described in section 2.2, which is where the placement decisions this study aims to inform are discussed.

Second, admissions with a primary diagnosis in the COVID-19 cluster (n = 979) were excluded to keep the study period comparable over time. COVID-19 patients form a distinct patient group whose numbers declined sharply over the study period, from 450 admissions in 2022 to 68 in 2025. The

cluster therefore has less relevance for the current situation, and a single threshold and case-mix model for it would mainly compare 2022 with later years rather than long with normal stays.

Operationalization of overflow bed exposure

The capacity dataset flags episodes in which a patient occupied an overflow bed. The flag is set by comparing the specialty responsible for the patient with the ward the patient occupies: each ward has a list of designated specialties, and an episode is flagged when the patient's specialty does not appear on that list.

In this thesis, overflow is defined at the level of the combination of specialty and ward. These specialty-ward combinations were validated with hospital staff, which led to correction of overflow flagging. Exceptions were established for the Acute Medical Unit, the Oncology Unit and the Day Care Unit, where patients are never flagged as overflow, because patients of all specialties can be admitted to these wards. On these wards, whether a patient is a guest patient depends on more detailed conditions, which were not operationalised as this would make the definition more complex. On the Acute Medical Unit, for example, a patient is only appropriately placed during the first 48 hours, after which they become a guest patient. The Oncology Unit is the home base only for patients with an oncological diagnosis, and the Day Care Unit only for patients in specific treatment or diagnostic programmes, such as (outpatient) surgery, endoscopy and the snoring clinic. Therefore, patient admissions on these wards were never flagged as overflow.

Furthermore, episodes were unflagged where a ward was confirmed to be appropriate for a patient's specialty. For example, orthopaedic patients were flagged as overflow on a ward that, after consultation, turned out to be appropriate. This resulted in 440 admissions being unflagged.

Another example is an correction that accounts for updated capacity management strategies. For example, due to highly fluctuating patient inflow in the cardiology department, the hospital began utilizing cardiology rooms for internal medicine patients in 2024. Consequently, internal medicine patients admitted to the cardiology ward are no longer classified as overflow, as this arrangement is by design. This correction resulted in 230 overflow admissions over the years 2024 and 2025 being unflagged.

Initially, 45.524 intake episodes (across 4.531 unique admissions) were flagged as overflow. After validating the medical specialties against the admission wards with hospital staff, this number was refined to 35.818 episodes (across 3.722 unique admissions).

Operationalisation of diagnosis-specific Long Length of Stay

For this thesis, length of stay (LOS) was analysed as a binary indicator of long length of stay (L-LOS), in line with the methodology of Lingsma et al. (2018). An admission was classified as L-LOS when its LOS belonged to the top quartile of stays within its diagnosis cluster, so that by design one in four admissions is long. A percentile rather than the mean was chosen as the threshold, because the mean is sensitive to extreme values, which are common in the strongly right-skewed distribution of LOS.

The binary outcome is easy to interpret and to apply in practice. Whether an admission was exceptionally long is a yes-or-no question that is easily recognisable, and a predicted probability of such a stay can be used more directly in a placement decision than a predicted number of days. DHD (2026b) made the same consideration when renewing the LBZ indicator for length of stay: after more extensive models proved difficult to translate into an understandable indicator, it opted for a simpler

approach that flags exceptionally long admissions, with explainability for care professionals as an explicit aim.

A binary outcome was also chosen because LOS is difficult to predict as a continuous variable. Studies that model LOS directly find that patient characteristics explain only a limited share of its variation. Austin et al. (2002) relate this to the shape of the distribution: because a small share of patients have very long stays, LOS is hard to capture in a model. A binary indicator asks less of the data: it does not have to predict how long a stay will last, only whether it will exceed the threshold.

The binary form also aligns the outcome with how prolonged admission is monitored in Dutch hospitals. Until 2025, the Landelijke Basisregistratie Ziekenhuiszorg (LBZ) used the indicator Onverwacht Lange Opnameduur (OLO): the share of admissions with a LOS at least 50 per cent longer than expected, based on the mean LOS of admissions with the same age group, principal diagnosis and, where applicable, principal procedure (DHD, 2026b). During this thesis, DHD replaced the OLO with a renewed Opnameduur-indicator. Because LOS is strongly skewed, this indicator flags an admission as exceptionally long when its LOS exceeds the 95th percentile of comparable admissions, grouped by diagnosis and operative profile.

The renewed indicator was published when the definition of L-LOS and the first analyses of this thesis had already been completed, so the definition of this thesis was retained. It shares the central principle of the renewed indicator, a percentile within diagnostically comparable groups rather than a deviation from the mean, but uses a lower percentile (95th percentile as threshold) and a different grouping. DHD notes that hospitals may choose a different percentile depending on the research question (DHD, 2026b). In addition, the sensitivity analysis repeats all phases with the 95th percentile of the renewed indicator (Section 4.5).

The binary form also has limitations. Two admissions just on either side of the threshold are treated as categorically different, although their duration is almost identical. Dichotomising also discards information (Altman & Royston, 2006). A stay just above the threshold counts the same as one that lasts weeks longer, so the measure does not show how many extra bed-days are associated with overflow exposure. Finally, because the threshold is a percentile of the observed distribution, the prevalence of L-LOS is fixed at approximately 25 per cent by construction. The measure is therefore relative and cannot show whether prolonged admissions become more or less common over time.

Diagnosis-specific thresholds

LOS is driven to a large extent by the reason for admission, as patients admitted for different conditions follow different care pathways and therefore have different normal durations of stay. A single threshold calculated across the whole population would mainly identify the diagnoses that are associated with longer admissions, rather than the admissions that are unusually long for the condition in question. The threshold was therefore calculated within diagnostically comparable groups, so that an admission is flagged as L-LOS only relative to admissions following a similar care pathway.

In this thesis, admissions were grouped at the level of the diagnosis main cluster (n = 14 after excluding 3 clusters), the highest level of the diagnosis classification used in the LBZ. Grouping at the more detailed level of the diagnosis group (n = 158) was also tested. Diagnosis groups are clinically more homogeneous, which is an argument in their favour, but with data from a single hospital, many contain too few admissions for a reliable threshold or case-mix model. The case-mix models fitted at this more detailed level had lower explanatory power and wider confidence intervals than the models per main

cluster. The main cluster was therefore chosen as the most detailed level at which both the thresholds and the case-mix models could be estimated reliably.

Calculating the threshold

To flag only stays that are long for their diagnosis, a separate threshold was set for each of the 14 remaining main clusters: the 75th percentile of LOS across all admissions in that cluster within the study population. Table 3.1 shows, per cluster, the number of admissions, the mean and median LOS, and the resulting threshold. The thresholds range from 3,72 days for diseases of the blood to 11,67 days for infectious and parasitic diseases. An admission was flagged as L-LOS when its LOS exceeded the threshold of its own cluster.

Main cluster	Number of admissions	Average LOS (days)	Median LOS (days)	Cut-off value for L-LOS (days)
Diseases of the circulatory system	11.178	5,26	3,48	6,79
Diseases of the digestive system	8.958	4,28	2,11	4,80
Neoplasms	7.932	4,18	2,17	4,32
Injury and poisoning	7.269	4,73	2,86	5,88
Diseases of the respiratory system	6.640	4,91	3,30	6,40
Diseases of the genitourinary system	6.351	3,53	1,84	4,00
Symptoms, signs and abnormal findings, and factors influencing health status	2.450	3,19	1,88	3,85
Diseases of the nervous system and sense organs	2.350	4,09	1,90	4,31
Endocrine, nutritional and metabolic diseases, and immunity disorders	1.739	5,12	3,05	6,64
Diseases of the musculoskeletal system and connective tissue	1.697	5,65	2,99	6,91
Diseases of the skin and subcutaneous tissue	1.500	5,26	3,71	6,44
Diseases of the blood and blood-forming organs	792	3,19	1,43	3,72
Mental disorders	690	5,91	2,76	7,53
Infectious and parasitic diseases	644	9,27	6,24	11,67

Table 3.1. The 14 main clusters with the number of admissions within that cluster, the average length of stay of admissions within that cluster, the median length of stay and the L-LOS cut-off value (or threshold)

The stricter threshold used in the sensitivity analysis (L-LOS95, Section 4.5) was determined in the same way at the 95th percentile of each main cluster; these thresholds are reported in F.

Calculating the expected Long-Length Of Stay

To examine how different dimensions of overflow exposure are associated with the risk of L-LOS, groups must be comparable in terms of the patients they contain. Groups differ systematically in case mix, and without correction any difference in observed risk would partly reflect those differences rather than exposure itself. An expected probability of L-LOS was therefore calculated for every admission, which can be used to adjust risks and odds for case-mix.

The approach is modelled on the methodology underlying the CBS Hospital Standardised Mortality Ratio (HSMR), in which an expected mortality is calculated per admission. Following that methodology, separate logistic regression models were fitted for each of the fourteen main clusters, allowing covariates to carry different coefficients across clusters. All models were estimated by maximum likelihood with the L-LOS flag as the dependent variable, and, in line with the HSMR methodology, non-significant variables were still included.

The covariates included are age (grouped into bins: 0–19, 20–29, ..., 90–99, 100+), sex, socioeconomic status, severity of the main diagnosis, urgency of admission, comorbidities, source of admission, year of registration, and month of admission. Variables are represented using dummy variables where for each categorical covariate, one category is selected as the reference category. A covariate with (n) categories is then represented by (n-1) binary indicator variables. The categories (and their reference groups) are presented in appendix G.

The case-mix models only include characteristics of the patient at the start of the admission. Variables reflecting the care delivered during the admission, such as the ward of an admission or the procedures performed, were not adjusted for, so that their influence remains visible in the observed outcome. The ward where a patient stayed, for example, reflects a decision made during the admission and adjusting for it would mean correcting for what the hospital had already done. The other side is that differences in treatment between exposed and unexposed admissions, such as an intervention, also remain part of the comparison (Section 6.2).

Two data-driven exclusion rules were applied before fitting each model. Each category of a covariate enters the model as a separate predictor, for example a specific age group, comorbidity or source of admission. First, a predictor was excluded from the model of a cluster when fewer than ten admissions in that cluster had both the characteristic and an L-LOS. Although CBS (2025) applies a threshold of one event per predictor, this produced models that failed to converge. Second, a predictor was excluded when the characteristic occurred fewer than fifty times in the cluster in total. Admissions with an excluded characteristic were thereby assigned to the reference category of that covariate. For example, the predictors for Charlson comorbidities 8 and 17 and for the highest severity class of the main diagnosis (0,4–1) were retained in only 2 of the 14 models, and those for the residual category of socio-economic status and the severity class 0,3–0,4 in only 3 of the 14 models.

Model performance was assessed per cluster with Nagelkerke's pseudo-R^2, which expresses on a scale from 0 to 1 how much better the model explains L-LOS than a model without predictors. Together, these models give a predicted probability of L-LOS for each admission, based on the patient's characteristics at the start of the admission.

Nagelkerke's pseudo-R^2 ranged from 0,11 to 0,46, indicating considerable variation in how well the covariates explained L-LOS across clusters. The best fit was found for diseases of the genitourinary system (R^2 = 0,46), mental disorders (0,38) and neoplasms (0,37), and the weakest for the residual cluster of symptoms, signs and ill-defined conditions (0,11). In appendix D the number of admissions, the frequency of L-LOS, the number of predictors and Nagelkerke's pseudo-R^2 are reported for each model. High values are not to be expected for clinical outcomes, which depend on many factors that administrative data do not capture. Gupta et al. (2024) conclude from a review of clinical studies that an R^2 above 0,15 can generally be considered clinically meaningful. Twelve of the 14 models exceed this value.

The expected probability of L-LOS95 was estimated with the same model specification; these models are also reported in Appendix F.

3.4 A.I. Disclosure

Generative artificial intelligence, namely Anthropic's Claude , was used in this thesis as a supporting tool. In writing, it was used to refine text, support the structure of chapters and paragraphs, improve grammar and check the consistency of terminology and reported figures. In the analysis, it supported the coding in Python, by helping to restructure code, debug errors and create figures. No patient-level or otherwise confidential hospital data were entered into the tool.

Examples of prompts:

> *"How can I create a bar graph in which each bar is divided according to the frequency of another variable?"*
>
> *"How can I improve the flow and structure of this section?"*
>
> *"Are there any grammatical or formulation errors in this paragraph?"*

The tool supported the work but did not replace it. The choices on, for example, research questions, methods, interpretation and conclusions were made by the researcher.

4

Research Methodology

This chapter describes the methods used to answer the three subquestions. Section 4.1 describes the methods used to characterize the study population and length of stay distribution. Sections 4.2, 4.3 and 4.4 then set out the method of each phase in turn. Finally, Section 4.5 describes the sensitivity analysis. The results are presented in Chapter 5 in the same order.

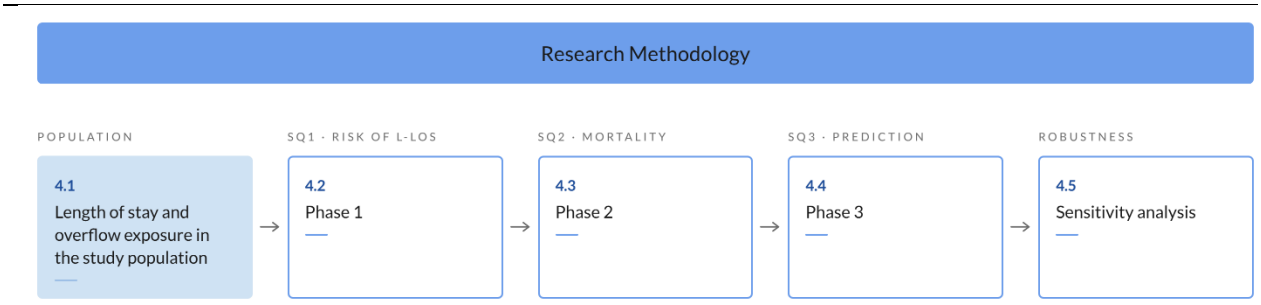

Figure 4.0: Overview of the structure of Chapter 4.

4.1 Length of stay and overflow exposure in the study population

Before the association between overflow exposure and L-LOS is examined, this section describes the outcome and the exposure in the study population of 60.190 admissions (Section 3.3). It gives insight in why length of stay is analysed as a binary outcome based on a percentile, and it shows how often exposure occurs and how length of stay differs between exposed and unexposed admissions before any adjustment. It gives also insights in how L-LOS should be interpreted using the threshold. Patient characteristics are not described here; differences between exposed and unexposed admissions are accounted for through the expected L-LOS (Section 3.3). The description consists of four steps, all without adjustment for case mix.

First, the number of admissions and the length of stay are reported per registration year, using the mean with standard deviation and the median. Comparing the mean with the median over time shows whether changes in length of stay concern the typical admission or the tail of the distribution. Second, the distribution of length of stay across all admissions is shown as a histogram, with the 25th, 50th, 75th and 99th percentiles marked and the share of admissions flagged as L-LOS shown per interval. This shows how skewed the distribution is and where the L-LOS thresholds fall within it. Third, the

number and percentage of admissions with at least one overflow episode are reported per registration year. Fourth, the distribution of length of stay is shown separately for exposed and unexposed admissions, with the median, interquartile range and 99th percentile for each group. This gives the unadjusted difference in length of stay that Phase 1 then corrects for case mix.

4.2 Phase 1

This phase examines the extent to which overflow exposure is associated with a higher risk of L-LOS. As concluded in Section 2.4, an uncorrected comparison gives a distorted picture. All comparisons are therefore corrected for case mix with the expected L-LOS (Section 3.3). The resulting increased risk of L-LOS can be measured in several ways, which are presented below.

Measuring risks

Three measures are used, because each answers a different question. The observed risk shows how often an L-LOS occurs in a group and serves as the starting point. The case-mix adjusted risk shows how often an L-LOS occurs relative to what the case mix of that group predicts. Comparing the two shows how much of the raw difference between exposed and unexposed admissions is explained by case mix. The odds ratio, estimated with logistic regression, compares the odds of an L-LOS between exposed admissions and unexposed admissions with a comparable risk profile.

All three measures are reported for general exposure, the main comparison. There, the difference between exposed and unexposed admissions is expressed in two ways. In absolute terms, the difference between their adjusted risks in percentage points. In relative terms, the odds ratio shows how many times higher the odds of an L-LOS are for patients with overflow exposure. Once this has been shown for the main comparison, the other three dimensions focus on how their categories differ from unexposed admissions. For these, only the odds ratio is reported, because a single model per dimension compares all categories with the same reference group and gives a confidence interval for each.

Observed risk.

The observed risk is the number of admissions with an L-LOS divided by the total number of admissions in a group, without adjustment.

Case-mix adjusted risk

The adjusted risk compares the observed number of L-LOS admissions in a group with the number expected given its case mix. The expected number is the sum of the expected probabilities of L-LOS of the admissions in that group (Section 3.3). For a group of admissions X:

$$Adjusted\ LLOS = Observed\ overall\ LLOS\ rate\ (25\%) \times \frac{\sum Observed\ LLOS(x)}{\sum Expected\ LLOS(x)}$$

An observed number above the expected number means more L-LOS admissions than the case mix predicts, and a lower number means fewer. Multiplying the ratio by the overall L-LOS rate of 25% returns it to the percentage scale, so that it can be read alongside the observed risk. This is the indirect standardisation also used in the CBS HSMR (CBS, 2025).

Odds ratio

Odds ratios were estimated with logistic regression, with one model per exposure dimension. The outcome of each model is whether an admission had an L-LOS. The model contains two covariates: the exposure variable of that dimension and the expected L-LOS of the admission. The expected L-

LOS is entered on the logit scale. Including the expected L-LOS adjusts for case mix, so that the odds ratio compares exposed and unexposed admissions with a comparable underlying risk. The odds of an L-LOS are the number of admissions with an L-LOS divided by the number without. Odds of 0,5, for example, mean that there is one admission with an L-LOS for every two without. The odds ratio divides the odds of an exposure group by the odds of unexposed admissions. Because the model includes the expected L-LOS, this comparison is made between admissions with a comparable expected L-LOS. An odds ratio of 1 indicates no difference; values above 1 indicate higher odds of an L-LOS, and values below 1 lower odds. Odds ratios are reported with their 95% confidence interval, and estimates that do not differ significantly from 1 at the 5% level are marked in the tables.

Risk dimensions

Overflow exposure can be operationalised in more than one way. Building on the dimensions introduced by Vloet (2025), *general exposure*, *duration of exposure* and *place of exposure*, this study adds a fourth: the moment at which exposure occurs. It also alters the place dimension. Where Vloet describes placement at an aggregated level, this study evaluates it against the flowcharts drawn up by hospital staff, which specify for each ward which alternatives are preferred when the appropriate ward is full.

The specific recoding steps, including the derivation of segments and episodes, the category boundaries used for each dimension, and the mapping of wards onto the flowcharts, are documented in Appendix H. Admissions without exposure serve as the reference category.

General exposure

The first dimension distinguishes admissions with any overflow exposure from those without. It is a binary indicator, set to 1 if at least one exposure episode occurred at any point during the admission. This dimension does not differentiate between a brief and a prolonged stay in an overflow bed, nor between one episode and several; it only flags whether the patient was exposed.

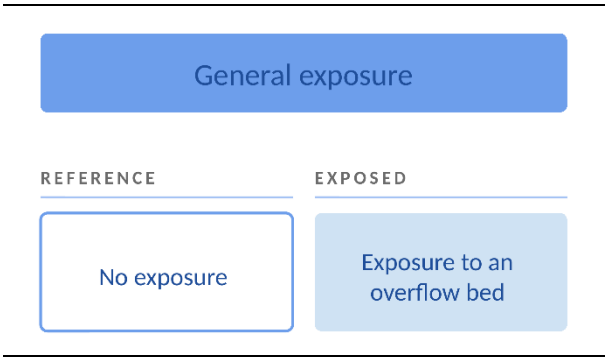

Figure 4.1: Exposure category in General exposure

Duration of exposure

The second dimension expresses how long a patient occupied an overflow bed, which can be measured in both absolute and relative terms. The absolute measure is calculated as the total time across all exposure episodes within a single admission. Time is measured in registered minutes and aggregated per admission; therefore, an admission with two six-hour episodes is treated identically to one with a single twelve-hour episode. This absolute time was categorized as follows: the first 48 hours

were divided into six 8-hour intervals, the period between 2 and 7 days was divided into 1-day intervals, and stays exceeding 7 days formed the final category. Intervals are narrower for shorter exposures because most exposure episodes occur within this range, so finer categories capture the bulk of the variation. This also allows the early peak in risk found by Vloet (2025), within the first 18 hours of exposure, to be captured.

The relative measure divides the exposure time by the total admission time, giving the proportion of the hospital stay spent in an overflow bed, as in Perimal-Lewis et al. (2013, 2016) and Santamaria et al. (2014). For example, a value of 0,25 indicates that a quarter of the admission was spent outside the appropriate ward, while a value of 1 means the entire admission was spent in overflow. Whereas Perimal-Lewis et al. used a single threshold of 70%, these relative values were grouped into deciles (10% increments).

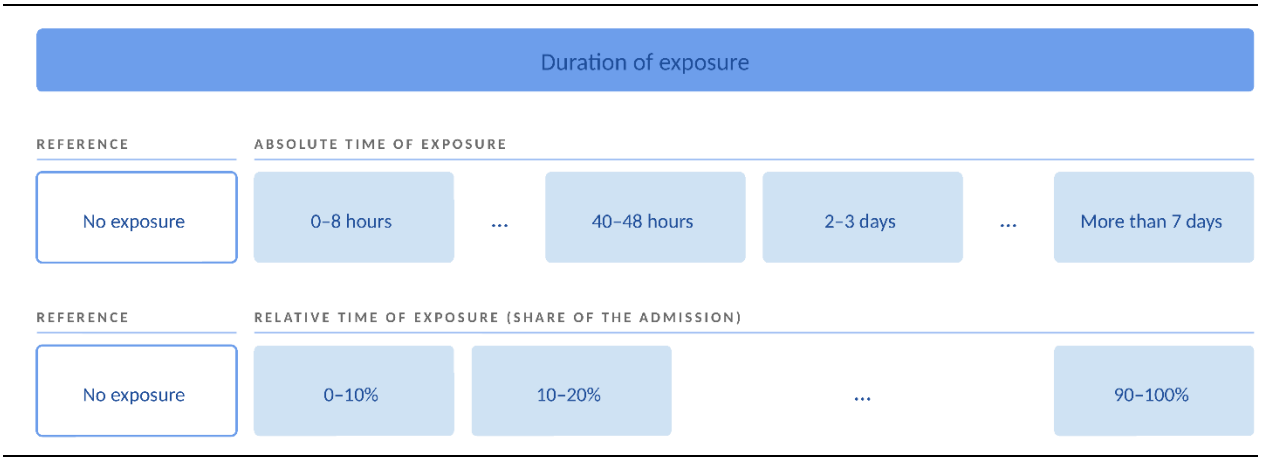

Figure 4.2: Exposure categories in Duration of exposure (Absolute & Relative)

Moment of exposure

The third dimension captures where in the admission the exposure occurred. The dimension is made categorical, based on the position of the first exposure episode within the admission. Three mutually exclusive groups are distinguished. *1st moment and onward* comprises admissions in which the patient occupied an overflow bed in the first segment; *2nd moment and onward* comprises admissions that began on an appropriate ward and involved overflow later; *last moment* comprises admissions in which overflow occurred only in the final moment, before discharge. The final group is separated first, and the remaining admissions with a later first episode are assigned to second and onward. Only the first episode determines group membership and later exposure episodes do not alter it.

The Acute Medical Unit, the EHH/CCU, the Day Care Unit and the Emergency Department are treated as flow-through wards (segment zero) and do not count as a patient's first ward. Patients on these wards are correctly placed, but not yet on the ward that matches their diagnosis. An overflow placement directly after a flow-through ward is therefore counted as exposure from the *first moment onward*.

Figure 4.3 segments the admission into intervals numbered 0 through N, where segment 0 represents time spent on one of the flow-through wards (CCU/EHH, Daycare, SEH, AMU). Segment 1 is classified as *1st moment and onward*, segments 2 through N-1 as *2nd moment and onward*, and segment N as *last moment*.

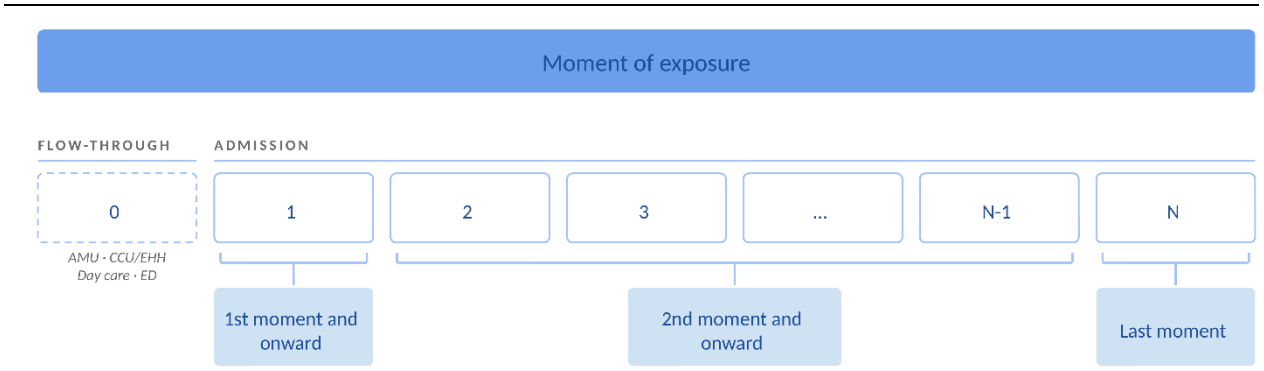

Figure 4.3: Exposure categories in Moment of exposure

Figure 4.4 illustrates how this classification rule is applied. It demonstrates that only the position of the first exposure episode determines the assigned category: an admission exposed in both the first and third segments is still classified as *1st moment and onward*, and an admission exposed throughout its entire length receives the same classification as one exposed only briefly at the start.

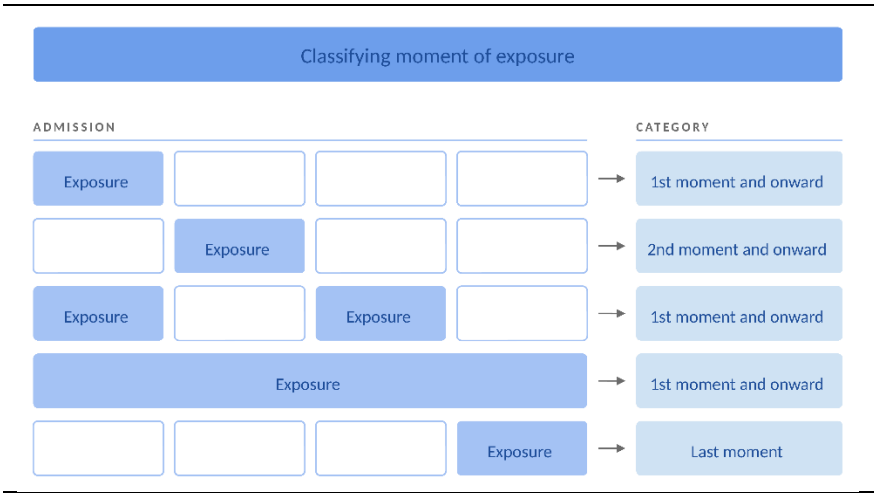

Figure 4.4: Illustration of classifying admissions in categories of moment of exposure

Place of exposure

The fourth dimension concerns where the patient was placed. As described in Section 2.2, the hospital maintains flowcharts that specify, for each ward (and for some wards per specialty), which alternative wards should be used when the appropriate ward is full. The order of these alternatives is based on how similar the care on the alternative ward is to that on the appropriate ward (Mans, 2025). The alternatives are grouped into three groups. The *1st preferred group of alternative wards* contains the most preferred alternatives, where the care delivered is most similar to that of the appropriate ward. The *2nd preferred group of alternative wards* contains the next alternatives, which are used when the wards in the first group are full. *The last preferred group of alternative wards* is the last resort, used when none of the wards in the other groups is available; in every flowchart, this is the ICU.

The flowcharts also contain exceptions. They list diagnoses for which overflow placement is not considered safe, for example because specialised equipment or expertise is required (Mans, 2025). In addition, specific patient groups, such as patients who need an isolation room, may be placed outside the regular order. These exceptions could not be identified in the data. The classification therefore reflects the position of the ward in the flowchart, not whether a placement followed an exception, and a placement in a less preferred group is not necessarily a deviation from the agreed procedure.

For each exposure episode, the intended ward and the responsible specialty determine which flowchart applies, and the ward on which the patient was actually placed determines its group. Each episode is classified accordingly as belonging to the 1st, 2nd or last preferred group of alternative wards.

Two further categories capture episodes that are not covered by a flowchart. The category *ward not in flowchart* contains episodes in which the patient was placed on a ward that does not appear in any of the three groups of the applicable flowchart. An example is a patient who needs care that only the team of a specific ward can provide, such as the placement of a midline on the cardiac emergency and coronary care unit (EHH/CCU), while that ward is not listed in the flowchart of the patient's intended ward. The placement thus falls outside the alternatives that the flowchart offers. The category *no applicable flowchart* contains episodes for which no flowchart exists for the combination of intended ward and specialty, so that the placement cannot be evaluated against a flowchart at all. An example of this is ICU patients, for whom there is no alternative because of the specialized care needed.



Figure 4.5: Exposure categories in Place of exposure

4.3 Phase 2

As discussed in the literature review, death can act as a competing event for L-LOS: a patient who dies during admission may not survive long enough to reach the L-LOS threshold, so that death can distort the observed association between exposure and the risk of L-LOS. This phase therefore examines the extent to which mortality distorts that association, in three steps. First, expected mortality and expected L-LOS are compared using correlation, to establish to what extent the same admissions are at high risk of both outcomes. Second, the raw distribution of mortality is presented among admissions with and without an L-LOS, and among admissions with and without overflow exposure. Finally, the general-exposure analysis from Section 4.2 is repeated for survivors only, and its odds ratio is compared with that of the full population, to assess whether excluding non-survivors changes the observed association between exposure and L-LOS.

Correlation between expected mortality and L-LOS risks

Both expected L-LOS and expected mortality are calculated using patient characteristics recorded at admission. A strong correlation would indicate that the two measures assign a high risk to the same patients.

The correlation is reported in three ways, because the two expected probabilities are distributed very unevenly. Most admissions have a low expected mortality, while a small group has a very high one. Pearson's correlation measures a linear association, and extreme values can have an undue influence on it (Schober et al., 2018). On the raw probability scale, it therefore gives most weight to admissions that lie far from the average, so it mainly reflects that small group of high-risk admissions rather than the population as a whole. Therefore, Pearson's correlation is also calculated on the logit scale. Transforming variables is a common way to bring them closer to a normal distribution and to linearise their relationship (Schober et al., 2018). The logit transformation stretches the compressed probabilities at the extremes into a continuous and unconstrained scale, allowing the linear measure to capture variation across the entire risk spectrum without being dominated by the highest values. Finally, Spearman's correlation compares the order in which the two models rank admissions, giving every admission equal weight regardless of how extreme its probability is. It is a Pearson correlation calculated on the ranks of both variables, does not require normally distributed data and is relatively robust to extreme values (Schober et al., 2018). Given the skewed distribution of expected mortality, Spearman's correlation is therefore taken as the main measure.

The strength of the correlations is interpreted following the conventional classification in Schober et al. (2018), in which a coefficient of 0,40–0,69 indicates a moderate, 0,70–0,89 a strong and 0,90 or higher a very strong correlation. These cut-off points are a rule of thumb rather than a strict boundary.

Distribution of mortality in exposure and L-LOS

To examine how mortality is distributed across the outcome and the exposure of interest, the raw proportion of admissions ending in in-hospital death is reported for admissions with and without an L-LOS, and this comparison is then repeated separately within the exposed and unexposed groups. Splitting this comparison by exposure status shows whether the relation between L-LOS and mortality is the same in exposed and unexposed admissions

Analysis of survivors

To assess whether the association between overflow exposure and L-LOS reported in phase 1 is affected by in-hospital mortality, the general-exposure logistic regression analysis from that section is repeated on a restricted population that excludes admissions ending in death. If death systematically shortens the admissions of the most severely ill patients before they can reach the L-LOS threshold, and this occurs to a different extent among exposed than among unexposed admissions, the odds ratio for L-LOS would be expected to differ between the full population and this survivor-only set.

4.4 Phase 3

Phase 3 addresses the third subquestion: which model best predicts L-LOS and in-hospital mortality? A prediction model estimates, for an individual patient at admission, the probability of an outcome, based on characteristics such as age and comorbidities. As Vloet (2025) developed logistic regression models for mortality, this phase extends prediction to L-LOS and compares logistic regression with more flexible machine learning models. The method is described in five parts: which outcomes are predicted, which models are compared, how the models were trained and tested, which patient characteristics were used, and how the performance of the models was evaluated.

Which outcomes are predicted

Two outcomes are predicted: whether an admission has a long stay (L-LOS), and in-hospital mortality as registered in the LBZ. Whether they are predicted in separate models or in one combined model depends on the answer to subquestion 2: whether the same admissions are at high risk of both outcomes (Section 4.3). If they are, a combined model would identify largely the same patients as two separate models, so each outcome is predicted in its own model. If they are not, both outcomes are combined into one model with three possible outcomes (normal stay, long stay or death), so that it shows which patients are at risk of which outcome. The models, the split into training and test data and the predictors described below are the same in both cases.

Which models are compared

Four models were compared: logistic regression (LR), decision tree, random forest and XGBoost. LR models the log-odds of the outcome as a linear combination of the predictors and served as the reference model, as it is used in Vloet (2025) and offers the most interpretable results for clinical professionals. The other three models are based on decision trees, which divide patients step by step into smaller groups using yes-or-no questions, such as whether a patient is older than 80. A random forest combines many such trees and averages their predictions, while XGBoost builds trees one after another, each correcting the errors of the previous ones (Section 2.6).

Unlike LR, the tree-based models can capture non-linear relationships and interactions between predictors without these being specified in advance, for example when the effect of comorbidities differs between age groups. They are, however, harder to interpret. The comparison therefore addresses whether these more flexible methods add predictive value over LR, which would justify their lower interpretability.

Training and testing the models

To test whether a model also works for future patients, the models were developed on earlier admissions and tested on later ones. The analyses were performed on the final dataset described in Section 3.3, which largely follows the procedure of Vloet (2025) but covers two more years (2022–2025) and excludes the wards listed in Section 3.3 and patients admitted with COVID-19. In line with Vloet, the models were restricted to acute admissions.

Admissions from 2022 to 2024 (about 75%) were used to develop the models (the training set), and admissions from 2025 (about 25%) to test them (the test set). A temporal split was chosen instead of a random one because it is in line with how a prediction model is used in practice: a model developed on historical admissions is applied to patients admitted later. The same approach was used for temporal validation in Vloet and is recommended for the evaluation of clinical prediction models (Riley et al., 2024).

Each model also has settings that determine how complex it is allowed to become, such as the number of trees in a random forest. These settings were chosen using only the training data: models were fitted on admissions from 2022 and 2023 and compared on admissions from 2024, after which the best settings were used to refit each model on all training data (Appendix F). The test set therefore played no role in developing the models and was used only once, for the final evaluation.

Which patient characteristics are used

Seven patient characteristics were used as predictors: age, sex, socio-economic status, comorbidities, severity of the main diagnosis, source of admission, and month of admission. These are all known at the moment of admission. Because the analysis was only done for acute admissions, urgency of admission was not included as a predictor. In addition, overflow exposure was included in its simplest form: whether a patient was exposed at all (general exposure). How each characteristic was coded for the models, including an additional version of XGBoost in which some characteristics were supplied in a different format, is described in Appendix H.

How the models are evaluated

A measure such as accuracy would be misleading as a model that predicts that nobody dies would be right for 96–97% of admissions, and a model that predicts that nobody has a long stay for 69%. The models were therefore evaluated on three aspects, following Steyerberg et al. (2010). Overall performance (scaled Brier score), discrimination (AUC) and calibration (calibration intercept and slope). Together, these show how close the predictions are to the observed outcomes, how well a model separates patients with and without the outcome, and whether the predicted risks are correct (Appendix D).

To examine whether the tree-based models made use of interactions, the contribution of each predictor to individual predictions was inspected with SHAP values.

Finally, the AUC does not show how often a high-risk prediction is correct, which matters for a rare outcome such as mortality. For the logistic regression models, it was therefore also calculated how many of the patients with the outcome are found, and how often a high-risk label is correct, when a given share of admissions with the highest predicted risk is flagged as high risk (Appendix I).

4.5 Sensitivity Analysis

Two sensitivity analyses were performed. The first assesses whether the conclusions depend on the definition of L-LOS. All three phases were repeated with a threshold at the 95th percentile instead of the 75th, in line with the renewed LBZ indicator (DHD, 2026b), using the corresponding expected probability. Everything else was kept identical to the main analysis, so that any difference in results can be attributed to the threshold. Because L-LOS95 is a rarer outcome, its odds ratios are not directly comparable in size with those of the main analysis. The comparison therefore focuses on whether the direction and pattern of the results hold under both thresholds. Appendix F repeats the data preparation steps of Section 3.3 for the 95th percentile, including the diagnosis-specific thresholds per main cluster and the logistic regression models per cluster used to estimate the expected probability of L-LOS95.

The second assesses whether the results of Phase 3 depend on the particular assignment of years to training, validation and testing. The models were refitted and evaluated in all twelve possible combinations of training, validation and test years, and the range of performance across these combinations was compared with the main split.

5 Results

This chapter presents the results of the analyses described in Chapter 4, in the same order. Section 5.1 describes the distribution of length of stay in the study population. Sections 5.2, 5.3 and 5.4 present the results of the three phases. Each phase concludes with a short summary of its main findings. The results of the sensitivity analysis with the 95th-percentile definition of L-LOS are presented in Section 5.5. The interpretation of the results follows in Chapter 6.

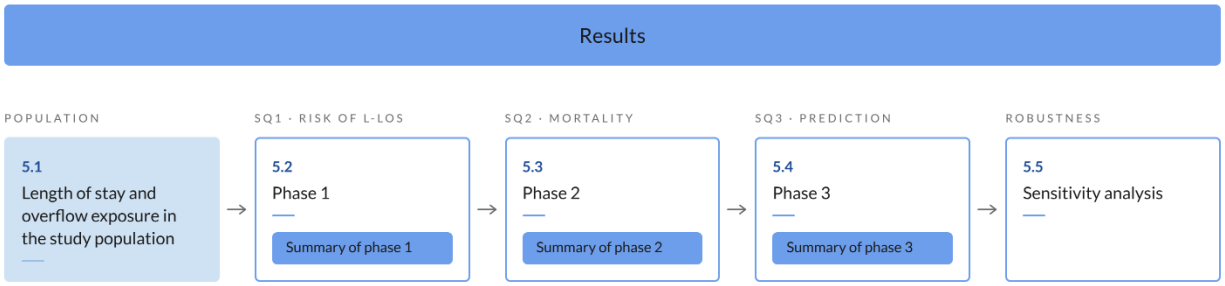

Figure 5.0: Overview of the structure of Chapter 5.

5.1 Length of stay and overflow exposure in the study population

This section presents the distribution of length of stay over the study period and across admissions. It then shows the length of stay of admissions with and without overflow exposure. All results in this section are descriptive and not adjusted for case mix, so differences between groups should not be read as effects of exposure.

Length of Stay

Table 5.1 reports the number of admissions and the length of stay (LOS) per year for the study population. Between 2022 and 2025 the annual number of admissions rose by approximately 1.000, an increase of about 7 per cent. Over the same period the mean LOS fell by 0,36 days (8,6 hours). While the median fluctuated slightly, it remained stable at roughly 2,6 days. The combination of a decreasing mean and a stable median points to a reduction in the tail of the distribution of length of stay rather than to a shorter stay for the typical patient.

		2022	2023	2024	2025	Total
Number of admissions		14.541	14.791	15.289	15.569	60.190
Length of Stay (days)	Mean (SD)	4,77 (6,66)	4,64 (6,32)	4,54 (5,89)	4,41 (5,39)	4,59 (6,07)
	Median	2,6	2,59	2,68	2,64	2,63

Table 5.1: Distribution of length of stay over the different recorded years

The gap between the mean and the median can be explained by the distribution of length of stay across patients. Figure 5.1 presents this distribution as a histogram of the number of admissions per length-of-stay interval, with the 25th, 50th, 75th and 99th percentiles marked. A quarter of all patients are discharged within 1,2 days, half within 2,6 days, and a quarter only after 5,6 days or more. The distribution is thus strongly concentrated at the left. The upper quarter extends across a range far wider: the 99th percentile lies at 28,7 days. This follows from the nature of LOS as a count variable: every admission starts at day zero, and from that point onward admissions are discharged at every subsequent day, so the number of admissions still ongoing can only shrink as the day count increases.

A second feature of the distribution is its succession of peaks at approximately whole-day intervals. These show the daily rhythm of the hospital: admissions and discharges are concentrated during the day and are rare at night. An admission is therefore far more likely to end at one of a limited number of moments in the day.

Figure 5.1 also shows which admissions are flagged as L-LOS, and thereby how the outcome in the following sections should be read. No admission shorter than 3,72 days is flagged, and every admission longer than 11,67 days is. These are the lowest and the highest threshold in Table 3.1. Between them, length of stay alone does not determine whether an admission counts as long; its diagnosis cluster does. A stay of six days, for example, is long for a patient with a neoplasm (threshold 4,32 days) but normal for a patient with a disease of the circulatory system (6,79 days).

Figure 5.1 also shows the limits of this reading. The flag covers stays from just under four days to more than a month. The 525 admissions (0,9%) that lasted longer than 30 days count the same as an admission just above its threshold. An odds ratio for L-LOS therefore shows how often admissions become long, not how much longer they last.

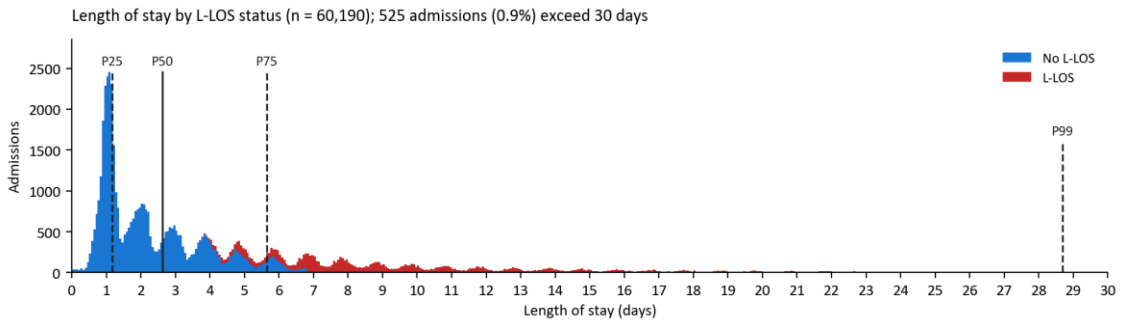

Figure 5.1: distribution of length of stay in the population. Admissions flagged as L-LOS are red while those admissions with no L-LOS are blue.

Overflow & length of stay

Table 5.2 shows the frequency of overflow exposure in the study population per registration year. Of the 60.190 admissions, 3.214 (5,3%) involved at least one overflow episode. The number of exposed admissions declined from 1.113 in 2022 (7,7% of admissions) to 637 in 2024 (4,2%), and remained at a similar level in 2025 (642; 4,1%).

	2022	2023	2024	2025	Total
Number of admissions with an exposure episode (% of total admissions)	1.113 (7,7%)	822 (5,6%)	637 (4,2%)	642 (4,1%)	3.214 (5,3%)

Figure 1. Table 5.2: frequency of admissions with an exposure fragment within each available registration year.

The distribution of length of stay differs between admissions with and without overflow exposure. Both distributions are presented in Figure 5.2. Among unexposed admissions the median length of stay was 2,5 days (IQR 1,2–5,3), whereas among exposed admissions it was 5,4 days (IQR 2,4–10,9), doubled at every quartile. The difference widens further in the tail: the 99th percentile lies at 26,7 days for unexposed admissions and at 51,4 days for exposed ones. Both distributions are right-skewed, but the exposed distribution is shifted to the right and has more weight in its tail. The figure shows both distributions up to 15 days; the number of admissions beyond that point is reported in the title. Because the vertical axes are adjusted for the difference in group size, the panels should be compared on shape and on the marked percentiles rather than on bar heights.

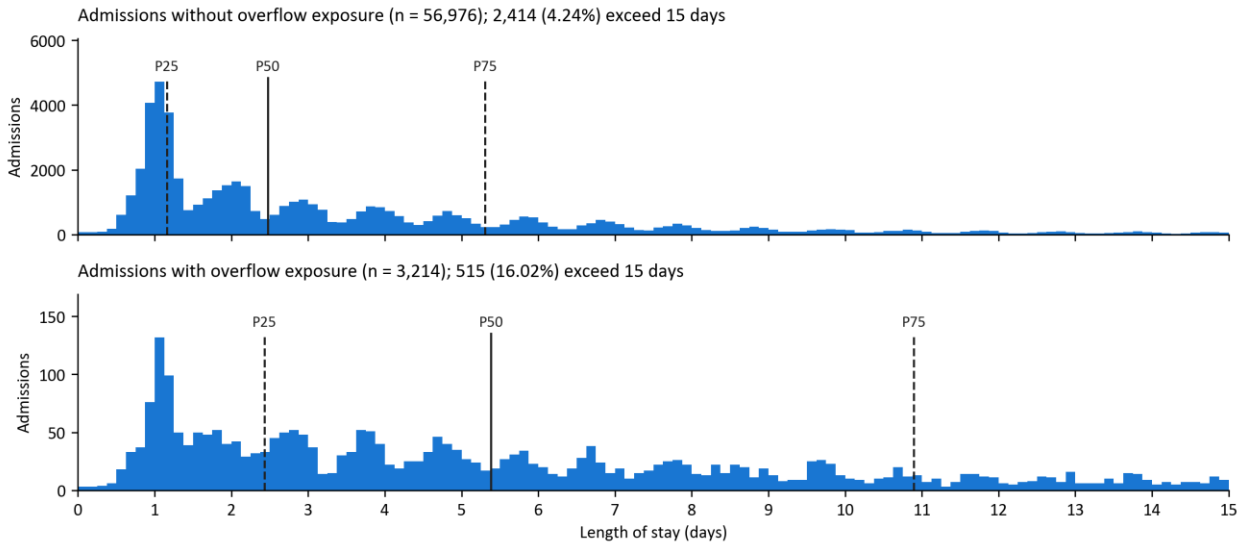

Figure 5.2: Length of Stay distribution for patients not exposed to overflow beds (upper graph) versus those exposed to overflow beds (lower graph)

5.2 Phase 1

This section presents the results of Phase 1, in which the case-mix-adjusted risk of L-LOS associated with overflow bed exposure is examined.

Risk dimensions

The results are presented for each of the four dimensions of exposure in turn: whether a patient was exposed to an overflow bed at all, how long the exposure lasted, at which moment in the admission it occurred, and on which ward the patient was placed. For *general exposure*, the observed risk, the case-mix-adjusted risk and the odds ratio are reported. Comparing the observed with the adjusted risk shows how much of the raw difference between exposed and unexposed admissions is explained by case mix, and the adjusted risks of both groups give this difference in absolute terms. Once this has been shown for the main comparison, the other dimensions focus on how their categories differ from unexposed admissions, for which the odds ratio from a single model is the most direct measure.

General exposure

The first dimension distinguishes admissions with any overflow exposure from those without. Table 5.3 gives the raw cross-tabulation. Of the 60.190 admissions in the study population, 3.214 (5,3%) involved at least one overflow episode. Within that group, 1.523 admissions (47,4%) exceeded the L-LOS threshold of their own main cluster, against 13.525 of the 56.976 unexposed admissions (23,7%). The unadjusted risk of a long admission is thus twice as high among exposed admissions.

	No Exposure	Exposure	Total
Normal Length of Stay	43.451 (76,3%)	1.691 (52,6%)	45.142 (75,0%)
Long Length of Stay	13.525 (23,7%)	1.523 (47,4%)	15.048 (25,0%)

Table 5.3: Distribution of L-LOS around the population for general exposure

The observed risk of 47,4% is the raw proportion. Part of this raw difference may reflect case mix rather than exposure. Table 5.4 therefore presents the case-mix-adjusted risks and the odds ratio. After correction for case mix, the risk of an L-LOS is 34,4% for exposed admissions and 24,4% for unexposed admissions. The latter is close to the observed 23,7%, because unexposed admissions make up most of the population on which the expected risks are based. The gap between the two groups thus narrows from 23,7 to about 10 percentage points: case mix explains more than half of the raw difference, but a substantial difference remains associated with exposure.

	Not exposed	Exposed patients
Number of admissions	56.976	3.214
Observed risk	23,7%	47,4%
Case-mix adjusted risk	24,4%	34,4%
Odds ratio		2,03 (CI:1,87-2,20)

Table 5.4: Adjusted risks and odds ratios on the risk of L-LOS in general exposure

The odds ratio expresses this remaining difference in relative terms. With an odds ratio of 2,03 (95% CI 1,87–2,20), exposed admissions have just over twice the odds of an L-LOS of unexposed admissions with a comparable risk profile. Even after correction for case mix, overflow exposure is thus associated with a long length of stay.

Duration of exposure

This section examines whether the risk of L-LOS depends on how long a patient was exposed, measured both in absolute time and as a share of the admission. The two measures can tell a different story: a short exposure may be a small part of a long admission, while a long exposure may cover almost an entire short one. The section therefore starts with a cross-tabulation of the two measures, which shows how they combine and is needed to interpret the odds ratios for absolute and relative exposure time that follow.

Two perspectives on duration of exposure

Table 5.5 cross-tabulates *absolute exposure time* against *relative exposure time* for the 3.214 exposed admissions. The columns describe how duration of overflow is distributed relative to the total admission time, the rows describe the overflow time in absolute terms. Cells are percentages of the total exposed admissions. Cells with an asterisk (*) have values of less than 1%.

As the following sections show, the odds ratios for absolute and relative exposure time sometimes point in opposite directions. The cross-tabulation shows their joint distribution, so that both sets of results can be interpreted together.

The table shows that the distribution of overflow is heavily concentrated in the two outer columns. Specifically, 27,4% of exposed admissions spent less than 10% of their stay in an overflow bed (with 23,7% spending less than a day), while 33,4% spent between 90% and 100% of their stay there. Together, these two groups account for 60,8% of the exposed population. Furthermore, for those who spent 90–100% of their admission in an overflow bed, half of these stays lasted between 1 and 3 days.

	Between 0 and 10%	Between 10 and 20%	Between 20 and 30%	Between 30 and 40%	Between 40 and 50%	Between 50 and 60%	Between 60 and 70%	Between 70 and 80%	Between 80 and 90%	Between 90 and 100%	Total
Between 0 and 8 hours	17,0%	*	*	*	*	*	*	*	*	*	18,2%
Between 8 and 16 hours	3,6%	1,7%	*	*	*	*	*	*	*	*	7,7%
Between 16 and 24 hours	3,1%	2,7%	1,6%	*	*	*	*	*	*	3,3%	13,4%
Between 24 and 32 hours	1,8%	1,5%	*	*	*	*	*	*	*	6,6%	12,0%
Between 32 and 40 hours	*	*	*	*	*	*	*	*	*	2,6%	4,8%
Between 40 and 48 hours	*	*	*	*	*	*	*	*	*	2,6%	6,7%
Between 2 and 3 days	*	1,3%	1,5%	*	1,0%	*	*	*	*	5,5%	12,2%
Between 3 and 4 days	*	*	*	*	*	*	*	*	*	3,0%	7,2%
Between 4 and 5 days	*	*	*	*	*	*	*	*	*	2,5%	5,4%
Between 5 and 6 days	*	*	*	*	*	*	*	*	*	1,8%	3,7%
Between 6 and 7 days	*	*	*	*	*	*	*	*	*	1,2%	2,2%
Longer than 7 days	*	*	*	*	*	*	*	*	*	3,2%	6,4%
Total	**27,4%**	**10,2%**	**7,2%**	**4,9%**	**4,6%**	**3,7%**	**3,4%**	**3,0%**	**2,1%**	**33,4%**	**100,0%**

Table 5.5: Cross-table of overflow exposure and its duration. The table's rows describe the duration an admission took in absolute terms, and the columns describe the exposure as a percentage of the total admission time. The cells are the percentages of the total number of exposed admissions. Cells with values less than 1% have an asterisk.

Absolute time of exposure

Table 5.6 reports odds ratios by *absolute exposure time*. Rows in italics did not reach significance at the 5% level. The highest odds are found at the two ends of the scale. Admissions with under eight hours of overflow time have 4,74 times the odds of a long stay of unexposed admissions. The estimates then fall towards and below one: the categories between 24 hours and four days do not differ significantly from the reference. After that, the odds rise steeply again, to 72,22 (95% CI 31,48–165,70) for admissions with more than seven days of overflow time.

The high odds and wide confidence intervals for the groups high on the scale can be explained: an admission cannot record seven days on an overflow bed without lasting at least seven days, and seven days already exceeds the L-LOS threshold in 12 of the 14 main clusters (Section 3.3). Admissions in the highest duration categories are therefore close to being flagged as L-LOS by construction.

The cross-table above indicates how absolute admission durations are composed in terms of relative overflow time. The group *between 0 and 8 hours* consists mostly (93%) of patients whose overflow time accounted for only 0 to 10 percent of the entire admission. This shows that while they are placed briefly in an overflow bed, this exposure is only a small part of a much longer overall admission. On the other hand, for patients whose absolute duration lasted between 1 and 4 days (with odds ratios with no significant effect), the relative time is mostly concentrated between 90 and 100%. This indicates that for these patients, almost their entire admission consisted of overflow exposure.

	Number of admissions	Odds ratios	
Between 0 and 8 hours	585	4,74	(CI: 3,93:5,72)
Between 8 and 16 hours	248	1,48	(CI: 1,12:1,97)
Between 16 and 24 hours	431	1,39	(CI: 1,12:1,73)
Between 24 and 32 hours	387	*1,19**	*(CI: 0,92:1,53, p = 0,18)*
Between 32 and 40 hours	155	*0,81**	*(CI: 0,55:1,20, p = 0,29)*
Between 40 and 48 hours	214	*0,97**	*(CI: 0,71:1,32, p = 0,82)*
Between 2 and 3 days	393	*1,04**	*(CI: 0,83:1,30, p = 0,74)*
Between 3 and 4 days	231	*1,33**	*(CI: 1,00:1,76, p = 0,05)*
Between 4 and 5 days	174	2,23	(CI: 1,61:3,08)
Between 5 and 6 days	120	3,31	(CI: 2,21:4,97)
Between 6 and 7 days	71	7,38	(CI: 4,05:13,46)
Longer than 7 days	205	72,22	(CI: 31,48:165,70)

Table 5.6: Odds ratios on the risk of L-LOS categorised for duration of exposure in absolute terms.

Relative time of exposure

Table 5.7 reports odds ratios by exposure time as a proportion of total admission time. The estimates of the odds ratios decline across the deciles, from 8,95 for admissions spending under 10% of their stay on an overflow bed, through 1,60 at 30–40%, to 0,40 for the 1.073 admissions spending between 90% and 100% of the admission there. Only the *80–90%* category does not differ significantly from the reference. The estimate in the category *90-100%* is the only one reported in this section that falls below 1.

The cross-tabulation in Table 5.5 helps to profile the admissions in the highest decile. As shown in the table, admissions spending 90–100% of their stay on an overflow bed are relatively short: half of these

admissions lasted between 1 and 3 days in total. In other words, patients who spend their entire admission on an overflow bed generally have a short overall hospital stay.

On the other hand, patients who spend only 0-10% of their admission on an overflow bed have high odds. The absolute overflow time of these admissions is concentrated at the low end. Almost two-thirds of these patients spend between 0 and 8 hours in overflow, while the remainder experience more than 8 overflow hours.

	Number of admissions	Odds ratios	
Between 0 and 10%	882	8,95	(CI: 7,51:10,65)
Between 10 and 20%	329	3,97	(CI: 3,09:5,10)
Between 20 and 30%	233	2,55	(CI: 1,92:3,39)
Between 30 and 40%	157	1,60	(CI: 1,14:2,26, p = 0,01)
Between 40 and 50%	147	2,28	(CI: 1,60:3,25)
Between 50 and 60%	118	1,52	(CI: 1,02:2,28, p = 0,04)
Between 60 and 70%	109	1,58	(CI: 1,04:2,40, p = 0,03)
Between 70 and 80%	97	1,86	(CI: 1,20:2,88, p = 0,01)
Between 80 and 90%	69	*1,34**	*(CI: 0,79:2,28, p = 0,27)*
Between 90 and 100%	1.073	0,40	(CI: 0,33:0,48)

Table 5.7: Odds ratios on the risk of L-LOS categorised for duration of exposure in relative terms.

This section shows that the risk of L-LOS does not increase with the time spent on an overflow bed. The highest odds are found for brief exposures, of less than a day or less than 10% of the admission, which mostly occur within long admissions. Admissions spent almost entirely on an overflow bed are mostly short and carry a lower risk than unexposed admissions, while the high odds for exposures of more than four days largely follow from the construction of the outcome. What matters for the risk is therefore not how long a patient stays on an overflow bed, but in what kind of admission the exposure occurs.

Moment of exposure

The admissions were divided into three groups based on the *moment of exposure*: *first moment onward*, *second moment onward*, and the *final moment*. Before the results of the associated risks are displayed, table 5.8 demonstrates that the admission profiles differ among the groups. The columns show, for each group, the number of admissions, the median length of stay and the mean percentage of the length of stay spent on an overflow bed. The last column of the table summarises notable differences from table 5.5.

Admissions exposed from the *1st moment onward* are the largest group (n = 2.276). This group has a median length of stay of 3,85 days and a high proportion of overflow time. Over 46% of these admissions spent 90-100% of their stay in an overflow bed, with the overall average sitting at 63% across the entire group. In contrast, admissions exposed from the *second moment onward* (n = 808) show the opposite profile, with a median length of stay of 11,77 days. Over 60% of the admissions in this group spent between 0-10% of their time in an overflow bed. On average, 10% of their total hospital stay was spent in overflow.

In short, exposure on the *1st moment and onward* typically characterizes a short admission dominated by overflow time, whereas overflow time for placement on the *2nd moment and onward* accounts for only a small fraction of a much longer hospitalization.

	Number of admissions	Median length of stay	Average % overflow time of admission	Notable differences
1st moment and onward	2.276	3,85 days	63%	46% of admissions spent 90-100% of their stay on an overflow bed
2nd moment and onward	808	11,77 days	10%	60% of admissions spent 0-10% of their stay on an overflow bed
Last moment	130	4,94 days	48%	

Table 5.8: Number of admissions, median length of stay and average relative time on overflow bed. For exposure categories of moment of admission

Figure 5.3 shows how the moment groups differ in when they occur. The left panel shows the number of exposed admissions per month of admission, summed over 2022–2025; the right panel shows the number per registration year. Exposure from the *first moment onward* follows a seasonal pattern: it is highest in the winter months, declines towards a low in summer and rises again in autumn. The other two groups remain relatively stable across the months of the year. Exposure from the first moment onward also decreased over the study period, from almost 900 admissions in 2022 to fewer than 400 in 2025. Exposure from the *second moment onward* remained stable and rose slightly in 2025, and exposure in the *final moment* remained small.

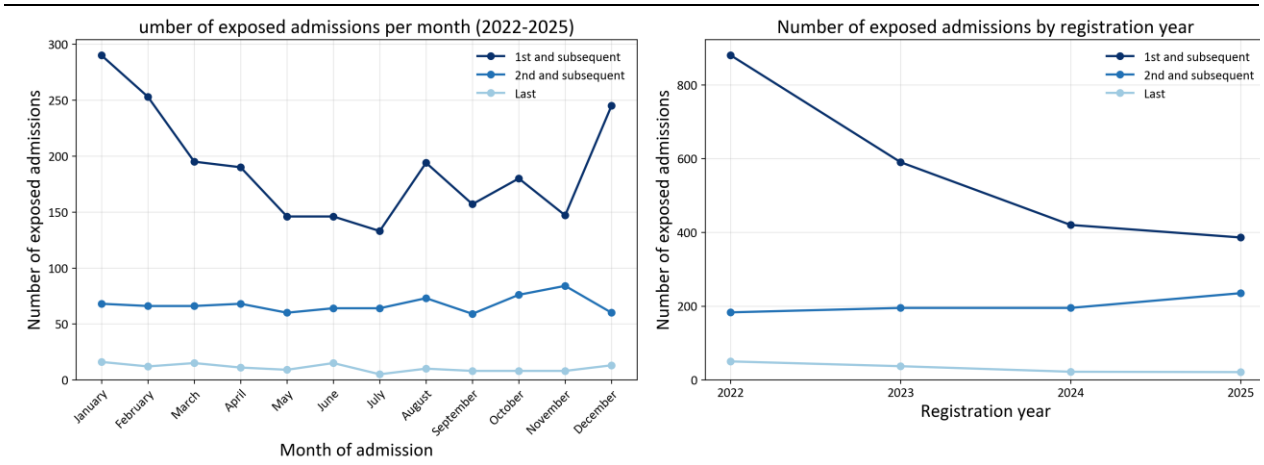

Figure 5.3: Number of exposed patients per month (left) and per registration year (right) divided into the exposure categories of moment of exposure.

Table 5.9 reports the corresponding odds ratios for each group. Exposure from the *first moment onward* gives an odds ratio of 1,17, exposure in the *final moment* 1,72, and exposure beginning at the *second moment or later* 11,27.

Odds ratios		
1st moment and onward	1,17	(CI: 1,06:1,29)
2nd moment and onward	11,27	(CI: 9,3-13,66)
Last moment	1,72	(CI: 1,17-2,54)

Table 5.9: Odds ratios on the risk of L-LOS categorised for moment of exposure

In 249 of the 808 (30,8%) admissions in *2nd moment and onward* group, the admission had already exceeded the L-LOS threshold of its main cluster before the first overflow episode began. Table 5.10 repeats the analysis, but with those admissions that already surpassed their L-LOS threshold excluded, so that only exposures occurring before the threshold was reached are counted. The estimate for exposure from the *first moment onward* is stable. The estimate for exposure beginning at the second moment or later falls from 11,27 to 7,28. The estimate for exposure in the *final moment* falls from 1,72 to 1,14 and no longer reaches significance.

	Number of admissions	Odds ratios	
1st moment and onward	2.264	1,14	(CI: 1,03:1,25)
2nd moment and onward	559	7,28	(CI: 5,93-8,93)
Last moment	112	*1,14**	*(CI: 0,74-1,76, p = 0,54)*

Table 5.10: Odds ratios on the risk of L-LOS categorised for moment of exposure after excluding admissions that already surpassed their L-LOS threshold on the moment of exposure

This section shows that the *moment of exposure* distinguishes groups of admissions with very different profiles and risks. Admissions exposed from the *first moment onward* are mostly short and spent most of their stay on an overflow bed, whereas admissions exposed from the *second segment onward* are long, with overflow time accounting for only a small share of the stay. Only exposure from the first segment onward follows a seasonal pattern, and it is also the only group that decreased over the study period. Its odds ratio for L-LOS is (modestly) above 1 (1,17). Exposure from the second segment onward

carries by far the highest odds ratio (11,27). Part of this estimate reflects admissions that had already exceeded their L-LOS threshold before the exposure began. After excluding these, the odds ratio falls to 7,28. Exposure in the *final segment* concerns a small group, and its association with L-LOS is no longer significant once these admissions are excluded. The association between overflow exposure and L-LOS depends strongly on the moment at which exposure occurs, and is concentrated in exposure that starts later in the admission.

Place of exposure

The fourth dimension concerns the ward on which the patient was placed, evaluated against the flowcharts described in Section 4.2. For each specialty and ward, these flowcharts group the alternative wards into tiers of preference, to be considered in turn when the appropriate ward is full. Placements are classified into five categories. Three follow the tiers of the flowchart: the *1st preferred group of alternative wards*, the *2nd preferred group of alternative wards* and the *last preferred group of alternative wards*. Two fall outside it: placement on a *ward not in the flowchart*, and placement for which there is *no applicable flowchart*.

Table 5.11 reports the odds ratios for each category. All five differ significantly from the reference. The *1st* and *2nd preferred groups of alternative wards* give the lowest estimates, 1,77 and 1,54 respectively, with overlapping confidence intervals. The *last preferred group of alternative wards* gives 3,54, and placement on a *ward not in the flowchart* gives 2,36. The highest estimate, 8,81, is found for the 162 admissions for which there was *no applicable flowchart*.

	Number of admissions	Odds Ratios	
No applicable flowchart	162	8,81	(CI: 6,09:12,74)
Ward not in flowchart	361	2,36	(CI: 1,87:2,98)
1st preferred group of alternative wards	1.322	1,77	(CI: 1,57:2,00)
2nd preferred group of alternative wards	1.054	1,54	(CI: 1,34:1,7)
Last preferred group of alternative wards	315	3,54	(CI: 2,78:4,51)

Table 5.11: Odds ratios on the risk of L-LOS categorised for place of exposure

The categories differ not only in their estimates, but also in the patients involved and in their intended and actual ward. The group with *no applicable flowchart* (odds ratio 8,81, n = 162) consists mostly (91%) of patients registered under the ICU who were placed on another ward.

The *last preferred group of alternative wards* is the fallback when none of the wards in the other groups is available. In every flowchart, this group consists of the ICU. All admissions in this category therefore concern patients from other specialties who were placed on the ICU.

The category *ward not in flowchart* covers admissions placed on a ward that does not appear in any of the three groups of the applicable flowchart. In 59% of these admissions, the ward of placement was the cardiac emergency and coronary care unit (EHH/CCU). Some patients need specialised care during their stay that only the team of that ward can provide; an example is the placement of a midline.

This section shows that most overflow placements follow the flowcharts: around three out of four exposed admissions were placed on a ward in the *1st* or *2nd preferred group of alternative wards*. These placements carry the lowest estimates, but both still remain above 1 and the two groups do not differ meaningfully from each other. The higher estimates are found for placements outside these groups. However, these categories involve specific populations: the two highest estimates concern the ICU, either as the ward of placement or as the specialty of the patient, and placements on a *ward not in the flowchart* mostly concern the EHH/CCU.

Summary of phase 1

Phase 1 addressed the first subquestion: *what is the risk of L-LOS associated with overflow bed exposure, after adjusting for case mix?* Without correction, 47,4% of exposed admissions had an L-LOS, against 23,7% of unexposed admissions. After correction for case mix, this was 34,4% against 24,4%. The analysis showed that exposed admissions had more than twice the odds of an L-LOS compared with unexposed admissions. The other three dimensions give a more refined explanation of where the risk comes from.

Duration and moment of exposure describe the same two groups from different angles. The first and largest group consists of short admissions spent largely on an overflow bed, usually from the start of the admission (n = 2.276; median 3,85 days). In this group, the odds of an L-LOS are 17% higher than for comparable unexposed admissions. The second group consists of long admissions in which the patient first stayed on the appropriate ward and was placed on an overflow bed only later, usually briefly (n = 808; median 11,77 days). This group shows the largest increase: even after excluding admissions that had already exceeded their L-LOS threshold before the exposure began, the odds of an L-LOS are still about seven times as high. The duration analysis shows the same pattern, with the largest increases for exposures shorter than eight hours, or shorter than 10% of the admission. Longer exposure is therefore not associated with a larger increase of risk. Moreover, only exposure from the start of the admission follows a seasonal pattern and decreased over the study period, whereas exposure later in the admission remained stable.

For place of exposure, placements on a ward in the 1st or 2nd preferred group of alternative wards (for 3 in 4 exposed patients) show the smallest increase, with odds about 1,8 and 1,5 times as high. The increase is larger outside these groups: about 2,4 times for a ward not in the flowchart, 3,5 times for the last preferred group, which is always the ICU, and 8,8 times for admissions with no applicable flowchart, mostly patients registered under the ICU but placed on another ward.

5.3 Phase 2

This section presents the results for Phase 2, in which the extent to which mortality affects the association between overflow exposure and L-LOS is examined. It first reports the correlation between the expected probabilities of mortality and L-LOS. It then turns to the distribution between the three variables: mortality, exposure and L-LOS. Finally, it repeats the analyses from Phase 1 on a population restricted to survivors.

Overlap between high risk on mortality and on L-LOS

In table 5.12 the calculated correlations between expected mortality and expected L-LOS are given. Spearman's rank correlation of 0z,81 indicates a strong correlation (Schober et al., 2018), meaning that the two models place admissions in almost the same order: those the CBS model identifies as having a high risk of death are largely the same admissions as those likely to exceed their length-of-stay

threshold. Pearson's correlation on the logit scale, at 0,77, is close to it and shows that the relationship is broadly linear on the scale on which both models are estimated. The lower value on the probability scale (0,54) is in line with the skewed distribution of expected mortality instead of a weaker relationship. Most admissions have an expected mortality below a few per cent, and a correlation calculated on that scale is influenced largely by the small group of admissions with the higher risks.

	Correlation
Pearson on probabilities	0,541
Pearson on logits	0,770
Spearman	0,811

Table 5.12: Correlations between expected L-LOS and expected mortality.

The correlations were also calculated within each main cluster. Spearman's ρ was at least 0.50 in every cluster and exceeded 0.80 in 9 of the 14 clusters. *The lowest correlations were found for infectious and parasitic diseases* (ρ = 0.50) and *diseases of the nervous system and sense organs* (ρ = 0.60).

Because the model producing expected mortality identifies the same admissions as high risk as the model producing expected L-LOS, it can also largely predict which admissions will exceed their threshold.

Association of mortality, exposure and L-LOS

In table 5.13 the observed mortality risk is given for patients with normal length of stay and L-LOS. Mortality is more than twice as high among admissions with a long length of stay: 698 of 15.048 L-LOS admissions ended in death (4,6%), against 862 of 45.142 admissions (1,9%). Across the study population, 1.560 admissions (2,6%) ended in in-hospital death.

	Normal LOS	L-LOS
Mortality	1,9%	4,6%

Table 5.13: distribution of mortality among admissions with L-LOS and those with normal LOS

Splitting the table further into exposure and no exposure results in Table 5.14. Within unexposed admissions, mortality is higher among admissions with an L-LOS than among those with a normal length of stay (4,3% versus 1,8%). This is an absolute difference of 2,5 percentage points, and a relative difference of 2,4 times as high. Within exposed admissions, the difference is smaller (7,6% versus 5,9%): 1,7 percentage points, or 1,3 times as high. The difference is thus somewhat smaller in absolute terms and considerably smaller in relative terms. This is mainly because mortality is already high among exposed admissions with a normal length of stay: at 5,9%, it is higher than among unexposed admissions with an L-LOS (4,3%). An L-LOS is therefore associated with higher mortality in both groups, but exposed admissions have a higher mortality regardless of their length of stay.

	No Exposure		*General exposure*	
	Normal LOS	**L-LOS**	**Normal LOS**	**L-LOS**
Mortality	1,8%	4,3%	5,9%	7,6%

Table 5.14: Distribution of mortality among admissions with L-LOS and normal LOS, stratified by general exposure.

Analysis of survivors

In table 5.15 the odds ratios are presented of overflow exposure within the general population and among survivors. In the full population the case-mix-adjusted odds ratio is 2,03; among survivors it is 2,07. The confidence intervals overlap almost entirely. This means that mortality does not seem to alter the association between overflow bed exposure and L-LOS.

	Survivors only		*Population*	
	Odds Ratio		**Odds Ratio**	
Overflow exposure	2,07	(CI: 1,91:2,25)	2,03	(CI: 1,87:2,20)

Table 5.15: Odds ratios of overflow exposure including the whole population (right table) and including only those surviving their hospital stay (left table).

The survivor analysis was also repeated for the other three dimensions of exposure from Phase 1: duration, moment and place. All showed the same pattern: among survivors, the odds ratios were slightly higher than in the full population, with overlapping confidence intervals.

Summary of phase 2

Phase 2 addressed the second subquestion: *to what extent is the relationship between overflow bed exposure and L-LOS distorted by mortality?* The first analysis compared the expected mortality from the CBS HSMR model with the expected L-LOS from the cluster models, both based on patient characteristics at admission. The two are strongly correlated (Spearman's $\rho = 0,81$): the admissions at high risk of dying are largely the same admissions as those at high risk of an L-LOS.

The second analysis looked at observed mortality. Exposed admissions had a higher mortality than unexposed admissions (6,7% versus 2,4%), as Vloet (2025) also found. In both groups, mortality was higher among admissions with an L-LOS than among those with a normal length of stay, but this difference was larger among unexposed admissions, in both absolute and relative terms. Moreover, exposed admissions with a normal length of stay had a higher mortality (5,9%) than unexposed admissions with an L-LOS (4,3%).

The final analysis repeated the Phase 1 analyses among survivors only. Excluding the 1.560 admissions that ended in death changed the odds ratio for general exposure only marginally, from 2,03 to 2,07, with almost fully overlapping confidence intervals. The other dimensions showed the same pattern.

In short, mortality hardly distorts the association between overflow exposure and L-LOS. Moreover, L-LOS and mortality largely identify the same admissions as being at high risk, so combining them in one

model would add little. Phase 3 therefore focuses on separate prediction models for L-LOS and in-hospital mortality.

5.4 Phase 3

Data selection & train and test

The prediction models were developed on acute admissions only. Table 5.16 shows, per registration year, the number of acute admissions with overflow exposure and the prevalence of both outcomes. The share of admissions with overflow exposure decreased from 10,2% in 2022 to 5,2% in 2025. The prevalence of L-LOS decreased slightly over the same period, from 32,5% in 2022 to 30,4% in 2025. In-hospital mortality ranged from 3,4% to 3,9%.

	2022	2023	2024	2025
Exposure to overflow	1.003 (10,2%)	712 (7,2%)	550 (5,3%)	556 (5,2%)
Prevalence of L-LOS	32,5%	32,2%	31,0%	30,4%
Mortality Rate	3,9%	3,6%	3,4%	3,5%

Table 5.16. Frequency of exposure in train/validation/test datasets, acute admissions only

In addition to the main split, the models for L-LOS were evaluated in all twelve possible combinations of training, validation and test years, as described in Section 4.4.

Predictor definition and encoding

The encoding of the seven predictor variables resulted in 45 predictors, listed with their reference categories in Appendix G.

Selecting the models

In appendix E the hyperparameters selected for each model after randomized search are presented. For these, registration year 2024 was used as validation set. For L-LOS, XGBoost selected a maximum tree depth of 2, the lowest value in the search space. For mortality, it selected a maximum depth of 3. The random forest selected a maximum depth of 10 for both outcomes, rather than 20 or unlimited depth.

Analysing the models

Table 5.17 presents the performance of the four models on the 2025 test set. For L-LOS, the four models performed very similarly. The AUC ranged from 0,68 (decision tree) to 0,70 (XGBoost), and the scaled Brier score from 0,083 (decision tree) to 0,099 (XGBoost). XGBoost performed best on both measures but with only a small advantage. All models had a negative calibration intercept between −0,097 and −0,089, meaning that each model overestimated the risk of L-LOS in 2025 to an extent.

For mortality, discrimination was higher than for L-LOS. Logistic regression achieved the highest AUC (0,85), followed by the random forest and XGBoost (both 0,84) and the decision tree (0,81). XGBoost achieved the highest scaled Brier score (0,086), compared with 0,081 for logistic regression. All models again overestimated the risk. Three out of four models had calibration slopes above 1 (1,14; 1,24 and 1,19 respectively) indicating predictions that were too moderate.

	AUC	Brier	Scaled Brier	Intercept α	Calibration slope β
Logistic Regression	0,69	0,192	0,093	-0,097	0,973
Decision Tree	0,68	0,194	0,083	-0,095	0,916
Random Forest	0,69	0,191	0,097	-0,089	1,123
XGBoost	0,70	0,191	0,099	-0,091	1,025

	AUC	Brier	Scaled Brier	Intercept α	Calibration slope β
Logistic Regression	0,85	0,030	0,081	-0,076	1,139
Decision Tree	0,81	0,031	0,063	-0,062	0,761
Random Forest	0,84	0,030	0,082	-0,062	1,239
XGBoost	0,84	0,030	0,086	-0,040	1,192

Table 5.17: Results of the predictions of L-LOS (upper graph) and mortality (lower graph)

Mortality was thus predicted with better discrimination than L-LOS, while the scaled Brier score was slightly lower for mortality than for L-LOS. The AUC only reflects how well a model ranks admissions. The scaled Brier score also reflects how close the predicted probabilities are to the observed outcomes, and for a rare outcome such as mortality the room for improvement over the prevalence is smaller.

Because mortality is a rare outcome, the AUC does not show how often a high-risk prediction is correct. Appendix H therefore shows sensitivity, false positive rate and precision when a given share of admissions with the highest predicted risk is flagged as high risk. When 10% of admissions are flagged, the mortality model identifies about 48% of the deaths, against 18% of the long stays for the L-LOS model. However, only about 17% of the admissions flagged by the mortality model end in death, against 56% for L-LOS. The mortality model thus separates admissions better, but most admissions it flags as high risk survive.

Supplying socio-economic status, month of admission, origin of admission and severity of the main diagnosis as categorical variables instead of dummy variables did not improve XGBoost.

To inspect what each predictor added, SHAP values were used to examine the contribution of each predictor to the predicted probability of each admission in the random forest and XGBoost models. For each predictor, the contributions formed two clearly separated groups: one for admissions with value 1 and one for admissions with value 0. Within each group, the contributions lay within a narrow range, so that admissions with the same value for a predictor received nearly the same contribution, regardless of the values of the other predictors, indicating few interactions.

Summary phase 3

Phase 3 addressed the third subquestion: *which model best predicts L-LOS and in-hospital mortality?*

The mortality models distinguished admissions at high risk from admissions at low risk better than the L-LOS models did, with an AUC of 0,85 for the best mortality model against 0,70 for the best L-LOS model. For both outcomes, the machine learning models did not perform substantially better than logistic regression: the decision tree, random forest and XGBoost reached AUC and scaled Brier scores similar to those of logistic regression, which even achieved the highest AUC for mortality. All

models overestimated the risk of both outcomes in the 2025 test year to a similar extent, and performance differed more between test years than between models. The SHAP values and the selected hyperparameters point in the same direction: the machine learning models found hardly any non-linear relationships, and the best-performing configurations were of low complexity.

In short, no model clearly outperformed logistic regression, which, given its interpretability, is the most suitable model for both outcomes. The limit to predicting L-LOS appears to lie in the information available at admission rather than in the choice of model.

5.5 Sensitivity analysis

The sensitivity analysis examined whether the results change under the stricter definition of L-LOS at the 95th percentile. In Phase 1, the direction and pattern of the results held: exposed admissions again had higher odds of an L-LOS95 than comparable unexposed admissions (odds ratio 2,92, against 2,03 in the main analysis), and exposure from the first and second moment onward and all place categories remained significantly associated with the outcome. In Phase 2, excluding deaths again changed the odds ratio only marginally (from 2,92 to 3,04), and in Phase 3, logistic regression again performed as well as the machine learning models (AUC 0,72 against 0,69–0,71).

The second sensitivity analysis examined whether the results of Phase 3 depend on the assignment of years to training, validation and testing. Across the twelve combinations, the AUC of logistic regression for mortality ranged from 0,82 to 0,85 and the scaled Brier score from 0,067 to 0,088, and the other models showed similar ranges. Performance differed more between test years than between models, and within a test year the choice of validation year hardly affected the results. The main configuration lay within these ranges, so the conclusion that no model clearly outperformed logistic regression does not depend on the choice of years.

6 Discussion

This chapter discusses the results presented in Chapter 5. Section 6.1 reflects on the key findings of this thesis, section 6.2 discusses the limitations of the research and how these may have affected the results. Next, section 6.3 gives an interpretation of the results and describes how the findings address the research gaps identified, Section 6.4 discusses implications for the hospital and Section 6.4 sets out recommendations for future research.

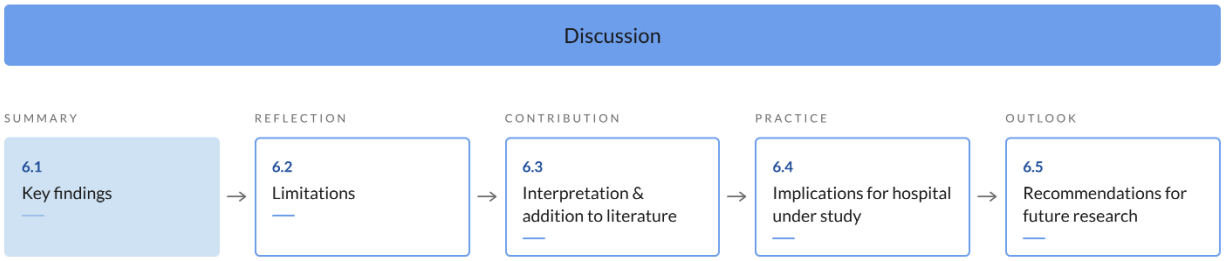

Figure 6.1: Overview of the structure of Chapter 6.

This thesis set out to answer the question: *to what extent is overflow bed exposure associated with long length of stay, and how can this risk be predicted alongside in-hospital mortality?* To answer it, the research was carried out in three phases.

Phase 1 addressed the first subquestion (SQ1): whether overflow bed exposure is associated with a higher risk of L-LOS after adjustment for case mix. Exposure was examined along four dimensions: whether a patient was exposed at all, for how long, at which moment during the admission, and on which ward, evaluated against the hospital's flowcharts. *Phase 2* addressed SQ2: to what extent this association is distorted by mortality, since death ends an admission and can prevent it from becoming long. It also examined whether the patients at high risk of L-LOS at admission are the same patients who are at high risk of dying. *Phase 3* addressed SQ3 by comparing logistic regression with three other machine learning models in predicting L-LOS and in-hospital mortality from the information available at admission.

6.1 Key findings

Phase 1 (SQ1) showed that overflow bed exposure was associated with a higher risk of L-LOS. After adjusting for case mix, exposed admissions had twice the odds of an L-LOS then comparable unexposed admissions (OR 2,03). Patients that are exposed directly after admission or after a flow through ward are the biggest exposure group (3 out of 4 exposed patients) and have modest increased odds of L-LOS. Patients that are exposed later in the admission have much higher increased odds. These concerned admission with short exposure duration and long total admission time. “Placements on the first and second preferred groups of alternative wards carried the smallest increase in the odds of a long stay. The increase was larger for placements on the ICU as the last resort in the flowcharts, on a ward not included in the applicable flowchart, and for admissions to which no flowchart applied.

Phase 2 (SQ2) showed that mortality did not distort this association as similar odds ratios were found among survivors as in the total population. Based on the information available at admission, the patients at high risk of L-LOS were largely the same patients as those at high risk of dying. Predicting L-LOS alongside mortality therefore flags the same patients of high risk as a single outcome model and phase 3 therefore focussed on predicting the metrics separately.

Phase 3 (SQ3) showed that no model clearly outperformed logistic regression. For both outcomes, the machine learning models performed close to logistic regression, with little sign of non-linear relationships or interactions in the data. Mortality was predicted with considerably better discrimination than L-LOS (AUC 0,85 against 0,70).

6.2 Limitations

This section discusses the limitations of this thesis and how they may have affected the results. For each limitation, it considers whether, and in which direction, it may have changed the results, and what was done to limit its influence.

Observational design and the direction of the association

This thesis is an observational study, so its results describe associations between overflow exposure and L-LOS rather than effects. Exposure was not measured at a single moment but over the course of the admission. So it could happen that exposure did not precede the outcome but was the consequence. To limit this effect and make sure that the exposure preceded the outcome, admissions that had already exceeded their L-LOS threshold before the first overflow episode were excluded in an additional analysis. Admissions also differed in how much time they had to become exposed: a long admission has more opportunity for an overflow placement than a short one. Exposure later in the admission is therefore more likely to occur in long admissions, which can inflate its association with L-LOS.

Delayed discharge

Delayed discharge may also influence the association. Patients who are medically ready for discharge sometimes remain in hospital while waiting for a place in a nursing home, a rehabilitation facility or home care. The cause of these extra days lies in the availability of follow-up care, not in the care delivered or in the overflow bed. Yet the L-LOS outcome counts them as part of the admission and thereby attributes them to the hospital. Exposed admissions included such a period more often: 13% against 5% of unexposed admissions, although this period did not necessarily coincide with the exposure. The direction may even be reversed if patients waiting for follow-up care are moved to an overflow bed to

free a bed on the appropriate ward. Part of the association between exposure and L-LOS may therefore reflect the availability of follow-up care rather than decisions of the hospital.

Confounding by indication

Patients are not placed on overflow beds at random. When the factors behind a placement decision are also related to length of stay, the comparison between exposed and unexposed admissions can be biased. The case-mix model captures only part of these factors, as the clinical judgement on which a placement is based is not recorded. The direction of this bias differs by type of exposure. For placements at the start of the admission, the Patient Flow Coordinator and the exclusion criteria in the flowcharts keep patients with complex care needs off overflow beds where possible (Mans, 2025). Although exposed patients are older and have more comorbidities on paper (Vloet, 2025), they may therefore be clinically more stable than these characteristics suggest, which would lead to an underestimation of the risk, as Bogler et al. (2021) suggested for their finding of a shorter stay. Other patients are placed on another ward precisely because of their condition: their condition or specialty changes later in the admission, or they need care that only another ward can provide, such as the EHH/CCU. This would lead to an overestimation. Reporting placements outside the flowcharts separately partly addresses the latter, as these mostly reflected such a need.

Residual confounding

The case-mix correction is limited by the patient characteristics available in the data. Characteristics present at admission that may influence both the placement decision and the length of stay are not recorded and could not be adjusted for. For example, Zannella et al. (2022) could adjust for physiological severity with a laboratory-based score, while this was unavailable for this study. Other examples are lifestyle factors, for example smoking is associated with a longer hospital stay (Hvidtfeldt et al., 2010).

The case-mix models also differed in how well they explained L-LOS. The mean explained variance (Nagelkerke's pseudo-R^2) was 0,26, ranging from 0,11 to 0,46 between main clusters. In the clusters with the weakest fit, more of the variation remains unexplained and the risk of residual confounding is larger. The case-mix models for the sensitivity analysis explained L-LOS95 less well, with a mean Nagelkerke's pseudo-R^2 of 0,15, so the risk of residual confounding is larger there.

Factors arising during the admission

The case-mix model only describes the patient at the start of the admission, so complications or new diagnoses that arise during the admission are not adjusted for. Their role depends on when they occur. Care on an overflow bed may lead to a complication, for example because deterioration is recognised later, and that complication may lengthen the stay. The complication is then an intermediate factor on the path from exposure to outcome, and adjusting for it would remove part of the effect being studied (Schisterman et al., 2009). A complication can, however, also precede the exposure: it can lead to a change of specialty and thereby to an overflow placement, while the complication itself lengthens the stay (Hoogervorst-Schilp et al., 2015). The overflow episode then marks the course of a complex admission rather than causing the long stay, although a patient with a complication who is placed on an overflow bed may still face additional risk there, as also discussed later. Both paths probably play a role, especially for exposure later in the admission. With the available data, it cannot be determined which part of the elevated risk stems from the overflow placement and which part from the course of the admission.

Administrative data and the classification of exposure

The analyses rely on routinely collected administrative data. Because administrative data are collected for purposes other than research, their completeness and accuracy depend on why and how they were recorded. Variables derived from such registrations are therefore only proxies for what they are meant to measure, and registration errors can lead to misclassification of exposures and outcomes (van Walraven & Austin, 2012). Exposure is derived from the combination of the treating specialty and the ward registered in the capacity dataset, so any error in either registration affects who is classified as exposed. Mans (2025) reported that staff at the hospital under study experienced delays and inaccuracies in the registration of patient data, such as discharges that were not yet processed in the patient administration system. The registration delay is most likely the case for Intensive Care patients placed on an overflow bed as hospital staff considered it highly unlikely that these patients were placed on an overflow bed given the specialised needs. The place dimension shows how this affects the classification of exposure. Most admissions without an applicable flowchart (91%) concerned patients still registered under the ICU after a transfer

Mortality as a competing event

To assess the influence of mortality, Phase 2 compared the odds ratio in the full population with the odds ratio among survivors only. It did not use a formal competing-risk model, as Bai et al. (2018) and Kohn et al. (2021) did. Such a model treats death and discharge as two different ways in which an admission can end and would estimate how large the competing effect of death is. The full-population analysis in this thesis counts patients who died early as admissions without L-LOS. The survivor-only analysis compares exposed survivors with unexposed survivors. Because exposure is associated with higher mortality, more exposed than unexposed patients are removed. Thus, the exposed survivors are a more selected group, which may introduce selection bias. Each analysis thus has its own source of bias. That the odds ratio differed marginally between the two (2,03 against 2,07) indicates that neither is large enough to change the conclusion. The aim was not to estimate the size of the competing effect, but to establish whether the association between exposure and L-LOS holds once mortality is taken into account.

Precision of the estimates

The expected probability of L-LOS was used as an input to the odds-ratio models. The estimation uncertainty is therefore not carried over into the odds ratios, so their confidence intervals may be too narrow. Most odds ratios lie far enough from one that this does not change their interpretation but estimates with a confidence interval close to one should be read with more caution.

Generalisability

The results apply to the hospital under study. Overflow was defined by the combination of specialty and ward. Moreover, it was evaluated against flowcharts drawn up by this hospital, and both depend on how the hospital organises its wards. As discussed in Section 2.4, what counts as an overflow bed differs between hospitals. The hospital also already applies several measures to limit the risks of overflow placement, such as the Acute Medical Unit, the transfer of overflow patients back to their appropriate ward as soon as possible, and explicit care plans for the nurses on the receiving ward (Section 2.2). In hospitals without such measures, or in other healthcare systems, the association between overflow exposure and L-LOS may differ. The approach of this thesis, decomposing exposure

into dimensions and evaluating placements against local flowcharts, can be applied in other hospitals, but its estimates cannot simply be transferred to them.

6.3 Interpretation & addition to literature

This section interprets the results in relation to the literature, following the three research gaps of Section 2.7.

The main finding, that patients placed on an overflow bed are more likely to experience a long length of stay, can be explained by the mechanisms described in Section 2.3. For instance, Goulding et al. (2012) describe how overflow patients are reviewed last on ward rounds and how nurses on the receiving ward are less familiar with their condition. Both may delay treatment, especially when complications arise, and thereby lengthen the stay. Staff at the hospital under study indicated that when the treatment plan is set late in the day, tests and treatment shift to the next day, with a longer length of stay as a direct consequence (Mans, 2025). The finding is also in line with Stylianou et al. (2017), who, like this thesis, counted any overflow placement during the admission as exposure and found a longer stay for exposed patients.

In addition, these results extend the literature to a setting that the reviewed studies did not cover. None of the studies in Section 2.4 was conducted in a Dutch hospital, and few considered the entire hospital: most focused on a single specialty, usually General Internal Medicine. Moreover, the hospital under study already takes several measures to limit the risks associated with overflow placement (Section 2.2): flowcharts guide patients to the most suitable alternative ward, the Patient Flow Coordinator keeps patients with complex care needs off overflow beds, and the Acute Medical Unit absorbs the first hours of an acute admission. This thesis therefore shows that an elevated risk of L-LOS remains, even in a hospital where such measures are in place.

Furthermore, La Regina et al. (2019) named differences in the definition of exposure as one explanation for the inconsistent findings in the literature. Most reviewed studies define exposure by the ward of admission or by the whole or most of the admission (Table 2.1), which largely corresponds to what this thesis calls exposure from the first moment onward (Section 4.2). For this form of exposure, this thesis found a modest increase in the odds of L-LOS. This is consistent with studies that found a longer stay (Alameda & Suárez, 2009; Kohn et al., 2021; Stowell et al., 2013), while others found slightly shorter stays (Perimal-Lewis et al., 2013; Zannella et al., 2022). Because the hospital under study keeps patients with complex care needs off overflow beds, this modest increase may underestimate the risk for patients who are not selected in this way (Section 6.2), and differences in such selection between hospitals may partly explain why studies of this form of exposure disagree (Bogler et al., 2021).

In contrast, this thesis also distinguishes a form of exposure that the reviewed studies did not, namely exposure later in the admission after a first stay on the appropriate ward, and this form showed by far the strongest association with L-LOS. These were mostly brief episodes, often less than 24 hours, within long admissions. The most likely explanation lies in why these patients are moved: a complication or comorbidity changes the treating specialty, and the admission is then registered as overflow until a bed on the new specialty's ward becomes available, or the patient needs specialised care from another ward (Mans, 2025). This explanation is supported by the finding that, unlike exposure at the start of the admission, this exposure showed no seasonal pattern and remained stable over time. This suggests that it is driven by the course of the admission rather than by peaks in demand, which is consistent with studies indicating that a complication or new diagnosis can cause a longer stay (Hoogervorst-Schilp et

al., 2015). Closeness to the responsible team might matter most for these patients: a complication needs staff who recognise its first signs (Goulding et al., 2012; Cheung et al., 2020), and a change of ward requires a handover that may in itself delay care (Mendelsohn et al., 2023). The strong association can therefore reflect both the course of a complicated admission and the overflow bed itself, but which part of the risk stems from which cannot be determined with these data (Section 6.2).

Finally, this thesis evaluated overflow placements against the hospital's own overflow protocol, which none of the reviewed studies did. Placements on preferred alternative wards showed a modest increase in the odds of L-LOS, whereas placement on the ICU, as the last resort, showed a clearly larger increase. That the preferred wards still carry an increased risk can be explained by the mechanisms that the flowcharts do and do not address. As described in Section 2.3, the flowcharts place patients on wards with similar care and keep patients who need specialised equipment off overflow beds, but they do not address later review, communication between teams or distance to the responsible team. In addition, part of the increase may stem from exposures later in the admission that are also placed on these wards. This first data-driven evaluation of the flowcharts, which were designed based on clinical judgement, thus supports their distinction between preferred alternatives and the ICU as a last resort, but also shows that placement on a ward with similar care does not remove the increased risk of L-LOS.

Not all placements, however, followed the flowcharts: one in six exposed admissions (523 of 3.214) fell outside them, and these showed higher odds of L-LOS than placements on the preferred wards. These mostly concern patients still registered under the ICU while located on another ward, and patients who needed specialised care that only another ward could provide, such as the EHH/CCU (Section 6.2). These placements thus mainly reflect registration delays or a specific care need rather than a shortage of beds. For these patients, the risk of a long stay is high, but the mechanism behind it is not yet clear.

Regarding the second research gap, mortality hardly distorted the association between overflow exposure and L-LOS. Although exposed admissions died more often, the odds ratio among survivors hardly differed from that in the full population. Early deaths, which may partly explain the shorter stay found by Perimal-Lewis et al. (2013), thus do not explain the association in this thesis. Moreover, at admission both outcomes largely identify the same patients as being at risk. L-LOS therefore adds little as a separate prediction target, but it does broaden what a mortality model can tell.

Regarding the third research gap, no machine learning model outperformed logistic regression in predicting mortality or L-LOS. This is in line with Christodoulou et al. (2019), who found no difference between the two in comparisons at low risk of bias. Two explanations can be given for this result. First, the predictors largely describe one underlying characteristic, namely how ill and how frail a patient is on arrival. This makes their effects largely additive, which is the ideal situation for logistic regression (Austin et al., 2021) and the machine learning models showed hardly any interactions. Second, the models appear to have reached the information ceiling of admission data: models with different structures arrived at nearly the same performance, and performance differed more between test years than between models. The limit to prediction thus appears to lie in the information available at admission rather than in the choice of model.

Moreover, this information ceiling lies lower for L-LOS: mortality was predicted with better discrimination, as Rajkomar et al. (2018) also found. Death depends mainly on how ill and how frail a patient is on arrival, which is what the predictors describe. Whether a stay becomes long, however, also

depends on what happens until discharge, partly outside the hospital (Rojas-García et al., 2018; Stone et al., 2022). L-LOS is therefore inherently harder to predict than mortality from the information available at admission. At the same time, a higher AUC does not mean that a high-risk label is more often correct. For a rare outcome such as death, a model can rank admissions well while most admissions it flags do not have the outcome (Saito & Rehmsmeier, 2015). Phase 3 showed that most admissions flagged as high risk by the mortality model survived, whereas most of those flagged by the L-LOS model did have a long stay (Appendix H). The mortality model thus ranks patients better, but its high-risk label is less often correct.

6.4 Implications for hospital under study

The introduction described two decisions for which a more complete picture of the risks associated with overflow placement is relevant: whether overflow beds are an acceptable way of absorbing peaks compared with expanding structural capacity, and where an individual patient can best be placed when their appropriate ward is full.

To begin with, the main results of this thesis provide a more comprehensive picture of the risks associated with overflow placement in the hospital under study, beyond in-hospital mortality. Death is the most severe outcome, but also a rare one: most patients whose admission does not go as planned, survive it. Such a course of admision can result in a longer stay, for example after an adverse event such as a hospital-acquired infection (Hoogervorst-Schilp et al., 2015). A long stay thus matters in two ways: for the patient, it marks an admission that deviated from its usual course, and for the hospital, it keeps a bed occupied that the next patient needs. L-LOS therefore captures consequences of overflow placement for both the patient and the hospital's capacity.

At the strategic level, these findings may also inform the trade-off between overflow beds and structural capacity. As occupancy is rarely equal across wards, overflow beds allow the hospital to use free beds elsewhere instead of expanding wards to meet peak demand. However, an admission that becomes long keeps a bed occupied for longer. If part of this association reflects an effect of the overflow placement itself, longer stays may cancel out part of the capacity gained by using these free beds. Moreover, not every exposure results from a shortage of beds: one in six exposed admissions fell outside the flowcharts and mostly concerned patients still registered under the ICU while located on another ward, or patients who needed specialised care that only another ward could provide (Section 6.3). When the hospital evaluates overflow as a consequence of insufficient capacity, for example in strategic decisions on the number of beds per ward, it should leave these admissions out.

At the operational level, as overflow beds are likely to remain part of how the hospital absorbs peaks, the relevant question for the hospital is not whether to use them, but how to limit the risks that come with them. This aligns with the recommendations of Zannella et al. (2022) to shift the focus from asking whether placing patients on overflow beds is harmful to exploring how it can be done safely. The results of this thesis discussed two of the means the hospital can use to this end: the flowcharts, which aim to place each patient on the most suitable alternative ward, and prediction models, which can separate high-risk from low-risk patients. Both fit this shift in focus, as they aim to make overflow placement safer rather than to avoid it altogether.

Regarding the flowcharts, the evaluation shows that current practice holds up in the data. Given the way patients are currently selected, placements within the preferred groups showed the smallest increase in the odds of L-LOS. Moving on to the second group when the first was full did not add to it,

whereas placements on the ICU as the last resort showed a clearly larger increase. The preferred groups still carried an increased risk, which may be explained by mechanisms that the flowcharts do not address, such as later review, communication between teams and distance to the responsible team (Section 6.3). If so, further gains in decreasing risks may come from addressing these mechanisms.

However, exposure later in the admission falls outside the situation the flowcharts were designed for: a patient who cannot be placed on the appropriate ward at admission because it is full. This exposure arises differently, after a change of specialty or a need for care on another ward and carries a strongly increased risk of L-LOS. The current measures to limit the risks associated with overflow placement, such as the flowcharts and the selection of patients by the Patient Flow Coordinator, are designed around the placement decision at the beginning of an admission. The Patient Flow Coordinator could therefore be part of a risk mitigation strategy by moving particularly these patients to the ward of their new specialty as soon as possible. Refining the current measures to include this group, for example by giving these patients priority in the daily Bed Availability Consultation, may limit the part of the risk that stems from the overflow bed. How much of this risk can be prevented depends on its mechanism, which therefore needs further study: whether the complication preceded the exposure, and whether the overflow bed affected its course.

Next to the flowcharts, prediction models are the second means to limit the risk of L-LOS, and for these the results imply that a single mortality model suffices. This thesis showed that, at admission, the patients at high risk of dying are largely the same patients as those at high risk of L-LOS. L-LOS is therefore not needed as a separate prediction target, but it does broaden what the model's output means. A patient with a high predicted risk of death is also likely to have a long stay, both as a marker of a complicated care path and as a bed that remains occupied for longer. The model does need periodic recalibration: all models overestimated the risk in the most recent year, as Vloet (2025) also found for 2023. For decision support at admission with current data available, the hospital can therefore build on the existing mortality model rather than develop a separate model for L-LOS.

Finally, the results show that the choice of model matters less than the data available. As no machine learning model outperformed logistic regression, the hospital can keep using logistic regression, which is easier to interpret and to explain to staff. To predict beyond what this model already captures, for example to better predict L-LOS or to distinguish it from mortality, information about the patient's care path is needed, such as interventions taken, medical tests needed or patient deterioration. As length of stay depends on more than the frailty of a patient, once data are available that can explain other causes of a prolonged stay (such as ward protocols or interventions taken), machine learning models other than LR may provide an advantage (Section 6.5). Whatever the model, better predictions first require more and different data than what is recorded at admission.

6.5 Recommendations for future research

The findings of this thesis raise several questions that could not be answered within the research timeline or with the available data. This section describes directions for future research. The first follows most directly from the results of this thesis.

The most direct next step is to distinguish exposures by the reason they occur. Besides a full ward, exposure can follow from a change of treating specialty or a need for a single room (Mans, 2025), and some may reflect a registration delay (Section 6.2). In the data, these are all flagged as overflow in the same way, while the results suggest that they differ in risk.

Distinguishing them matters because each reason is relevant to a different decision and may call for a different mitigation. Exposures caused by a full ward matters for strategic decisions on the number of beds per ward, whereas exposure later in the admission matters for the daily Bed Availability Consultation, which decides which overflow patient moves first. Likewise, a placement on a full ward may call for fixed review moments, whereas exposure after a change of specialty may call for a fast transfe.

A first step could be to review a sample of medical records for the two groups whose interpretation is least certain: admissions registered under the ICU while located on another ward, and admissions exposed from the second moment onward. This would show whether the first reflect registration delays, and whether the second more often involved a complication that came with a change of specialty. Future research could then distinguish the reasons for all admissions by profiling the routes through which exposure arises, based on characteristics such as the moment and duration of exposure and the preceding and following ward.

A second direction is to estimate the costs of overflow placement, which are needed to weigh overflow beds against structural capacity. Structural capacity carries a direct and budgeted cost, whereas the costs of overflow placement are less visible. They might arise through additional days in hospital, through adverse events, which are associated with higher costs per admission (Adler et al., 2018), and through beds that cannot be used for another patient (Bayley et al., 2005). An exploratory comparison at the hospital under study showed that in about [redacted] of overflow placements, the patient was placed on a ward with a higher cost price than their appropriate ward. These are cost prices and are ward averages that largely consist of fixed costs such as staff and buildings (Karabatsou et al., 2016), so they do not show whether the placement raises the cost of an admission.

A third direction concerns the prediction models, which could be improved in two ways. First, the prediction models in this thesis used only characteristics recorded at admission, and appeared to have reached the information ceiling of these data. Future models could include richer information, both on the patient's clinical condition, such as a patients health record and on choices made earlier in the admission, such as the ward of placement and the treatments given, which would give flexible models more to work with. Rajkomar et al. (2018) reached higher discrimination for length of stay using the full health record. Second, future models could predict at the moment a placement decision is made, rather than only at admission. The strongest association in this thesis was found for exposure later in the admission, and a model that updates its prediction with the information available at that moment could match the decision of the Patient Flow Coordinator more closely.

Finally, this thesis shows where the association between overflow exposure and L-LOS arises, but not what goes wrong in individual admissions. Future research could therefore examine exposed admissions that ended in a long stay, death or other risks, through a review of medical records. Complemented by interviews with the staff involved. Such a review could show whether care was delayed, a complication was recognised late or a handover went wrong, and thus which of the mechanisms in Section 2.3 actually occur in this hospital. Comparing these cases with exposed admissions that went as planned would show what distinguishes them. This would help the hospital target its mitigations at the mechanisms that matter most in its own practice.

7 Conclusion

Dutch hospitals face a growing demand for care with a limited number of beds and staff. When a patient's appropriate ward is full, an overflow bed allows the hospital to admit the patient elsewhere instead of expanding wards to meet peak demand. Whether this is an acceptable way of absorbing peaks depends on the risks it brings for patients. This thesis therefore set out to answer the question:

To what extent is overflow bed exposure associated with long length of stay, and how can this risk be predicted alongside in-hospital mortality?

The results show that the risks associated with overflow placement extend beyond mortality. After adjustment for case mix, exposed admissions had twice the odds of L-LOS. This did not change when deaths were excluded, and the increase remained when a long stay was defined more strictly. The association remained in a hospital that already takes measures to limit these risks, and thereby adds evidence from a Dutch hospital, across specialties, to a literature that mostly studied single specialties elsewhere. Most exposed patients survive their admission, but they more often have a long stay, which matters both for the patient, whose admission deviates from its usual course, and for the hospital, whose beds remain occupied for longer.

The association is not spread evenly across exposures, which helps to explain why the literature does not agree. Exposure at the start of an admission carried a modest increase, in line with comparable studies. The largest increase lay in brief exposures later in long admissions, after a first stay on the appropriate ward, a form of exposure that most definitions leave out. A complex course of admission may explain both this exposure and the longer stay, but how much of the risk stems from the overflow bed remains unclear. Where patients were placed also mattered: placements on the preferred alternative wards of the flowcharts showed the smallest increase and the ICU as a last resort a clearly larger one. Whether overflow placement is associated with a long stay thus depends on when, where and why a patient ends up on an overflow bed.

For decision support, a single model suffices to predict this risk alongside mortality. At admission, the patients at high risk of a long stay were largely the same as those at high risk of dying. A logistic regression model for mortality (such as that of Vloet (2025)) therefore already identifies the patients at highest risk of both outcomes, provided it is recalibrated over time. L-LOS thus adds little as a separate prediction target, but it does broaden what a high predicted risk means: besides the risk of death, it signals a likely complicated course and a bed that is likely to remain occupied for longer. More flexible machine learning models did not predict either outcome better, and L-LOS was predicted

considerably less well than mortality. As admission data describe how ill a patient is on arrival, while whether a stay becomes long depends largely on what happens afterwards. Better prediction of L-LOS therefore requires information on the course of the admission, not a more complex model.

For the hospital under study, these findings have several implications. First, they broaden the picture of the risks associated with overflow placement beyond mortality, as a long stay affects both the patient and the hospital's capacity. Not every exposure, however, results from a shortage of beds: one in six exposed admissions fell outside the flowcharts and should be left out when overflow is evaluated as a capacity problem, for example in decisions on the number of beds per ward. The flowcharts themselves hold up in the data, but do not cover exposure later in the admission. Including these patients in current measures, for example by giving them priority in the daily Bed Availability Consultation, may limit the part of the risk that stems from the overflow bed.

The dimensions of exposure show where the risk lies, but not why. Future research could distinguish exposures by the reason they occur, such as a full ward or a change of specialty, as these may carry different risks through different mechanisms and therefore call for different decisions and mitigations. Another is to review the medical records of exposed admissions, to see which mechanisms occur in this hospital, such as staff on the receiving ward being less familiar with the patient's condition or distance to the responsible team. With such knowledge, the hospital can move from asking whether overflow placement carries risks to deciding how to limit them where they arise.

References

Ackroyd-Stolarz, S., Read Guernsey, J., Mackinnon, N. J., & Kovacs, G. (2011). The association between a prolonged stay in the emergency department and adverse events in older patients admitted to hospital: A retrospective cohort study. *BMJ Quality & Safety, 20*(7), 564–569. https://doi.org/10.1136/bmjqs.2009.034926

Adler, L., Yi, D., Li, M., McBroom, B., Hauck, L., Sammer, C., Jones, C., Shaw, T., & Classen, D. (2018). Impact of Inpatient Harms on Hospital Finances and Patient Clinical Outcomes. *Journal of Patient Safety, 14*(2), 67–73. https://doi.org/10.1097/PTS.0000000000000171

Alameda, C., & Suárez, C. (2009). Clinical outcomes in medical outliers admitted to hospital with heart failure. *European Journal of Internal Medicine, 20*(8), 764–767. https://doi.org/10.1016/j.ejim.2009.09.010

Altman, D. G., & Royston, P. (2006). The cost of dichotomising continuous variables. *BMJ, 332*(7549), 1080. https://doi.org/10.1136/bmj.332.7549.1080

Austin, P. C., Rothwell, D. M., & Tu, J. V. (2002). A comparison of statistical modeling strategies for analyzing length of stay after CABG surgery. Health Services and Outcomes Research Methodology, 3(2), 107–133. https://doi.org/10.1023/A:1024260023851

Austin, P. C., Harrell, F. E., Jr., & Steyerberg, E. W. (2021). Predictive performance of machine and statistical learning methods: Impact of data-generating processes on external validity in the "large N, small p" setting. *Statistical Methods in Medical Research, 30*(6), 1465–1483. https://doi.org/10.1177/09622802211002867

Bai, A. D., Srivastava, S., Tomlinson, G. A., Smith, C. A., Bell, C. M., & Gill, S. S. (2018). Mortality of hospitalised internal medicine patients bedspaced to non-internal medicine inpatient units: Retrospective cohort study. *BMJ Quality & Safety, 27*(1), 11–20. https://doi.org/10.1136/bmjqs-2017-006925

Bayley, M. D., Schwartz, J. S., Shofer, F. S., Weiner, M., Sites, F. D., Traber, K. B., & Hollander, J. E. (2005). The financial burden of emergency department congestion and hospital crowding for chest pain patients awaiting admission. *Annals of Emergency Medicine, 45*(2), 110–117. https://doi.org/10.1016/j.annemergmed.2004.09.010

Bekker, R., Koole, G., & Roubos, D. (2017). Flexible bed allocations for hospital wards. *Health Care Management Science, 20*(4), 453–466. https://doi.org/10.1007/s10729-016-9364-4

Bogler, O., Liu, J., Cadesky, B., & Bell, C. M. (2021). Quality of care and outcomes in internal medicine patients bedspaced to noninternal medicine units. *Medicine, 100*(18), Article e25737. https://doi.org/10.1097/MD.0000000000025737

Bueno, H., Ross, J. S., Wang, Y., Chen, J., Vidán, M. T., Normand, S. L., Curtis, J. P., Drye, E. E., Lichtman, J. H., Keenan, P. S., Kosiborod, M., & Krumholz, H. M. (2010). Trends in length of stay and short-term outcomes among Medicare patients hospitalized for heart failure, 1993–2006. *JAMA, 303*(21), 2141–2147. https://doi.org/10.1001/jama.2010.748

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). Association for Computing Machinery. https://doi.org/10.1145/2939672.2939785

Cheung, J., West, S., & Boughton, M. (2020). The frontline nurse's experience of nursing outlier patients. *International Journal of Environmental Research and Public Health, 17*(14), Article 5232. https://doi.org/10.3390/ijerph17145232

Christodoulou, E., Ma, J., Collins, G. S., Steyerberg, E. W., Verbakel, J. Y., & Van Calster, B. (2019). A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *Journal of Clinical Epidemiology, 110*, 12–22. https://doi.org/10.1016/j.jclinepi.2019.02.004

Couronné, R., Probst, P., & Boulesteix, A.-L. (2018). Random forest versus logistic regression: A large-scale benchmark experiment. *BMC Bioinformatics, 19*, Article 270. https://doi.org/10.1186/s12859-018-2264-5

Dutch Hospital Data. (2026a). *Naslagwerk LBZ-indicatoren* (Versie 2.1). https://www.dhd.nl/assets/uploads/DHD-Naslagwerk-Rapportage-LBZ-indicatoren-v2.1.pdf

Dutch Hospital Data. (2026b). *Opnameduur-indicator: Naslagwerk* (Versie 1.0). https://www.dhd.nl/assets/uploads/20260420-Naslagwerk-Opnameduur-indicator.pdf

Gómez-Rosado, J.-C., Li, Y.-H., Valdés-Hernández, J., Oliva-Mompeán, F., & Capitán-Morales, L.-C. (2019). Analysis of frequency, type of complications and economic costs of outlying patients in general and digestive surgery. *Cirugía Española (English Edition), 97*(5), 282–288. https://doi.org/10.1016/j.cireng.2019.05.004

Goulding, L., Adamson, J., Watt, I., & Wright, J. (2012). Patient safety in patients who occupy beds on clinically inappropriate wards: A qualitative interview study with NHS staff. *BMJ Quality & Safety, 21*(3), 218–224. https://doi.org/10.1136/bmjqs-2011-000280

Goulding, L., Adamson, J., Watt, I., & Wright, J. (2015). Lost in hospital: A qualitative interview study that explores the perceptions of NHS inpatients who spent time on clinically inappropriate hospital wards. *Health Expectations, 18*(5), 982–994. https://doi.org/10.1111/hex.12071

Gupta, A., Stead, T. S., & Ganti, L. (2024). Determining a meaningful R-squared value in clinical medicine. *Academic Medicine & Surgery*. https://doi.org/10.62186/001c.125154

Han, T. S., Murray, P., Robin, J., Wilkinson, P., Fluck, D., & Fry, C. H. (2022). Evaluation of the association of length of stay in hospital and outcomes. *International Journal for Quality in Health Care, 34*(2), Article mzab160. https://doi.org/10.1093/intqhc/mzab160

Hofstede, S. N., van Bodegom-Vos, L., Kringos, D. S., Steyerberg, E., & Marang-van de Mheen, P. J. (2018). Mortality, readmission and length of stay have different relationships using hospital-level versus patient-level data: An example of the ecological fallacy affecting hospital performance indicators. *BMJ Quality & Safety, 27*(6), 474–483. https://doi.org/10.1136/bmjqs-2017-006776

Hoogervorst-Schilp, J., Langelaan, M., Spreeuwenberg, P., de Bruijne, M. C., & Wagner, C. (2015). Excess length of stay and economic consequences of adverse events in Dutch hospital patients. *BMC Health Services Research, 15*(1), Article 531. https://doi.org/10.1186/s12913-015-1205-5

Hvidtfeldt, U. A., Rasmussen, S., Grønbaek, M., Becker, U., & Tolstrup, J. S. (2010). Influence of smoking and alcohol consumption on admissions and duration of hospitalization. *European Journal of Public Health, 20*(4), 376–382. https://doi.org/10.1093/eurpub/ckp153

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). *An introduction to statistical learning: With applications in R* (2nd ed.). Springer. https://doi.org/10.1007/978-1-0716-1418-1

Kaboli, P. J., Go, J. T., Hockenberry, J., Glasgow, J. M., Johnson, S. R., Rosenthal, G. E., Jones, M. P., & Vaughan-Sarrazin, M. (2012). Associations between reduced hospital length of stay and 30-day readmission rate and mortality: 14-year experience in 129 Veterans Affairs hospitals. *Annals of Internal Medicine, 157*(12), 837–845. https://doi.org/10.7326/0003-4819-157-12-201212180-00003

Karabatsou, D., Tsironi, M., Tsigou, E., Boutzouka, E., Katsoulas, T., & Baltopoulos, G. (2016). Variable cost of ICU care, a micro-costing analysis. *Intensive & Critical Care Nursing, 35*, 66–73. https://doi.org/10.1016/j.iccn.2016.01.001

Kohn, R., Harhay, M. O., Bayes, B., Song, H., Halpern, S. D., Kerlin, M. P., & Greysen, S. R. (2021). Influence of bedspacing on outcomes of hospitalised medicine service patients: A retrospective cohort study. *BMJ Quality & Safety, 30*(2), 116–122. https://doi.org/10.1136/bmjqs-2019-010675

La Regina, M., Guarneri, F., Romano, E., Orlandini, F., Nardi, R., Mazzone, A., Fontanella, A., Campanini, M., Manfellotto, D., Bellandi, T., Gussoni, G., Tartaglia, R., & Squizzato, A. (2019). What quality and safety of care for patients admitted to clinically inappropriate wards: A systematic review. *Journal of General Internal Medicine, 34*(7), 1314–1321. https://doi.org/10.1007/s11606-019-05008-4

Lingsma, H. F., Bottle, A., Middleton, S., Kievit, J., Steyerberg, E. W., & Marang-van de Mheen, P. J. (2018). Evaluation of hospital outcomes: The relation between length-of-stay, readmission, and mortality in a large international administrative database. *BMC Health Services Research, 18*(1), Article 116. https://doi.org/10.1186/s12913-018-2916-1

Liu, J., Griesman, J., Nisenbaum, R., & Bell, C. M. (2014). Quality of care of hospitalized internal medicine patients bedspaced to non-internal medicine inpatient units. *PLOS ONE, 9*(9), Article e106763. https://doi.org/10.1371/journal.pone.0106763

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In I. Guyon, U. von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), *Advances in Neural Information Processing Systems 30* (pp. 4765–4774). Curran Associates.

Mans, S. E. (2025). *Managing outlying patients: Exploring hospital staff's challenges and safety considerations in acute patient placement decisions* [Master's thesis, Delft University of Technology]. TU Delft Repository. https://resolver.tudelft.nl/uuid:fc71f75e-16b2-412a-884b-6130a07125b2

Mendelsohn, E., Honeyford, K., Brittin, A., Mercuri, L., Klaber, R. E., Expert, P., & Costelloe, C. (2023). The impact of atypical intrahospital transfers on patient outcomes: A mixed methods study. *Scientific Reports, 13*, Article 15417. https://doi.org/10.1038/s41598-023-41966-w

Perimal-Lewis, L., Bradley, C., Hakendorf, P. H., Whitehead, C., Heuzenroeder, L., & Crotty, M. (2016). The relationship between in-hospital location and outcomes of care in patients diagnosed with dementia and/or delirium diagnoses: Analysis of patient journey. *BMC Geriatrics, 16*(1), Article 190. https://doi.org/10.1186/s12877-016-0372-5

Perimal-Lewis, L., Li, J. Y., Hakendorf, P. H., Ben-Tovim, D. I., Qin, S., & Thompson, C. H. (2013). Relationship between in-hospital location and outcomes of care in patients of a large general medical service. *Internal Medicine Journal, 43*(6), 712–716. https://doi.org/10.1111/imj.12066

Rajkomar, A., Oren, E., Chen, K., Dai, A. M., Hajaj, N., Hardt, M., Liu, P. J., Liu, X., Marcus, J., Sun, M., Sundberg, P., Yee, H., Zhang, K., Zhang, Y., Flores, G., Duggan, G. E., Irvine, J., Le, Q., Litsch, K., … Dean, J. (2018). Scalable and accurate deep learning with electronic health records. *npj Digital Medicine, 1*, Article 18. https://doi.org/10.1038/s41746-018-0029-1

Riley, R. D., Archer, L., Snell, K. I. E., Ensor, J., Dhiman, P., Martin, G. P., Bonnett, L. J., & Collins, G. S. (2024). Evaluation of clinical prediction models (part 2): How to undertake an external validation study. *BMJ, 384*, Article e074820. https://doi.org/10.1136/bmj-2023-074820

RIVM. (2025). *Choosing a healthy future: Public Health Foresight Study 2024* (RIVM report 2024-0203). National Institute for Public Health and the Environment. https://www.volksgezondheidtoekomstverkenning.nl/vtv-2024

Rojas-García, A., Turner, S., Pizzo, E., Hudson, E., Thomas, J., & Raine, R. (2018). Impact and experiences of delayed discharge: A mixed-studies systematic review. *Health Expectations, 21*(1), 41–56. https://doi.org/10.1111/hex.12619

Saito, T., & Rehmsmeier, M. (2015). The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLOS ONE, 10(3), Article e0118432. https://doi.org/10.1371/journal.pone.0118432

Santamaria, J. D., Tobin, A. E., Anstey, M. H., Smith, R. J., & Reid, D. A. (2014). Do outlier inpatients experience more emergency calls in hospital? An observational cohort study. *Medical Journal of Australia, 200*(1), 45–48. https://doi.org/10.5694/mja12.11680

Sari, A. B., Sheldon, T. A., Cracknell, A., Turnbull, A., Dobson, Y., Grant, C., Gray, W., & Richardson, A. (2007). Extent, nature and consequences of adverse events: Results of a retrospective casenote review in a large NHS hospital. *Quality and Safety in Health Care, 16*(6), 434–439. https://doi.org/10.1136/qshc.2006.021154

Schisterman, E. F., Cole, S. R., & Platt, R. W. (2009). Overadjustment bias and unnecessary adjustment in epidemiologic studies. *Epidemiology, 20*(4), 488–495. https://doi.org/10.1097/EDE.0b013e3181a819a1

Schober, P., Boer, C., & Schwarte, L. A. (2018). Correlation coefficients: Appropriate use and interpretation. *Anesthesia & Analgesia, 126*(5), 1763–1768. https://doi.org/10.1213/ANE.0000000000002864

Serafini, F., Fantin, G., Brugiolo, R., Lamanna, O., Aprile, A., & Presotto, F. (2015). Outlier admissions of medical patients: Prognostic implications of outlying patients. The experience of the Hospital of Mestre. *Italian Journal of Medicine, 9*(3), 299–302. https://doi.org/10.4081/itjm.2015.528

Shmueli, G. (2010). To explain or to predict? *Statistical Science, 25*(3), 289–310. https://doi.org/10.1214/10-STS330

Song, H., Tucker, A. L., Graue, R., Moravick, S., & Yang, J. J. (2020). Capacity pooling in hospitals: The hidden consequences of off-service placement. Management Science, 66(9), 3825–3842. https://doi.org/10.1287/mnsc.2019.3395

Statistics Netherlands. (2025). *HSMR 2024 methodological report*. https://www.cbs.nl/en-gb/longread/diversen/2025/hsmr-2024-methodological-report

Steyerberg, E. W., & Vergouwe, Y. (2014). Towards better clinical prediction models: Seven steps for development and an ABCD for validation. *European Heart Journal, 35*(29), 1925–1931. https://doi.org/10.1093/eurheartj/ehu207

Steyerberg, E. W., Vickers, A. J., Cook, N. R., Gerds, T., Gonen, M., Obuchowski, N., Pencina, M. J., & Kattan, M. W. (2010). Assessing the performance of prediction models: A framework for traditional and novel measures. *Epidemiology, 21*(1), 128–138. https://doi.org/10.1097/EDE.0b013e3181c30fb2

Stone, K., Zwiggelaar, R., Jones, P., & Mac Parthaláin, N. (2022). A systematic review of the prediction of hospital length of stay: Towards a unified framework. *PLOS Digital Health, 1*(4), Article e0000017. https://doi.org/10.1371/journal.pdig.0000017

Stowell, A., Claret, P. G., Sebbane, M., Bobbia, X., Boyard, C., Genre Grandpierre, R., Moreau, A., & de La Coussaye, J. E. (2013). Hospital out-lying through lack of beds and its impact on care and patient outcome. *Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine, 21*, Article 17. https://doi.org/10.1186/1757-7241-21-17

Stylianou, N., Fackrell, R., & Vasilakis, C. (2017). Are medical outliers associated with worse patient outcomes? A retrospective study within a regional NHS hospital using routine data. *BMJ Open, 7*(5), Article e015676. https://doi.org/10.1136/bmjopen-2016-015676

Thomas, J. W., Guire, K. E., & Horvat, G. G. (1997). Is patient length of stay related to quality of care? *Hospital & Health Services Administration, 42*(4), 489–507.

Van Calster, B., McLernon, D. J., van Smeden, M., Wynants, L., & Steyerberg, E. W. (2019). Calibration: The Achilles heel of predictive analytics. *BMC Medicine, 17*, Article 230. https://doi.org/10.1186/s12916-019-1466-7

Van der Ploeg, T., Austin, P. C., & Steyerberg, E. W. (2014). Modern modelling techniques are data hungry: A simulation study for predicting dichotomous endpoints. *BMC Medical Research Methodology, 14*, Article 137. https://doi.org/10.1186/1471-2288-14-137

van Walraven, C., & Austin, P. (2012). Administrative database research has unique characteristics that can risk biased results. Journal of clinical epidemiology, 65(2), 126–131. https://doi.org/10.1016/j.jclinepi.2011.08.002

Verburg, I. W. M., de Keizer, N. F., de Jonge, E., & Peek, N. (2014). Comparison of regression methods for modeling intensive care length of stay. *PLOS ONE, 9*(10), Article e109684. https://doi.org/10.1371/journal.pone.0109684

Vloet, M. J. (2025). *Integrating safety in hospital capacity management: Predicting the impact of patient exposure to overflow beds on patient mortality* [Master's thesis, Delft University of Technology]. TU Delft Repository. https://resolver.tudelft.nl/uuid:b5b74b1d-4173-4608-9da2-9ffba9a4d00b

Wetenschappelijke Raad voor het Regeringsbeleid. (2021). *Kiezen voor houdbare zorg: Mensen, middelen en maatschappelijk draagvlak* (WRR-Rapport 104). WRR. https://www.wrr.nl/publicaties/rapporten/2021/09/15/kiezen-voor-houdbare-zorg

Zannella, V. E., Jung, H. Y., Fralick, M., Lapointe-Shaw, L., Liu, J. J., Weinerman, A., Kwan, J., Tang, T., Rawal, S., MacMillan, T. E., Bai, A. D., Gill, S., Shi, J., Bell, C. M., Razak, F., & Verma, A. A. (2022). Bedspacing and clinical outcomes in general internal medicine: A retrospective, multicenter cohort study. *Journal of Hospital Medicine, 17*(1), 3–10. https://doi.org/10.1002/jhm.2734

A

Description of dimensions

Table A.1 shows the exposure categories for each of the four dimensions: general exposure, duration, moment and place. Each row is one category, and the right column describes which admissions it contains.

Dimension	Category	Description
General exposure	Exposed	At least one overflow episode at any point during the admission
Duration of exposure (absolute)	0–8 h, 8–16 h, 16–24 h, 24–32 h, 32–40 h, 40–48 h	Total overflow time across all episodes, in 8-hour intervals during the first 48 hours
	2–3 d, 3–4 d, 4–5 d, 5–6 d, 6–7 d	1-day intervals between 2 and 7 days
	> 7 d	More than seven days of overflow time
Duration of exposure (relative)	0–10%, 10–20%, ... , 90–100%	Overflow time as a share of the total length of stay, in deciles
Moment of exposure	1st moment and onward	First overflow episode in the first segment of the admission, including directly after a flow-through ward (AMU, EHH/CCU, Day Care Unit, ED)
	2nd moment and onward	First overflow episode in a later segment, after a stay on the appropriate ward
	Last moment	Overflow only in the final segment before discharge
Place of exposure	1st preferred group of alternative wards	Ward in the first tier of the applicable flowchart, where care is most similar to the appropriate ward
	2nd preferred group of alternative wards	Ward in the second group
	Last preferred group of alternative wards	Last resort in every flowchart (ICU)
	Ward not in flowchart	Ward that does not appear in any tier of the applicable flowchart
	No applicable flowchart	No flowchart exists for the patient

Table A.1: Description of the dimensions used in phase 1

B

Charlson's comorbidity groups

Table B.1 lists the 17 Charlson comorbidity groups used to describe a patient's comorbidities, taken from the HSMR methodology (CBS, 2025). A group counts as present when it is recorded as a secondary diagnosis during the admission, and one admission can have several groups at once.

No.	Charlson comorbidity groups
1	Acute myocardial infarction
2	Heart failure
3	Peripheral vascular disease
4	Cerebrovascular disease
5	Dementia
6	Chronic pulmonary disease
7	Connective tissue disease
8	Peptic ulcer disease
9	Liver disease
10	Diabetes
11	Diabetes complications
12	Paraplegia and other paralysis
13	Kidney disease
14	Cancer
15	HIV
16	Metastases
17	Severe liver disease

Table B.1: Charlson comorbidity groups where the number is linked to a diagnosis

C

Metrics for model performance and evaluation

Nagelkerke's pseudo R2

Nagelkerke's pseudo-R^2 expresses how much better a model fits than a null model with only an intercept (Steyerberg et al., 2010). It builds on the Cox & Snell pseudo-R^2, whose maximum lies below 1, and rescales it to range from 0 to 1, where 1 indicates a perfect fit. L_0 is the likelihood of the null model, L_M that of the model with predictors, and n the number of admissions.

$$R^2_{Cox\&Snell} = 1 - (\frac{L_0}{L_M})^{\frac{2}{n}}, \qquad R^2_{Nagelkerke} = \frac{1 - (\frac{L_0}{L_M})^{\frac{2}{n}}}{1 - L_0^{\frac{2}{n}}}$$

Brier Score

The Brier score is the mean squared difference between the predicted probability f and the observed outcome o (0 or 1) over N admissions. Unlike accuracy, it also accounts for how confident a prediction is. It ranges from 0 (perfect) to 1 (worst).

$$Brier\ score = \frac{1}{N}\sum_{i=1}^{N}(f_i - o_i)^2$$

The Brier score depends on the prevalence p of the outcome: a model that assigns p to every admission already scores p(1 − p). In the test set, this is about 0,21 for L-LOS (acute admissions in 2025, prevalence 30,4%) and 0,034 for mortality (3,5%). The scaled Brier score therefore expresses the Brier score relative to this uninformative model (Steyerberg et al., 2010): Scaled Brier score = 1 − Brier score / p(1 − p)

A value of 0 means no improvement over assigning the prevalence, 1 means perfect predictions, and a negative value means worse than the prevalence. This makes models for outcomes with different prevalences, such as L-LOS and mortality, comparable.

Calibration

Calibration measures whether predicted probabilities match observed frequencies: if a model predicts a risk of 25% for a group of admissions, about 25% of them should have the outcome. It is summarised with the calibration intercept and slope (Van Calster et al., 2019), each estimated with a logistic regression of the observed outcome on the log-odds of the predicted probability. For the intercept, the slope is fixed at 1, so that it shows whether predictions are on average too high (negative intercept) or too low (positive intercept), with 0 being ideal. The slope shows whether predictions are too extreme (below 1, typical of overfitting) or too moderate (above 1), with 1 being ideal.

Discrimination (AUC)

The AUC measures how well a model separates admissions with and without the outcome. It is the area under the ROC curve, which plots the true positive rate against the false positive rate across all possible decision thresholds. It can be read as the probability that a randomly chosen admission with the outcome receives a higher predicted risk than one without. An AUC of 1 means perfect separation, and 0,5 means the model does no better than chance.

D

Stratified L-LOS models

Table D.1 summarises the 14 cluster-specific logistic regression models used to calculate the expected probability of L-LOS (Section 3.3). For each main cluster it shows the number of admissions, the number flagged as L-LOS, the number of predictors and Nagelkerke's pseudo-R^2. The covariates are listed in Appendix G.

Nagelkerke's pseudo-R^2 ranged from 0,11 to 0,46 (mean 0,26). The best fit was found for diseases of the genitourinary system (0,46), mental disorders (0,38) and neoplasms (0,37). The weakest fit was found for Symptoms, signs and abnormal findings, and factors influencing health status (0,11) and for diseases of the circulatory system (0,14).

Model of main cluster	Number of admissions	Frequency of L-LOS	Number of predictors in model	Nagelkerke's pseudo-R^2
Diseases of the circulatory system	11.178	2.794	49	0,14
Diseases of the digestive system	8.958	2.239	47	0,17
Neoplasms	7.932	1.983	44	0,37
Injury and poisoning	7.269	1.817	46	0,24
Diseases of the respiratory system	6.640	1.660	45	0,25
Diseases of the genitourinary system	6.351	1.588	43	0,46
Symptoms, signs and abnormal findings, and factors influencing health status	2.450	613	40	0,11
Diseases of the nervous system and sense organs	2.350	588	43	0,29
Endocrine, nutritional and metabolic diseases, and immunity disorders	1.739	435	40	0,21
Diseases of the musculoskeletal system and connective tissue	1.697	424	39	0,34
Diseases of the skin and subcutaneous tissue	1.500	375	37	0,24
Diseases of the blood and blood-forming organs	792	198	31	0,26
Mental disorders	690	173	29	0,38
Infectious and parasitic diseases	644	161	35	0,17

Table D.1: Number of admissions, number of L-LOS admissions, number of predictors and Nagelkerke's pseudo-R^2 per main cluster.

E

Model hyperparameter settings

The hyperparameters that determine the complexity of each machine learning model were tuned with a randomised search within the training set. Each candidate combination was fitted on admissions from 2022 and 2023 and compared on the Brier score in 2024. Up to 30 combinations were evaluated per model and outcome; for the decision tree and the random forest, the search space contained 30 and 24 combinations, so all were evaluated. Each model was then refitted with its selected values on the complete training set (2022–2024) and evaluated once on the 2025 test set. Table E.1 shows the search space and the selected values. Hyperparameters not listed were kept at their default values.

Model	Hyperparameter	Search space	Mortality	L-LOS
Decision Tree	min_samples_leaf	20, 50, 100, 200, 500	100	100
	max_depth	3, 4, 6, 8, 10, 12	12	10
Random Forest	min_samples_leaf	10, 25, 50, 100	10	10
	max_features	0,3; 0,5	0,3	0,3
	max_depth	10, 20, unlimited	10	10
XGBoost	subsample	0,7; 0,85; 1,0	0,7	0,7
	n_estimators	100, 200, 400, 800	100	200
	min_child_weight	1, 5, 10, 50	5	1
	max_depth	2, 3, 4, 6	3	2
	learning_rate	0,01; 0,03; 0,05; 0,1	0,1	0,1

Table E.1: Search space and selected hyperparameter values per model, for mortality and L-LOS.

F

Data preparation sensitivity analysis L-LOS95

This appendix repeats the data preparation steps of Section 3.3 for the stricter definition of L-LOS used in the sensitivity analysis (Section 4.5). Only the threshold differs: an admission is flagged as L-LOS95 when its length of stay exceeds the 95th percentile of its main cluster. All other steps are identical to the main analysis.

Diagnosis-specific thresholds

For each of the 14 main clusters, the 95th percentile of length of stay was calculated over all admissions in that cluster. Table F.1 shows, per cluster, the number of admissions, the L-LOS95 threshold in days, and the number and percentage of admissions flagged as L-LOS95. By construction, about 5% of the admissions in each cluster are flagged.

Main cluster	Number of admissions	Average LOS (days)	Median LOS (days)	Cut-off value for L-LOS95 (days)
Diseases of the circulatory system	11.178	5,26	3,48	15,57
Diseases of the digestive system	8.958	4,28	2,11	14,57
Neoplasms	7.932	4,18	2,17	14,95
Injury and poisoning	7.269	4,73	2,86	14,95
Diseases of the respiratory system	6.640	4,91	3,30	14,48
Diseases of the genitourinary system	6.351	3,53	1,84	11,23
Symptoms, signs and abnormal findings, and factors influencing health status	2.450	3,19	1,88	9,90
Diseases of the nervous system and sense organs	2.350	4,09	1,90	15,08
Endocrine, nutritional and metabolic diseases, and immunity disorders	1.739	5,12	3,05	16,04
Diseases of the musculoskeletal system and connective tissue	1.697	5,65	2,99	19,46
Diseases of the skin and subcutaneous tissue	1.500	5,26	3,71	15,29
Diseases of the blood and blood-forming organs	792	3,19	1,43	10,96
Mental disorders	690	5,91	2,76	19,78
Infectious and parasitic diseases	644	9,27	6,24	27,42

Table F.1: Number of admissions, average LOS, Median LOS and the L-LOS95 cut-off value

Expected probability of L-LOS95

A new expected probability was estimated for every admission with a separate logistic regression model per main cluster, using the same covariates, reference categories and exclusion rules for

sparse predictors as the main case-mix models (Section 3.3; Appendix G). Because L-LOS95 is a rarer outcome, fewer admissions combine a characteristic with the outcome, so at least as many predictors fall below the threshold of ten such admissions as in the main analysis, and generally more. Table F.2 shows, per cluster, the number of admissions, the number of L-LOS95 admissions, the number of predictors retained and Nagelkerke's pseudo-R^2.

Model of main cluster	Number of admissions	Frequency of L-LOS	Number of predictors in model	Nagelkerke's pseudo-R^2
Diseases of the circulatory system	11.178	559	41	0,09
Diseases of the digestive system	8.958	448	40	0,13
Neoplasms	7.932	397	36	0,18
Injury and poisoning	7.269	364	39	0,13
Diseases of the respiratory system	6.640	332	41	0,14
Diseases of the genitourinary system	6.351	318	38	0,28
Symptoms, signs and abnormal findings, and factors influencing health status	2.450	123	25	0,16
Diseases of the nervous system and sense organs	2.350	118	22	0,25
Endocrine, nutritional and metabolic diseases, and immunity disorders	1.739	87	18	0,12
Diseases of the musculoskeletal system and connective tissue	1.697	85	19	0,19
Diseases of the skin and subcutaneous tissue	1.500	75	15	0,15
Diseases of the blood and blood-forming organs	792	40	9	0,14
Mental disorders	690	35	6	0,08
Infectious and parasitic diseases	644	33	4	0,05

Table F.2: Number of admissions, number of L-LOS admissions, number of predictors and Nagelkerke's pseudo-R^2 per main cluster.

G

Recoding case-mix covariates

This appendix describes the case-mix covariates used to estimate the expected probability of L-LOS (Section 3.3). Table H.1 describes what each covariate represents, and Table H.2 shows how each covariate was coded, including the reference category used in the models.

Variable	Description
Age	The patient's age at admission, grouped into age bands
Sex	The patient's sex
Socio-economic status	Socio-economic class based on the patient's postcode, from 1 (lowest) to 5 (highest), with a separate category when unknown
Severity of the main diagnosis	The severity class of the main diagnosis, based on the historical mortality rate of that diagnosis, as used in the HSMR model
Urgency of admission	Whether the admission was acute (care needed within 24 hours) or non-acute
Comorbidities	Chronic conditions recorded as secondary diagnoses, grouped into the 17 Charlson groups (Appendix B)
Source of admission	Where the patient came from before admission: home, another hospital, or a nursing home or other institution
Year of registration	The year of registration, to account for differences in the baseline risk of L-LOS between years
Month of admission	The month of admission

Table G.1. Description of the case-mix covariates.

All covariates were coded as dummy variables. For a covariate with n categories, one category served as the reference and the remaining $n - 1$ categories were each represented by a binary indicator.

Variable	Categories	Reference category
Age (years)	20–29, 30–39, 40–49, 50–59, 60–69, 70–79, 80–89, 90–99, 100+	< 20
Sex	Female	Male
Socio-economic status	Below average, Average, Above average, Highest, Unknown	Lowest
Comorbidities	17 Charlson comorbidity groups, each coded as present or absent (Appendix B)	Absent
Severity of the main diagnosis	[0,01; 0,02), [0,02; 0,05), [0,05; 0,1), [0,1; 0,2), [0,2; 0,3), [0,3; 0,4), [0,4; 1], Other	[0; 0,01)
Urgency of admission	Urgent	Non-urgent
Source of admission	Other hospital, Nursing home or other institution	Home
Month of admission	February to December	January
Year of registration	2023, 2024, 2025	2022

Table G.2. Categories and reference categories of the case-mix covariates.

In the prediction models of Phase 3, age was included as a continuous variable and the other characteristics were one-hot encoded as in Table G.2. Charlson group 15 (HIV) was left out, as it occurred in only five admissions in the training set, none with the outcome, so that its coefficient could not be estimated. Together with general exposure, this gives 45 predictors.

Tree-based models do not require one-hot encoding. An additional XGBoost model was therefore fitted in which socio-economic status, month of admission, source of admission and severity of the main diagnosis were each supplied as a single categorical variable, using XGBoost's native support for categorical features.

H

Sensitivity, FPR & Precision of LR predictions

Figure F.1 shows, for the logistic regression models in the 2025 test set, what happens when a given share of admissions with the highest predicted risk is flagged as high risk. The x-axis shows this share, from 1% to 50% of admissions.

The left panel shows the sensitivity: the share of admissions with the outcome that are flagged, in other words how many of the patients who die, or who have an L-LOS, the model identifies. The middle panel shows the false positive rate: the share of admissions without the outcome that are nevertheless flagged, in other words how many patients receive a high-risk label without needing it. The right panel shows the precision: the share of flagged admissions that actually have the outcome, in other words how often a high-risk label is correct.

The dashed lines show the result of flagging admissions at random. For sensitivity and the false positive rate, this equals the share flagged and is the same for both outcomes; for precision, it equals the prevalence of each outcome. The further a model's line lies from the dashed line, the better the model performs than chance. The dotted vertical line marks the point at which 10% of admissions are flagged.

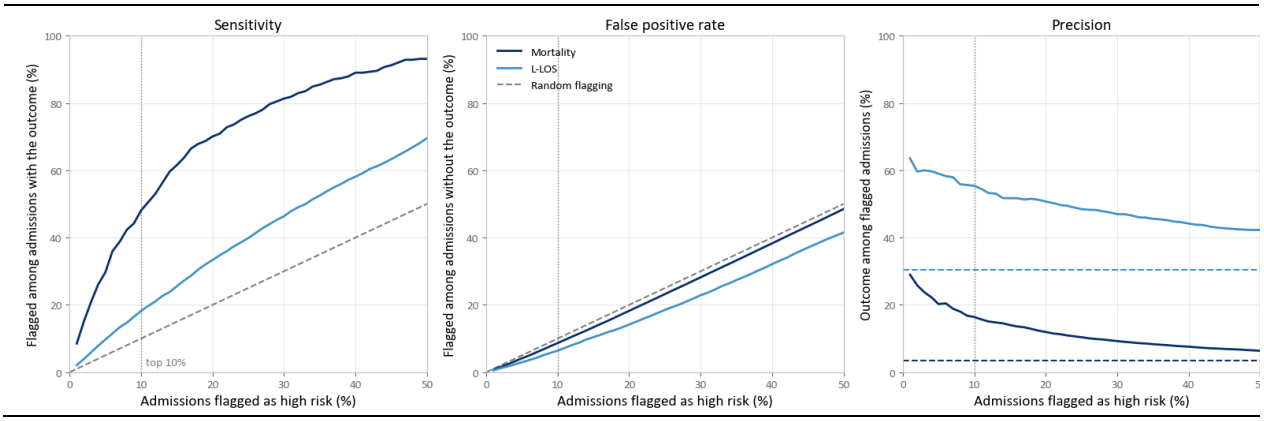

Figure F.1: Sensitivity, false positive rate and precision of the logistic regression models in the 2025 test set, by the share of admissions with the highest predicted risk that is flagged as high risk. Dashed lines show random flagging; the dotted line marks 10% flagged.