UNDERSTANDING SYSTEM FAILURE IN HEALTH CARE: A MENTAL MODEL FOR DEMAND MANAGEMENT

D. Hartmann[1,3]*, J. Bicheno[2], B. Emwanu[3] & T.S. Hattingh[4]

ARTICLE INFO

Article details
Submitted by authors 25 Mar 2020
Accepted for publication 12 Jul 2021
Available online 31 Aug 2021

Contact details
* Corresponding author
dieter.hartmann@gmail.com

Author affiliations
1 Hazaa! Consulting, Johannesburg, South Africa

2 Business School, University of Buckingham, Buckingham, United Kingdom

3 School of Mechanical, Industrial and Aeronautical Engineering, University of the Witwatersrand, Johannesburg, South Africa

4 Department of Industrial Engineering, The North West University, Potchefstroom, South Africa

ORCID® identifiers
D. Hartmann
http://oricd.org/0000-0001-9641-0095

J. Bicheno
http://oricd.org/0000-0001-8555-0439

B. Emwanu
http://oricd.org/0000-0001-9906-7245

T.S. Hattingh
http://oricd.org/0000-0001-5930-2546

DOI
http://dx.doi.org/10.7166/32-2-2344

ABSTRACT

The load on health systems caused by systemic overburden leads to heightened costs, longer waiting times, a reduced quality of care, and associated problems. This may be caused by 'failure demand'; however, its definition is inadequate for a complex hierarchical system. Although accounting for a significant proportion of load in other industries, the academic assessment of *failure demand* in health care remains limited. We present a novel way of identifying repeat consumption, which we loosely equate with failure demand. We present a framework that can be used to identify 'system failure', the trigger for later repeat consumption. This provides new insight into understanding whether common events represent system failure. A diagnostic framework was developed from observations, the literature, and brainstorming. Commonly observed exit scenarios in health care were tested against the framework to create a system-failure list. The framework and the categorisation table were shared with eight international Lean health-care experts. Following feedback, the framework and categorisations were fine-tuned and consensus was achieved via member-checking. Identifying and managing *failure demand* for these settings can lead to a reduced system load, thus reducing costs and increasing system efficiency and quality.

OPSOMMING

Die las op gesondheidstelsels as gevolg van sistemiese oorlading lei tot hoër kostes, langer wagtye, swakker versorgingsgehalte en ander verwante probleme. Dit mag dalk veroorsaak word deur 'falings-aanvraag', maar dié konsep se definisie is onvoldoende vir ingewikkelde hiërargiese stelsels. Daar is beperkte literatuur beskikbaar oor falings-aanvraag in die gesondheidsektor. Hierdie artikel bied 'n nuwe manier om herhaalde verbruik te identifiseer – ons assosieer dit losweg met falings-aanvraag. 'n Raamwerk wat gebruik kan word om stelsel mislukking, die afsetter vir herhaalde verbruik, te identifiseer, was ontwikkel. Dit verskaf nuwe insig om te verstaan of algemene gebeure stelsel mislukking verteenwoordig. 'n Diagnostiese raamwerk is ontwikkel aan die hand van waarnemings, die literatuur en dinkskrums. Algemene uitgangscenario's is getoets deur middel van die raamwerk om 'n lys van stelsel mislukkings te genereer. Die raamwerk die kategoriseringstabel is gedeel met agt internasionale lenige gesondheidsorg kenners. Hul terugvoer is gebruik om die raamwerk en die kategorisering te verfyn totdat konsensus bereik is. Die identifiseer en bestuur van falings-aanvraag vir hierdie scenario's mag lei tot verminderde stelsel lading en dus onkostes verminder en sodoende doeltreffendheid en gehalte verbeter.

1 INTRODUCTION

1.1 Study context

The genesis of this study lies in trying to understand the phenomenon of failure demand, how it presents in health systems, and the impact that this has on service delivery.

The first publication in a series of three publications (shown in Figure 1) identified five demand modalities in health systems, of which failure demand was one. Recognising that there are gaps in defining and identifying failure demand in more complex hierarchical organisations, a greater depth of investigation was required.

This paper forms the second part of a larger study that was conducted with the intention to understand certain aspects of demand in health systems. In this paper we present a framework that can be used to assess system failure that could lead to failure demand. Common events in the provision of health care were tested against this algorithm to validate it. We summarise our findings with a list of events that could be root causes of failure demand. The totality — framework and findings — has been validated by experts.

In the third publication, one of the categories responsible for failure demand identified in this paper — poor supply chain management — was explored in greater depth in an empirical study of a national pharmaceutical supply chain in a developing country.

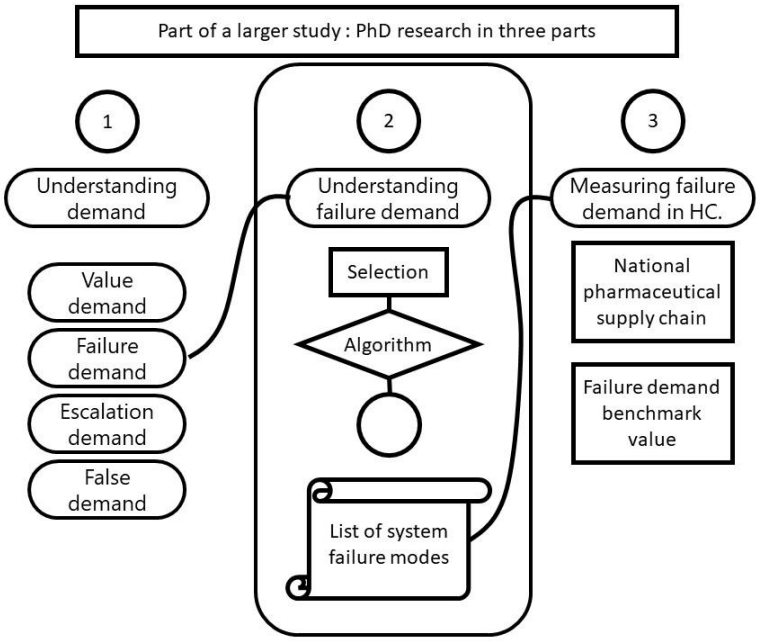

Figure 1: Context of this paper

1.2 Seeing failure demand everywhere

The Institutional Review Board (also referred to as the ethics or review board or committee[1]) at the authors' home institution recently circulated the following memo:

We are aware of the concern of applicants about the length of time to receive responses to their applications submitted through the Research Office. The reason is that the committee and office are overwhelmed by workload. A recent audit shows that 891 new applications were submitted for review in 2016. Only 107 (12%) were approved at the first evaluation, 784 (88%) had to be resubmitted — this means that the new application workload was 891 + 784 = 1675 over and above other work. In 2008 217/586 (37%) were approved at first evaluation after which there has been a steady deterioration. We apologize for the delays that are influenced by the workload.

This memo reveals a fascinating phenomenon that raises workload, increases demand on limited resources, and so increases waiting time and affects the quality of work. '*Failure demand' is* the customer interaction that occurs more than once because a previous interaction with the system that provides the service was unsuccessful.

[1] Nomenclature differs across the world; the North American standard form is 'Institutional Review Board', while the British (and broadly Commonwealth) naming is some variation that contains the word 'Ethics' [1]. Their functions are indistinguishable, and are governed by the declaration of Helsinki [2].

1.3 Failure demand and resolution

Seddon [3] proposes two forms of demand: *value demand*, which (in the above example) can be thought of as the initially successful 12 per cent of ethics applications, and *failure demand* — the remainder of the cases, which need the work elements to be repeated with subsequent reprocessing. He introduced the idea [4] as an evolution of his earlier thinking on 'demand that we do not want' [3]. The definition of failure demand is:

demand caused by a failure to do something or do something right for the customer. [4 Pg. 27]

The remainder of this section unpacks Seddon's wording to visualise failure demand and to identify its nuances. Consider Figure 2, which shows the current thinking.

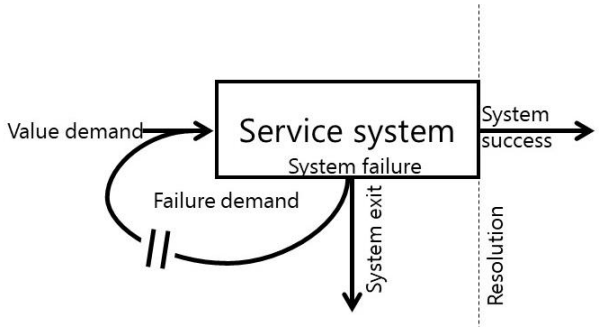

Figure 2: Generic system showing system failure, failure demand, and value demand

The system receives a certain *demand*, which consists of *value demand* and *failure demand*. This *total demand* is addressed by the system. A certain proportion of work passes through the system unimpeded, and that we refer to as *system success*. This is the normal case as proposed as the 'conservation of material — law' by Hopp and Spearman [5], and is the flow-through system that forms the basis for Little's law [6].

What is resolution?
The notion of resolution is complex. Whereas in manufacturing it is clear when the product is delivered at its nominal value [7], in service systems the customer decides whether the nominal value of the service has been achieved. This idea is implied in the service-directed work by Womack and Jones [8] and explicitly stated by Seddon [4]. Whether a service has been delivered to completion is therefore dynamic, and no static value defining an acceptable 'product' exists [4]. Attempting to standardise service offerings escalates cost without generally achieving greater levels of customer satisfaction [9]. This is further complicated in health care, as the objective of the work often differs. Lillrank *et al*. differentiate between the purpose of treatment being either 'cure', where the target outcome is healing, while 'care' by contrast does not target a healed state, but rather focuses on maintenance [10]. For this reason, patients may need to return repeatedly for treatment without its being thought of as a system failure, if resolution has been achieved. Analytically this means that failure demand cannot be measured by simply counting every instance of a patient returning (for the same reason).

Some work elements do not successfully progress through the system for a variety of reasons, and at various stages of the process. We call the trigger for this unsuccessful process '*system failure*'. Once *system failure* has occurred, the work element may depart the system forever — which we refer to as 'system exit' — or the work element may return to the demand queue at a later stage after a time delay. The work elements that undergo *system failure*, yet return, represent *failure demand*. This means that the total demand queue is elongated by *failure demand* elements, which place the same burden on the system as they did previously.
hile the classic definition of *failure demand* may suffice in some industries, we view health care as an environment in which Seddon's definition requires refinement.

In this paper we will discuss *failure demand* broadly. We will identify where *failure demand* comes from, what causes it, and what effect it has on overall system performance. We will describe the moment when the system is unable to serve the customer, and introduce the terminology of *system failure* to describe this moment. We introduce the drivers of *system failure*, and explore how this moment of *system failure* relates to the occurrence of *failure demand*. We will show that, although *system failure* is always the trigger for *failure demand*, not all instances of *system failure* result in *failure demand*. We will then show

how the current definitions of *failure demand* are not adequate to describe health care systems, given their higher levels of complexity, their complicated hierarchical structure, and the characteristics of health care, including independent patient behaviour. We will propose an algorithm that will be used to assess system exits prior to resolution. This algorithm will be used to establish whether a particular modality of system failure leads to failure demand, and tabulate our findings.

Examples of failure demand

Seddon [4] primarily focused the introduction of his idea of *failure demand* in the call centre industry. An example can be found in the case of a UK bank, in which it appeared that demand was increasing because the call volumes were increasing; so the bank in turn increased the number of call centres to cater for this demand. However, the reality was that call volumes had increased owing to *failure demand* in addition to *value demand*. *Failure demand* in this case was found to account for 46 per cent of the total demand. The remedy was to manage first-time interactions and, as a result, the number of calls reduced substantially in the long term, leading to the ability to reduce the number of call centres. These studies on call centres have been retested and explained in numerous research papers over the years, with *failure demand* levels ranging from 12 per cent to 80 per cent [11]. Marr and Neely [12] found that most studied organisations spend at least half of their time dealing with *failure demand*. This is also shown in our Institutional Review Board processing ethics applications, where the bulk of time is spent on reprocessing applications, and which has led to a stagnation of actual throughput. This is predictable, using Little's law [6], which shows that more feedback loops into the system lower the throughput rate.

The implications of *failure demand* are that most organisations are either overcapitalised to deal with such unnecessary demand, or that their service delivery is compromised, which Piercy and Rich [13] identified as major opportunities for improvement.

1.4 Introducing *system failure* as a driver of failure demand

Recalling our example of the Institutional Review Board: when an application is rejected, the system has failed. We propose that the applicant forms part of this complex system, and that the agency for the 'failure' may lie there. The trigger that caused the system exit is the point of *system failure*. Applicants may wish never to resubmit their application, in which case the *system failure* was the last act in the interaction. If applicants resubmit their ethics documents, then, by joining the queue and forming a repeated burden on the system, that application becomes *failure demand* — in other words, demand because of a prior failure of the system.

Failure demand is the consequence of an event, and does not exist in isolation from systemic triggers. Understanding these triggers allows for the measurement and understanding of *failure demand*, which may lead to increases in effectiveness and efficiency [14]. Although implied in Seddon's definition, a specific event triggers *failure demand*, and that is the instance of the *system* failing to do something or failing to do something right [4].

These triggers have previously been described using the terminology 'failure' or 'service failure' [12]. We prefer to refer to these triggers as '*system failure' (terminology used, almost in passing, by Piercy and Rich* [13]) to identify events that contribute to unsuccessful work completion. When a user engages a system, but does not get resolution, and departs the system, the system has failed to meet the user's needs. *System failure* becomes *failure demand* only when customers return to the system in need of the same service.

The relationship between *system failure* and *failure demand*, therefore, is one of cause and effect. This paper builds on the traditional model, considering the idiosyncrasies of health care, presenting a framework for identifying *system failure* in health care, which, if the patients return, results in *failure demand*.

1.5 Unique characteristics of health care systems

Health systems in the 21st century are complex organisms. Plsek and Greenhalgh [15] identify many elements that categorise health systems as complex adaptive systems, which have characteristics that are distinct from simple systems. This complexity stems from the nature of health care as a commodity, the interaction between health care and patients, and the nature of disease.

Unlike manufactured products, services cannot buffer completed products for release when demand occurs, meaning that full-service offerings can only be started when they are needed [16]; while Sasser wrote that

service demand tends to be personal and in person [16], meaning that nobody can receive a health care service other than the patient, and usually the care provider cannot have a proxy.

Moreover, because health systems tend to be capacity-led, little attention is paid to demand management [17]. Walley found that this was particularly true in public services. Describing them as 'resource-driven', he concluded that service delivery could be meaningfully improved through the adoption of private-sector-inspired demand-driven strategies [18].

When trying to understand or improve a service system, demand must first be investigated and understood [4, 18-20]. Designing solutions that do not consider demand first are likely to result in incorrect solutions or solutions to incorrect problems [4, 19]. To understand demand, it is important not only to know how much demand a system experiences [16], but also the frequency or distribution of its arrival, what type of demand it is, and in which units the demand is measured.

The nature of disease makes health care an even more complex service environment. Disease can present in many ways, responses to treatment vary, and mistakes are made [21]. Often medical practice, although conservative by nature [22], is experimental, and correct treatment regimens are decided upon through trial and error.

1.6 Understanding how demand is measured

Although demand can be reduced to a simple measure such as 'the number of people in a queue', which is a view that we have taken in earlier work [23], and also in some of the classic Lean health care literature [24, 25], this approach does not recognise that individuals cannot be equated with the load on a system. In later work we made use of time consumed as the measure of demand [26]. But this did not fully address the true load on the system, differentiating between levels of specialisation, resource scarcity, and work complexity. Wagstaff introduces the idea of the 'stock of health capital'. This is the resource that is being depleted when a load is placed on a system [27]. To understand how this 'stock' is structured requires that demand be seen in terms of the underlying complexity of task (which guides the decision of which resources are required to perform the work) and the time that is required to complete the work.

Therefore, the view we take of demand in this paper is a composite that considers the amount of time in which resources are consumed, and the type of resources. Doing so describes the 'stock of health capital' that is being depleted in an interaction and, by extension, describes the load on the system.

1.7 The relationship between demand and capacity

'Utilisation' can be broadly defined as the proportion of capacity being used for economic purposes [28]. This is shown in Equation 1.

$$U = \frac{r}{C} \tag{1}$$

where U is the utilisation of the system, r is the rate of entry of product (which can be seen as the demand or load), and C is the system capacity.

Caution must be applied to avoid the intuitive 'rule' that utilisation must be as high as possible. Designing systems by targeting absolute utilisation creates the risk of their going unmanageably out of control [19]. Low utilisation means that a system is less capable of delivering a service; however, as utilisation nears a hundred per cent (a practical impossibility, as constrained by Hopp an Spearman's laws of utilisation and capacity [5]), the system's ability to respond to demand falls drastically, and queues are elongated uncontrollably [19]. Conscious of the harmful effects of very high utilisation, the NHS has introduced an 85 per cent bed occupancy 'rule' [29] as a way to keep the system in control.

We argue that demand in health care is structured as shown in Equation (2):

$$D_t = D_v + D_e + D_f + D_n \tag{2}$$

where D_t: total demand, D_v: value demand, D_e: escalation demand, D_f: failure demand, D_n: false demand.

The elements (or, as we prefer, 'modalities') of demand shown in Equation (2) suggest a way to understand how the queue in a health system is constituted. We use Seddon's view that demand is an aggregation of value demand and failure demand [4]. However, we include modalities outside of his binary classification.

We include *false demand* (which emerges from the healthy population) and *escalation demand* (the load placed on the system owing to delayed treatment; for example, in the USA, it was found that one in eight cases became more severe because of delayed treatment [30]).

The capacity of a system is its ability to execute a function [28]. This capacity cannot be exceeded, and can only be sustained for transient timeframes, as a multitude of limitations, called 'detractors', reduce this capacity from the base capacity to the process capacity [5, 28].

$$C = C_b - D \tag{3}$$

where C: process capacity, C_b: *base capacity*, D: detractors

Equation (3) [28] shows how detractors reduce the base capacity of a system to its true capacity, which becomes the target, and which is in many cases the arbitrary product of average utilisation, one good shift, and other factors [28]. Detractors may include failures, breakdowns, resource unavailability, and start-up effects, which Bicheno calls 'equipment losses' [31]. These elements can be managed to achieve zero losses [32]. Another category is are so-called 'dispensable-time-losses' [31], which include meaningless meetings, capturing and reporting on data that is never scrutinised, bureaucratic clutter, or what Graeber refers to as "BS jobs" [33][2].

2 OBJECTIVES AND APPROACH

Recognising the impact that *failure demand* has in many industries, we set out to understand and define this phenomenon in health care. To do so, we aim to:

- build a logical framework that can be used to assess events to classify them as either 'system failure' or not;
- use the developed framework to classify commonly occurring events in health care and identify them by modality; and
- test the framework and the classifications with a panel of experts to validate the utility, primarily of the framework and secondarily of the classifications.

2.1 Method

Roy *et al*. say that concepts are incremental and build upon existing knowledge, ideas, observations, and their synthesis [34]. To ensure that the framework presented in this paper is a credible tool, we surveyed a panel of experts to validate the usefulness of the framework and to ensure that it was a comprehensive treatment capable of delivering a valid conclusion about system failure.

To achieve our aims, we reviewed applicable global literature sources to identify the major drivers for patients leaving health services. We augmented these sources with several exploratory studies and general observations.

This study consists of the four major parts shown in Figure 3.

First, a framework was constructed ① making use of literature sources, general observations, and brainstorming [35]. The purpose of this framework was to serve as a logical test of a scenario to evaluate whether or not an event is system failure.

Second, a selection of eighteen common events was compiled, based on observations of failure in health systems from literature sources and our own experience ②. These events were presented to the framework developed above and, in so doing, tested the ability of the framework to conclude correctly whether or not an event was system failure. Based on the application of each of the events to the model, a categorised list of events was presented ③.

[2] Graeber is less restrained in his use of the unabbreviated form, which the curious reader may find in the reference list.

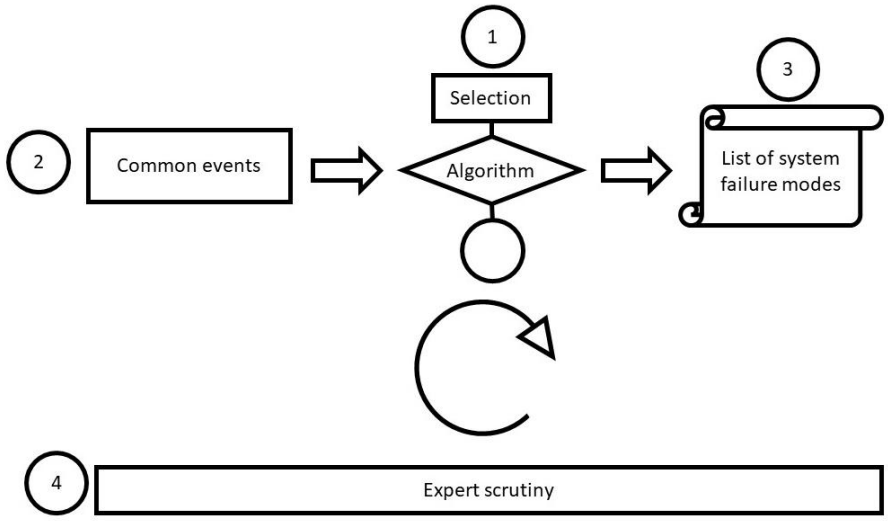

Figure 3: The phases of this study

Model validation
Using Dijkstra *et al.*'s expert validation approach [35], a panel of experts was engaged to validate the framework as well as the findings of this study ④. Ten international experts in Lean healthcare were contacted by email to request their support in validating the model. Two of the experts ignored the request for support, and one declined participation, citing a high COVID-19 workload. One of the experts re-shared the survey with an additional expert, bringing the total response rate to eight.

What was shared with the experts
The experts were able to watch a recorded video that was hosted on a private YouTube channel. The video described all the main elements presented in this paper, with major sections devoted to 'a recap of failure demand', 'hierarchical models', 'the development of the algorithm', 'identifying the activities to be tested against the algorithm', 'demonstration of the algorithm in use', and a 'conclusion of the categorisation of events'. The video was 27 minutes long; however, watching speeds of up to 1.5x remained feasible — and the experts were encouraged to do so. A draft of this paper was also shared on request.

After watching the video, the experts accessed a Google Sheet that automated the collection of their opinions. It had three major sections: the collection of key information about the experts, gathering their opinions and inputs about the algorithm, and their conclusions about the findings of the study.

The outputs shown in this paper reflect two rounds of interactions with the panel of experts, and represent the endpoint of many iterations to arrive at consensus.

Experts' credentials
All the consulted experts were prominent professionals in Lean healthcare. They were sourced from academia (five), private consulting (two), and health system management (one). The experts' self-reported experience in Lean healthcare averaged sixteen years. All of the experts had published extensively in the field, with high-impact journal publications, keynote appearances, and at least five published books between them. Four of the experts held a PhD or were completing one. Two had a Master's degree, and two had undergraduate degrees. Six of the experts had been involved in Lean training, developing bespoke material for academic and private organisations. Six of the experts had led a major Lean project. Although the credentials of the experts were beyond question, it was interesting that most of them were modest when assessing their own expertise, with their average self-reported score of expertise being 3.75 on a five-point Likert scale.

2.2 Expert assessment

The framework presented in this paper is the final product, and has undergone expert validation. Modifications have been incorporated into Figure 4 and, by extension, Table 1.

2.2.1 Changes to the framework

The language of this framework initially used the term 'delinquent act'. Although this was meant to refer to the act, two experts were concerned that the pejorative implications of the word 'delinquent' might create the impression of a 'bad' patient or a 'bad' doctor or nurse. This was not the intention for two reasons: first, that the systems view does not focus on the individual 'fault' [36, 37], but rather tries to assess systemic questions; and second, the wording unwisely implied that the focus lay on 'fault' rather than on 'the act' [4]. As this was by no means the intention of this model, the term was dropped and replaced by 'triggering act'.

All the experts agreed that the framework was useful. In a separate question, they awarded it an average score of 3.75 for being able to be used for other, untested scenarios. This score was lower than we had hoped, and seemed to emerge from a concern that the framework needed revision to be more generalisable. One expert suggested that, to be more generalisable, clinical language should be removed from the formulation. This would allow the framework to be useful not only in clinical settings, but perhaps also in other complex hierarchical systems such as government and the legal system. Clinically specific language was thus deleted and generalised.

The nodes

One of the experts suggested that an emphasis on capacity planning and resource allocation [38] should be split out as an additional node, and not be covered under the 'catch-all system-design' node. Doing so also brings the strategic layer [39] into the algorithm as an equal contributor to system failure.

A previous category, 'support services', was eliminated, as the support services are inside the service system boundary and are covered by existing nodes. Similarly, a previously existing node for incompetence was deleted and merged with the 'delinquent act' — later the 'triggering act' — because there was no meaningful distinction between them.

Two of the more subjective nodes were enriched by creating smaller sub-frameworks. One was the 'triggering act', making use of the traditional failure demand definition [4]; and a further framework was added to assess whether resolution was achieved [4, 8, 40]. This was necessary to adjudicate the difference between what Lillrank *et al.* call 'care' and 'cure' [10]. This should ensure that returning for care should not be interpreted as system failure.

The addition of scenarios for testing

Only one expert identified a common scenario that should be added to the assessment of the model: the patient leaving without being seen. This was included in Section 4.3.5.

3 THE SYSTEM FAILURE FRAMEWORK

The framework is shown in Figure 4. The main framework is shown between the two subordinate frameworks, indicated by ① and ②. These smaller frameworks are used to assist in reducing the subjectivity on two nodes, as indicated. The 'triggering act' is derived from the pure definition for failure demand: not doing something, or doing something wrong [4]. The next node speaks to the complex idea of resolution. This merges the thinking of Womack [8] and Seddon [41], that the completion of a service is to be interpreted from the point of view of the customer, but that, equally, it needs to be measured against a good practice standard in order to determine whether this is appropriate for the type of service required [10].

On the main framework, the next node assesses system design on the operational and strategic levels [4, 39, 42], followed by a node that assesses poor sharing of information, which again could be systemic in nature. The final node assesses whether departures were the result of issues with capacity planning and resource allocation [38].

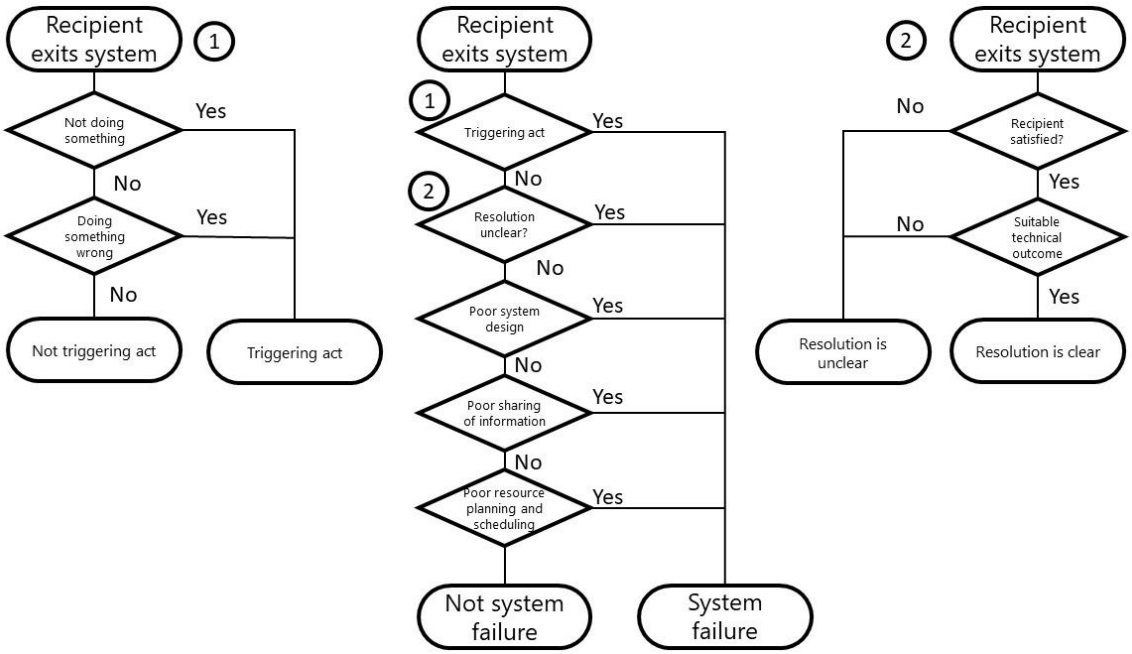

Figure 4: Framework for assessing system failure

4 TESTING THE FRAMEWORK

This section explores the modes of *system failure* in health care, which — if accompanied by returning patients — can be classified as *failure demand*. *System failure* is the root cause of *failure demand*, and it is the manageable element in reducing failure demand.

Sections 4.1 to 4.4 show the four key themes for which clarity is required to define *system failure* in health care. These deal with:

- the hierarchical nature of health systems,
- unsuccessful medicine,
- the errors made by patients,
- the overall operational environment for health provision.

4.1 Hierarchical structure of health care systems

A prominent difference between health and other systems is their complex, intentionally structured hierarchies. Underlying this structure are ideological, political, and economic models in support of the health system [43], ranging from day-to-day management to strategic, system-wide design [39].

Most health systems are structured so that simpler care is provided at lower-cost entities [44], such as in community information programmes [45] and at minor clinics, general practitioner practices, health management organisations (HMOs) [46], and outpatient departments. Higher skill and specialisation coincides with more costly and generally larger facilities such as specialised clinics and hospitals [21]. This protects highly specialised hospitals from being overburdened (Muri) [37], which is wasteful [47]. This structure also reduces the overall system cost, as primary interventions cost less.

To benefit from hierarchical levels, a referral system is used in which patients arriving at a point of care of the incorrect level are referred to the correct facility. Patients who arrive at a level above their need are referred downwards, usually after some diagnostic and administrative work [48]. Similarly, patients who arrive at a more primary point of care are referred upwards through the health system until they arrive at the correct level of care, without unduly burdening more skilled, costly, and restricted higher diagnostic and administrative levels.

These hierarchies are probably necessary to provide health care, as up- and down-referral is a cost-limiting mechanism by which patients arrive at the correct care level. It is plausible that patients will repeatedly exit and re-enter the health system until they reach the correct level. Although this does unburden higher levels of care (and is favoured as the future direction of health care by Hopp and Lovejoy [49]), we regard it as *system failure*, as upon their return, the health system must repeat the work already done on these patients.
We propose the relationship in Figure 5 to show *system failures* in complex systems, such as a health system. We do this by modifying the previous model, which represents the simple case shown in Figure 2.

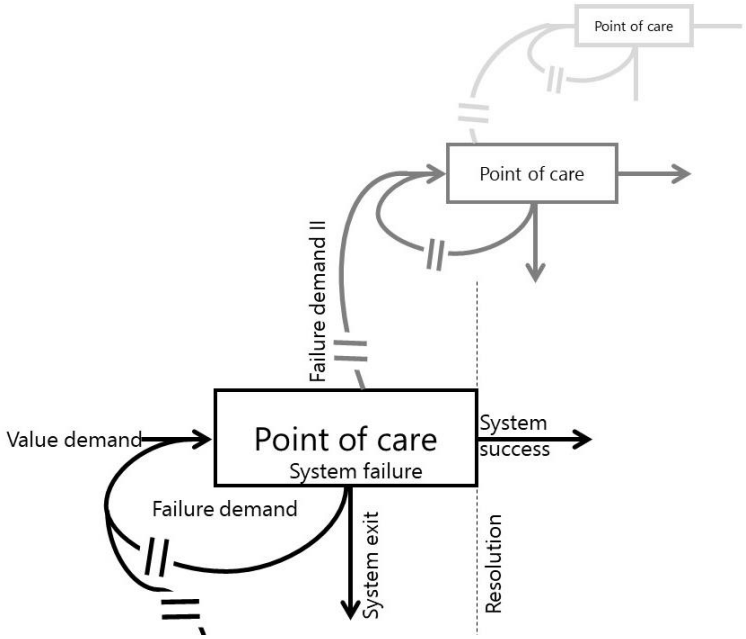

Figure 5: The occurrence of system failure[3] and failure demand in hierarchical service systems, such as health care

The *failure demand* model introduced in Figure 2 is repeated as the starting unit of the model above. The simple case assumes that *system failure*s can only exit the system or return to the same point of care. The expanded model shows that *system failure*s can cause *failure demand* at points of care other than the ones that created the *system failure*. Because the nature of this type of *system failure* is considerably different to the traditional *system failure — failure demand* relationships, we introduce the notation *failure demand II*.
Failure demand II emphasises that load should be seen with a systemic lens [50] and that just because work has been moved from one point to another may not benefit the system as a whole.

Our expert-panel supported this concept, with only one outlier, citing practicality rather than correctness as their major concern. The expert assessment of this idea delivered an average score of 4 out of a possible 5.

Conclusion: System failure — system design -> repeated consumption.

4.2 Unsuccessful[4] medicine

Medicine can be unsuccessful for many reasons. Some forms are *system failures*, while others are not. This section describes different forms of unsuccessful medicine from literature and personal experience and

[3] For easier reading, we only indicate 'failure demand II' moving from a lower level point of care to a higher one; but it must be assumed that it moves just as much down the hierarchy. The reader may also assume further levels of care above and below the model presented here.

[4] We introduce the concept of 'unsuccessful' medicine with caution. As this section will show, the success of a medical intervention is not always defined by cure, as in the case of chronic or palliative care. We cannot, however, head this section with the phrase 'non-curative' medicine, as that implies the *intention* not to cure, which belies categories such as experimental or trial-and-error interventions, which may not cure, yet strive to.

provides arguments for categorisation. We discuss chronic care, experimental medicine, trial and error medicine, palliative care, and medical mistakes.

4.2.1 Chronic Care

Chronic care is the care of diseases that '… are not passed from person to person. They are of long duration and generally slow progression. The four main types of non-communicable diseases are cardiovascular diseases (like heart attacks and stroke), cancers, chronic respiratory diseases (such as chronic obstructed pulmonary disease and asthma) and diabetes.' [51] [52, p. 2]

The nature of chronic care is that patients repeat on the system and as such, *resolution* in the simple sense has not been achieved. We propose that chronic care is not a *system failure* because the purpose of such care is not striving for a cure, but rather prioritising management of an ongoing condition [10]. *Resolution* in this case should be defined as the administration of an appropriate disease — management or diagnostic event — patient-exits are not *system failures*, but management events seen to resolution, that are programmed to return for further management.

Although chronic care is not an example of *system failure*, we propose that it should strive towards lower contact frequency, subject to the proviso that the health outcomes are not altered [53].

Conclusion: Not system failure — resolution is defined by 'care', not 'cure'.

4.2.2 Experimental medicine

Medicine errs towards familiar approaches [22] yet treatment ranges from conservative interventions to invasive, and often unnecessary, but costly treatments [54]. Non-conservative treatment is sometimes inappropriate and even morally questionable, however, at other times, it is the only response to an unfamiliar medical condition.

When confronted with unseen problems and unfamiliar cases, clinicians may need to innovate in their treatment approaches. In Bangladesh, where innovation is encouraged, health indicators have the best trajectory in South Asia [55].

Nevertheless, the innovating clinician must choose approaches that fit the condition. It is unreasonable to deem experimental medicine as *system failure*, unless it is done contrary to good clinical practice. We raise the caveat that the experimental approaches should be limited to unfamiliar conditions and even then, benchmarked against best clinical practices.

Conclusion: Not system failure — scientific limitations may limit the ability of medicine to cure.

4.2.3 Trial and error medicine

We view trial and error medicine as a nuanced type of experimental medicine. This relates specifically to clinicians refining an intervention through iterative methods. An example is depression medication dosage. In general, patients will receive anti-depressant medication, and the dosage will be modified until a level is reached at which a clinical response is achieved [56]. This trial-and-error approach is common and good clinical practice, and in our view is not a *system failure*.

We add the caveat that we view trial-and-error approaches as necessary but non-value-adding activities [57], and the system must strive for greater knowledge and data to reduce the amount of trial-and-error required in medicine.

Conclusion: Not system failure — scientific limitations may limit the ability of medicine to cure.

4.2.4 Palliative care

At times, patients have no remaining prospect for improved health. Nevertheless, they repeat on a health care system for palliative care — the maintenance of the best possible standard of living, which centres on comfort and dignity with no long-term survival expectations [58].

Conclusion: Not system failure — resolution is defined by 'care', not 'cure'.

4.2.5 Medical errors

Some deaths are avoidable. Kohn, Corrigan [21] speak about the burden of medical errors in the United States; they mention two studies that show that around three per cent of patient interactions contain what they refer to as adverse events. Between 44 000 and 98 000 patients die in US hospitals each year because of medical errors. Although Hayward and Hofer [59] argue that the number of deaths is exaggerated, they

do not dispute that medical errors are significantly problematic. Toussaint and Gerard [24] claim that US clinicians make up to 15 million medical errors annually, ranging from incorrect drugs or dosages[5] to incorrect site surgeries or infection. In the United Kingdom, the government reported that a million patients are 'put in hospitals' annually as a result of medical errors [24, 60].

Attention to reducing medical errors in a systematic way can have dramatic results, as was famously shown at Allegheny General Hospital in the United States, where the prevalence of central line infections was virtually eliminated [61] by using the principles of the Toyota production system [37]. Similar improvements were made at Virginia Mason Hospital (also in the US), simply by raising awareness of errors that medical professionals were making unwittingly [62].

Even though they are seen as important, and are often deadly [63], incorrect diagnoses are generally not included in the definitions of medical errors in the literature. However, we regard this as wrong, because it ignores root causes in favour of symptoms. We include diagnostic mistakes, and therefore conclude that errors are even higher than stated.

The majority of failures occur as a result of poorly designed systems [64]. This emphasises the value of designing systems intentionally compared with poorly or non-designed systems that allow (or even cause) errors. Although mistakes are 'human' [21], Deming's 94-6 principle [64] makes the point that the smaller proportion of mistakes (six per cent) is the result of human incompetence, negligence, or malice, while the bulk of errors (94 per cent) are accounted for by systemic issues.

The above reasoning tries to show that medical mistakes are *system failures,* and that there is evidence that strengthened systems lead to sustainably fewer errors and less harm. Thus patients who are fortunate enough to return after having experienced a medical error should be counted as *failure demand*.

Conclusion: System failure — system design does not minimise errors.

4.3 Patient errors

Patients are vital actors and stakeholders in a health system. Their actions are due as much scrutiny as those of clinicians. Although the patient is not included in a strict interpretation of Seddon's definition, we view this as a logical extension. A large portion of *system failure* is the result of patient error of some sort. This section will explore five typical patient errors: patients arriving at the wrong site, or at the wrong time, patients who disobey pre-treatment instructions, those who arrive with the incorrect paperwork, and those who take medication contrary to instructions.

4.3.1 Patients who arrive at the incorrect point of care

Some health systems (such as those in South Africa and Spain) are structured into health districts where patients have to be seen by the facility that serves the district in which they live [48, 65]. This means that a patient's address determines their health facility. When patients go to an 'incorrect' point of care, they are eventually turned away. Patients might also arrive at the incorrect *level*[6] of care; however, this is generally a clinical mode of *system failure,* and is separately addressed in Section 4.1.

Conclusion: System failure — system design and poor communication.

4.3.2 Patients who arrive for care at the incorrect time

Many health systems provide regularly scheduled care for certain conditions — for example, regular Aids clinics in Entebbe, Uganda [66], or regular diabetes clinics for military veterans in Los Angeles in the US [67]. In general, these clinics are scheduled for particular days, and patients who arrive on a different day do not receive the service.

Similarly, patients in scheduled care who arrive at a time other than their appointment time will often not be seen. In the Spanish system, some primary emergency departments only operate during business hours, requiring patients to move to general hospital accident and emergency departments after hours [48]. We have frequently observed an end-of-day migration of many dozens of un-served patients from the outpatients section to the emergency department.

---

[5] Drug errors can have two origins: incorrect scripts can be written, and the incorrect drugs can be given. In both cases, the drug itself or the dosage may be wrong [60]. Patients who take incorrect medicines are treated separately in this study.

[6] As opposed to *point* of care, which has a geographic element.

Conclusion: System failure — system design and poor communication.

4.3.3 Patients who disobey pre-treatment instructions

Many treatments require patient behaviours in preparation for treatment. For example, pre-surgical patients are required to 'starve' prior to surgery [68]. In a pilot study in a surgical theatre, we found that two per cent of procedures were cancelled because patients had eaten inside the 'starvation' period. In such cases patients must return on a future date for the same treatment. Similarly, some laboratory tests require a certain diet leading up to the actual test [69]. Non-conformance leads to repetition or incorrect results.

Conclusion: System failure — the system designed allows non-adherence; and also proper practice is inadequately communicated. Occasionally there is a wilful triggering act.

4.3.4 Patients who arrive with incorrect paperwork

Patients are seldom seen without suitable forms of national identification. This is particularly true in countries that have essentially free (from the point of view of the patient) health care. Equally, patients subject to '*failure demand* II' generally need to bring referral documents. These letters serve as the primary communication across levels in a health system [70], without which patients will often not be seen.

Conclusion: System failure — system design.

4.3.5 Patients who take medication contrary to instructions

The US Surgeon General reported that 75 per cent of Americans have trouble taking their medications as directed [71, 72]. This is worrying, especially when read together with Di Matteo *et al.*'s finding that clinical outcomes are three times worse in patients who have poor adherence [73]. These poor outcomes can include common adverse clinical side effects [74] or worse. McCarthy found that as many as 125 000 Americans die annually owing to improperly taken medication [76], although Meredith emphasises that the true scale of the problem is unknown [76]. A variety of reasons exists for poor adherence [78], which include polypharmacy (taking more than five medications daily [79], forgetfulness, unclear clinical instructions, and high costs [79]. Non-adherence is particularly common among people who live alone [74] or are elderly [80].

Conclusion: System failure — instructions not adequately communicated, or lack of remedies to compensate for patient-driven non-adherence.

4.3.6 Conclusion on patient errors

We propose that all the instances of patient errors presented above are a form of *system failure*. According to Deming [64], a system should be designed in such a way that people cannot make mistakes. Therefore, although the error is that of the patient, the fault lies with the health system that did not adequately inform patients of their duties, obligations, and functions.

The patient errors shown here are a consequence of patients being given inadequate information or an inadequate understanding of the available information. A system-wide intervention to inform patients, familiarise them with the operational modes of the health system, and encourage compliance is a systemic intervention that could reduce *failure demand*. The health system should be simultaneously re-engineered so that it is more difficult to make errors.

4.4 Operational environment

Medicine is, in many ways, idiosyncratic. This section evaluates these idiosyncrasies as potential systemic drivers of failure. We explore six common elements: financing, queues, supply-chain, support services, staffing, and infrastructure.

4.4.1 Patients who have insufficient financial means for treatment

Article 25 of the Universal Declaration of Human Rights [81] states that the right to health and health care is inviolable. Backman *et al.* argue not only that the right to health is 'good management, justice or humanitarianism', but also that providing such care is indeed an 'obligation under human rights law' [82, p. 2047]. Nevertheless, only a few nations can provide fair access to health care, despite more than sixty years of this principle being universally accepted.[7]

[7] Regrettably, of the thirty articles in the Universal Declaration of Human Rights, not a single one has been universally adopted.

Fairness is one of the three pillars of a health system, according to the World Health Organization [44]. The notion of fairness includes access. Economic exclusion means that a health system is unfair.

The cost of medical treatment is a matter of global concern. In many Organisation for Economic Co-operation and Development (OECD) states, the objective of the health system is to base care on need and not on means [83]. This study found that, generally, European states are more able to provide care on this basis, while in the United States, notably, health-seeking behaviour was significantly biased towards the wealthy.

In the United States, 29 per cent (Sarnak reports 33 per cent [84]) of patients take medications incorrectly, owing to the cost of consuming the medication at the correct rate [30], while 12 per cent of Americans cannot afford their medical bills [30]. Indeed, 64 per cent of Americans report 'the fear of unexpected medical expenses' as their greatest financial worry, ahead of mobility, heat, utilities, or having somewhere to stay [30]. As a result, half of Americans report being discouraged from seeking medical care [30], leading to self-medication and conditions being ignored [85].

The introduction of the Affordable Care Act in the United States [86], even in its naming, tried to address the high costs of care. Whether or not it has achieved this objective is unfortunately mired in political disagreement (see, e.g., [87]). However, the intention to reduce health care costs is noted as a priority.

Compared with developed countries, developing countries face an even greater burden from insufficient finance, which leads to even more pronounced exclusion from health services. For example, the Chinese health system is designed with the intention that the patient pays for services —, although a rapidly emerging health insurance industry exhibits a complexity that is beyond the scope of this paper [88].

Hsiao [89] reported that the Chinese system resulted in the economic exclusion of poorer people from care. Hall, Thomsen [90] showed how diabetes treatment in Sub-Saharan Africa impoverishes families, with Sudanese families spending up to two-thirds of their income on caring for a diabetic child. Leive and Xu [91] show how, in fifteen African states, between 30 per cent and 40 per cent (70 per cent in Burkina Faso) of people need to borrow money or sell property to cover their health care expenses at some time. In Burkina Faso it was found that up to 15 per cent of households suffer from catastrophic health costs, and this for relatively low levels of care [92].

Perhaps no single measure so completely reflects the failure of the entire system, from its intent to its functionality, as the exclusion, for financial reasons, of those who need care. Nevertheless, this is a daily reality in many countries globally, leading to *failure demand,* particularly because those so excluded could later return with even worse conditions. This places an increased demand on the health system.

Conclusion: System failure in the specific sense and, more broadly, as it refers to the overall purpose of health care in the first place.

4.4.2 Queues

The management of queues in health care systems is a significant research field. Many studies have attempted to shorten these queues, either through improved efficiency or by planning capacity better [93]. One reason to focus on queue length is that patients may balk and choose to leave if the queue for care is too long (leaving without being seen). In a simulated study at a real site in the United States, the losses attributable to balking amounted to over US$680 000 per month [95]. Methodologically, this is a transferable figure, suggesting that significant losses probably hold true in many similar environments.
Bottlenecks in system designs further accumulate load and reflect poor approaches to resource management.

Strategically, Allder *et al*. [95] identified demand-lag strategies — that is, where queues exist, but they remain stable and are, in effect, a buffer against system variation. Strategically [39], queues are only problematic if they continue to grow, which would indicate a systemic under-capacity. Tactically, however, balking as a result of predictable queue length is problematic [4], and steps should be taken to improve this.

The question of interest for this paper is whether long queues and the consequent balking is *system failure*. The reason for long queues is definitive in answering this question: queues are caused either by poor planning or by variation in load.

Conclusion: Both. Seddon speaks of 'predictability' [4]. If the queue length is predictable, then the cause is systemic, and the problem is system failure. Unpredictable failure, however, (such as a bus accident or a stadium stampede) is not system failure.

4.4.3 Supply chains and inventory management

Managing inventories of drugs, consumables, and other health care resources is of considerable importance.

In Blantyre, Malawi, it was found that a major reason for anti-retroviral non-adherence was frequent stock-outs in the pharmacy [96]. The returning patient was not only a *failure demand*, but was also quite possibly immune-compromised, thus potentially escalating the severity of the illness.

In a pilot study in a surgical ward, we found that more than 15 per cent of surgeries were cancelled because clean linen was not available. The absence of competent supply chains for necessary items led to the delay of treatment and the escalation of the severity of the condition; and we view it as *system failure*.

Conclusion: System failure — poor system design and triggering act.

4.4.4 Staff unavailability

In another study, we found that the late arrival or the non-availability of nurses, doctors, and anaesthetists accounted for roughly 30 per cent of surgical delays. We found that, in general, surgical days started more than 90 minutes after their scheduled start [97], delaying the whole schedule and often leading to cancelled procedures. We have observed similar issues in general practitioner (GP) and other practices, where the absence of a medical professional leads to *system failure*.

Conclusion: System failure — triggering act.

4.4.5 Delays from support or diagnostic services

Delayed laboratory results increased waiting times and reduced overall system efficiency [98]. A study found that introducing laboratories into an emergency department improved the unit's productivity by the same quantum as hiring an additional nurse and clinician [99]. Similarly, improved laboratory turnaround times reduce overall waiting times and improve health outcomes, as clinicians can make evidence-based decisions.

Long waiting times for laboratory results are therefore doubly *system failures*, as waiting times are increased and clinicians may make mistakes because of occasional time-pressured interventions, such as choosing to proceed with treatment before receiving delayed lab results.

Conclusion: System failure — poor system design.

4.4.6 Lack of infrastructure

Many health providers can only perform certain treatments if it can be foreseen that the patient can be admitted to a hospital bed either prior to or after treatment. We found that 23 per cent of surgical cancellations were the result of shortages of beds in high- or intensive-care units that were required for post-operative admissions. These patients had often been admitted to general wards and starved in preparation for elective surgery before it was cancelled.

Conclusion: System failure — system design and badly planned capacity and resource allocation.

5 SYSTEM FAILURE IN HEALTH CARE

5.1 Summary of system failure events

In the preceding sections we have differentiated between events that amount to *system failure* and those that do not.

Table 1 summarises the reasoning from these sections. The reader may follow the section headers in parentheses as a reminder of the argument in each case.

Recalling Equation (3), non-system-failure events represent detractors as much as they do value demand. In both instances, managing these cases and limiting their occurrence is possible and necessary; however, they do not represent a failure of the system, and their returning demand would be considered value demand.

Broadly speaking, a well-designed system is one that, in our view, has not failed. For example, Hopp and Lovejoy [49] show a variety of benchmarks for health-system design, on the basis of which bed numbers and system capacity are mandated. Should the demand on a well-designed system be lumpy or erratic [100] beyond the best judgement for its design, this could be excused as a capacity concern, and its classification as system failure would be unreasonable.

Table 1: Framework defining system failure in health care

System failure	Not **System failure**
Moving up or down the hierarchy (4.1)	**Unsuccessful medicine** (4.2)
Medical errors[8] (4.2.5)	**Chronic care** (4.2.1)
Patient errors (4.3)	**Experimental medicine** (4.2.2)
Wrong site (4.3.1)	**Trial-and-error medicine** (4.2.3)
Wrong time or date (4.3.2)	**Palliative care** (0)
Disobeyed pre-treatment instructions (4.3.3)	**Predicable long queues** (4.4.2)
Incorrect paperwork (4.3.4)	
Operational environment (4.4)	
Insufficient finances (4.4.1)	
Patients who take medication contrary to instruction (4.3.5)	
Unpredictable queues (4.4.2)	
Supply chain and inventory management (4.4.3)	
Staff unavailability (4.4.4)	
Support and diagnostic service delays (4.4.5)	
Lack of infrastructure (4.4.6)	

6 IMPLICATIONS

6.1 Theoretical contributions

We expand *failure demand* to introduce *system failure* to explain *failure demand* better, and then present researchers and practitioners with a model for identifying and measuring *system failure* and (therefore by extension, more holistically) *failure demand* in health care systems.

We emphasise the causal relationship between *system failure* as the root cause and *failure demand* as the symptom. In our model, derived from the literature and from our experience, events that are *system failure* are identified that, upon a patient's return, become *failure demand*.

The framework comments on the hierarchical nature of *system failure* in health care, and so introduces the idea of *failure demand* II — the mode of *failure demand* that crosses hierarchical bridges. This enables an overall systems view of health care facilities, including interconnectedness among facilities, which has implications for regional and national policy [39].

On a philosophical level, one expert asked whether any distance from the ideal should be seen as system failure. We suggest that the answer to this is 'yes', and that this logic must be applied in the same way that non-value-adding activities must be strictly identified [57], even if they are necessary or unavoidable. Our approach in categorising events was inspired by Lean best practice, which urges one to err on the side of severity when considering whether or not an activity is wasteful [57]. Defining an activity as value-adding leads to its long-term classification, and remedies to improve it are not sought. Following this reasoning, we tend to categorise ambiguous cases as *system failure*, rather than protecting such behaviours from future scrutiny.

6.2 Managerial implications

Failure demand has been found to range between 40 per cent and 80 per cent across a variety of industries. It would be interesting to establish the impact of *failure demand* in the health care industry. If the incidence of *failure demand* is high, that would provide an interesting insight into how much demand is avoidable.

Systematic interventions can reduce demand, meaning that capitalisation and staffing can be reduced or, more usefully, service levels could improve at no additional cost. This, in the face of considerable cost pressures on health systems globally, is desirable.

[8] Including incorrect diagnoses, and drug errors, which include wrong prescriptions and prescriptions incorrectly filled and, in both cases, the incorrect drug or inappropriate dosages. This category further includes wrong-site surgery and other medical errors.

Given that reduced *failure demand* could dramatically impact demand patterns, we advocate that health system strengthening policies incorporate the reduction of *system failure*, and considers failure demand as an important design consideration for strategic system design [39].

6.3 Societal impact

From ordinary observation, we have identified four cases that occur after *system failure*. The first case is the classic case of *failure demand*: patients who have experienced *system failure* return for repeat service in pursuit of resolution. The second case represents those patients who do not return but get better by themselves — so-called 'self-limiting conditions' [101] The third case represents patients who do not return to care, and do not get better, but also do not get worse [93], — what we refer to as the 'suffering in silence' population. The last group of patients do not return for medical care, and their conditions worsen and may be as severe as being fatal.

In the first case, the burden, which includes all the costs of *system failure*, is carried by the funder of medical care. In the other three cases, the full burden of *system failure* is carried by society.

To understand the reasons why patients transfer the burden to society, we propose that most patients who do not return do so because they are poorly informed, and do not realise that their condition can be improved. Even if they do believe that their condition can be improved, they have often come to mistrust the capabilities of health systems, or have become discouraged about seeking care. Further considerations include the time they have to take off work to seek care, and the financial means they require to pay for it.

6.4 Recommendation for further study

- This work should be expanded in field trials that explore the areas of concern.
- The impact of failure demand can now be measured in health care settings; so this should be done.
- Events that are not system failure — for example, the 'care spectrum' — should be examined in greater depth to find opportunities to create systems that unburden the health system, even when 'cure' is not the aim.

6.5 Conclusion

This paper contributes a clarified model for understanding system failure in complex hierarchical systems, such as health care. This model provides several opportunities for future work. Researchers could use the model to identify and measure *failure demand* in selected health care settings, while policy makers could incorporate thinking about failure demand into health system planning and design.

7 REFERENCES

[1] **Levine, R.J.** 1989. Institutional review boards. *British Medical Journal*, 298(6683), pp. 1268-1269.
[2] **World Medical Association.** 2001. World Medical Association Declaration of Helsinki: Ethical principles for medical research involving human subjects. *Bulletin of the World Health Organization*, 79(4), pp. 373-374.
[3] **Seddon, J.** 1992. *I want you to cheat! The unreasonable guide to service and quality in organisation*. New York: Vanguard Press.
[4] **Seddon, J.** 2005. *Freedom from command and control: Rethinking management for lean service*. Boca Raton: cRC Press.
[5] **Hopp, W.J. & Spearman, M.L.** 2011. *Factory physics*. Long Grove: Waveland Press.
[6] **Little, J.D.** 1961. A proof for the queuing formula. *Operations Research*, 9(3), pp. 383-387.
[7] **Roy, R.K.** 2010. *A primer on the Taguchi method*. Michigan: Society of Manufacturing Engineers.
[8] **Womack, J.** 2005. *Lean consumption*. Manufacturing Engineer, 84(4), pp. 5-5.
[9] **Seddon, J. & Brand, C.** 2008. Debate: Systems thinking and public sector performance. *Public Money & Management*, 28(1), pp. 7-9.
[10] **Lillrank, P., Groop, P.J. & Malmström, T.J.** 2010. Demand and supply–based operating modes — A framework for analyzing health care service production. *Milbank Quarterly*, 88(4), pp. 595-615.
[11] **Wallenius, J.J.** 2011. *Failure demand at telefonbanken, SEB*. Master of Science thesis, Chalmers University of Technology.
[12] **Marr, B. & Neely, A.** 2004. *Managing and measuring for value: The case of call centre performance*. Cranfield: Cranfield School of Management & Fujitsu Corporation.
[13] **Piercy, N. & Rich, N.** 2009. High quality and low cost: The lean service centre. *European Journal of Marketing*, 43(11/12), pp. 1477-1497.
[14] **Pidgeon, C.** 2010. *Systems thinking and public sector efficiencies*. Research Paper 050/10. Belfast: Northern Ireland Assembly, Research and Library Service.
[15] **Plsek, P.E. & Greenhalgh, T.** 2001. The challenge of complexity in health care. *British Medical Journal*, 323(7313), pp. 625-628.
[16] **Sasser, W.E.** 1976. Match supply and demand in service industries. *Harvard Business Review*, 54(6), pp. 133-140.
[17] **Radnor, Z.J., Holweg, M. & Waring, J.** 2012. Lean in healthcare: The unfilled promise? *Social Science & Medicine*, 74(3), pp. 364-371.
[18] **Walley, P.** 2013. Does the public sector need a more demand-driven approach to capacity management? *Production Planning & Control*, 24(10-11), pp. 877-890.
[19] **Bicheno, J.** 2012. *The service systems toolbox: Integrating lean thinking, systems thinking and design thinking*. Buckingham: Picsie Books.
[20] **Walley, P., Found, P. & Williams, S.** 2019. Failure demand: A concept evaluation in UK primary care. *International Journal of Health Care Quality Assurance*, 32(1), pp 21-33.
[21] **Kohn, L.T., Corrigan, J.M. & Donaldson, M.S. (eds).** 2000. *To err is human: Building a safer health system*. Washington: National Academies Press.
[22] **Sox, H.C., Lurie, J.D.** 1988. *Medical decision making*. Sydney: ACP Press.
[23] **Hartmann, D. & Mandavha, R.** 2010. *Lean healthcare, a casualty of inefficiency*. SAIIE Conference. Pretoria: Southern African Institute of Industrial Engineering.
[24] **Toussaint, J. & Gerard, R.** 2010. *On the mend: Revolutionizing healthcare to save lives and transform the industry*. Cambdisge, Massachusetts: Lean Enterprise Institute.
[25] **Graban, M.** 2008. *Lean hospitals*. Boca Raton, London, New York: CRC Press.
[26] **Hartmann, D.** 2012. *Improvement in operating theatre efficiency through better measurement and scheduling*. MSc research report. Johannesburg: University of the Witwatersrand Press.
[27] **Wagstaff, A., Van Doorslaer, E. & Paci, P.** 1991. On the measurement of horizontal inequity in the delivery of health care. *Journal of Health Economics*, 10, pp. 169-205.
[28] **Hopp, W.J.** 2011. *Supply chain science*. Long Grove: Waveland Press.
[29] **Allder, S., Walley, P. & Silvester, K.** 2011. Is follow-up capacity the current NHS bottleneck? *Clinical Medicine*, 11(1), pp. 31-34.
[30] **Dijulio, B., Kirzinger, A., Wu, B. & Brodie, M.** 2017. Data note: Americans' challenges with health care costs. *Kaiser Family Foundation*. Available at: https://www.kff.org/health-costs/issue-brief/data-note-americans-challenges-health-care-costs/
[31] **Bicheno, J.** 2018. *Towards reducing queues: Muri, mura, muda*. European Lean Educator Conference (ELEC2018), Braga, Portugal.
[32] **McCarthy, D. & Rich, N.** 2015. *Lean TPM: A blueprint for change*. Oxford: Butterworth-Heinemann.
[33] **Graeber, D.** 2013. *On the phenomenon of bullshit jobs: A work rant*. Strike Magazine (3) pp. 1-5
[34] **Roy, R.B., Lillrank, P., Sreekanth, V. & Torkki, P.** 2019. *The conceptual tools* in *Designing Service Machines: Translating Principles of System Science to Service Design* (2019). Singapore:Springer. pp. 5-19.
[35] **Dijkstra, J., Galbraith, R., Hodges, B.D., McAvoy, P.A., McCrorie, P., Southgate, L.J., Van der Vleuten, C.P., Wass, V. & Schuwirth, L.W.** 2012. Expert validation of fit-for-purpose guidelines for designing programmes of assessment. *BMC Medical Education*, 12(1), pp. 1-8.
[36] **Senge, P.M.** 1997. The fifth discipline. *Measuring Business Excellence*, 1(3), pp. 46-51.
[37] **Liker, J.K.** 2004. *The Toyota way*. New York: McGraw Hill.
[38] **Hans, E.W., Van Houdenhoven, M. & Hulshof, P.J.** 2012. A framework for healthcare planning and control. In *Handbook of healthcare system scheduling* (2012). Hall, R. (ed). Boston, Massachusetts:Springer. pp. 303-320.

[39] **Hines, P., Holweg, M. & Rich, N.** 2004. Learning to evolve: A review of contemporary lean thinking. *International Journal of Operations & Production Management*, 24(10), pp. 994-1011.
[40] **levitt, T.** 1972. Production-line approach to service. *Harvard Business Review*, 50(5), pp. 41-52.
[41] **Seddon, J., O'Donovan, B. and Zokaei, K.** 2011. Rethinking lean service. In *Service design and delivery (2011)* Macintyre, M., Parry, G., Angelis, J.(Eds.). Boston, Massachusetts: Springer. pp 41-60.
[42] **Seddon, J. & Caulkin, S.** 2007. Systems thinking, lean production and action learning. *Action Learning: Research and Practice*, 4(1), pp. 9-24.
[43] **Frenk, J.** 1994. Dimensions of health system reform. *Health Policy*, 27(1), pp. 19-34.
[44] **World Health Organization.** 2000. *The world health report 2000: Health systems: Improving performance*. Geneva: World Health Organization.
[45] **United States Department of Health Human Services.** 1991. Healthy people 2000: National health promotion and disease prevention objectives. In *Healthy people 2000: National health promotion and disease prevention objectives*. Centres for Disease Control and Prevention. Washington D.C.: US Government Printing Office.
[46] **Tajeu, G.** 2014. *Health maintenance organization (HMO)*. In *The Wiley Blackwell Encyclopedia of Health, Illness, Behavior, and Society* (2014), *Cockerham, W., Dingwall, R., Quah S.R.* (eds.). Hoboken NJ.: John Wiley & Sons.
[47] **Womack, J.P. & Jones, D.T.** 1996. *Lean thinking: Banish the waste and create in your corporation*. London: Simon and Schuster.
[48] **Sempere-Selva, T., Peiró, S., Sendra-Pina, P., Martínez-Espín, C. & López-Aguilera, I.** 2001. Inappropriate use of an accident and emergency department: Magnitude, associated factors, and reasons — an approach with explicit criteria. *Annals of Emergency Medicine*, 37(6), pp. 568-579.
[49] **Hopp, W.J. & Lovejoy, W.S.** 2012. *Hospital operations: Principles of high efficiency health care*. Upper Saddle River, NJ: FT Press.
[50] **Checkland, P.** 1999. Systems thinking. In *Rethinking management information systems (1999), Currie W.L., Galliers, b. (Eds.)*. Oxford: Oxford University Press, pp. 45-56.
[51] **World Health Organization**. 2015. *Noncommunicable diseases*. Health Topics 2015 2017-06-26 [date accessed 2017 2017-06-26]. https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwiIns_AwNjyAhULG-wKHSjFAH4QFnoECAUQAQ&url=https%3A%2F%2Fwww.who.int%2Fnews-room%2Ffact-sheets%2Fdetail%2Fnoncommunicable-diseases&usg=AOvVaw33J15SJYEdDlJSqjwurHYf
[52] **Bernell, S. & Howard, S.W.** 2016. Use your words carefully: What is a chronic disease? *Frontiers in Public Health*, 4, p. 159.
[53] **Wagner, E.H., Grothaus, L.C., Sandhu, N., Galvin, M.S., McGregor, M., Artz, K. & Coleman, E.A.** 2001. Chronic care clinics for diabetes in primary care: A system-wide randomized trial. *Diabetes Care*, 24(4), pp. 695-700.
[54] **Wennberg, J.E.** 1984. Dealing with medical practice variations: A proposal for action. *Health Affairs*, 3(2), pp. 6-32.
[55] **Das, P. & Horton, R.** 2013. Bangladesh: Innovating for health. *The Lancet*, 382(9906), pp. 1681-1682.
[56] **Paykel, E.S. & Priest, R.G.** 1992. Recognition and management of depression in general practice: Consensus statement. *British Medical Journal*, 305(6863), pp. 1198-1202.
[57] **Bicheno, J. & Holweg, M.** 2016. *The lean toolbox: A handbook for lean transformations*, 5th edition. Buckingham: PICSIE Books.
[58] **Jin, J.** 2013. Clinicians examine advances and challenges in improving quality of end-of-life care in the ICU. *JAMA*, 310(23), pp. 2493-2495.
[59] **Hayward, R.A. & Hofer, T.P.** 2001. Estimating hospital deaths due to medical errors: Preventability is in the eye of the reviewer. *JAMA*, 286(4), pp. 415-420.
[60] **Smyth, C.** 2017. Million patients a year put in hospital by medicine mistakes. London: *The Sunday Times* (September 6 2017) https://www.thetimes.co.uk/article/million-patients-a-year-put-in-hospital-by-medicine-mistakes-96h5kgb3t.
[61] **Shannon, R.P., Frndak, D., Grunden, N., Lloyd, J.C., Herbert, C., Patel, B., Cummins, D., Shannon, A.H., O'Neill, P.H. & Spear, S.J.** 2006. Using real-time problem solving to eliminate central line infections. *The Joint Commission Journal on Quality and Patient Safety*, 32(9), pp. 479-487.
[62] **Kenney, C.** 2012. *Transforming health care: Virginia Mason Medical Center's pursuit of the perfect patient experience.Boca Raton:* CRC Press.
[63] **Kirch, W. & Schafii, C.** 1996. Misdiagnosis at a university hospital in 4 medical eras: Report on 400 cases. *Medicine*, 75(1), pp. 29-40.
[64] **Deming, W.E.** 1986. *Out of the crisis*. Cambridge, MA: Massachusetts Institute of Technology, Center for Advanced Engineering Study.
[65] **Van Rensburg, H.** 2004. *Health and health care in South Africa*. Pretoria, South Africa: Van Schaik Publishers.
[66] **Watera, C., Todd, J., Muwonge, R., Whitworth, J., Nakiyingi-Miiro, J., Brink, A., Miiro, G., Antvelink, L., Kamali, A., French, N. & Mermin, J.** 2006. Feasibility and effectiveness of cotrimoxazole prophylaxis for HIV-1-infected adults attending an HIV/AIDS clinic in Uganda. *Journal of Acquired Immune Deficiency Syndromes*, 42(3), pp. 373-378.
[67] **Ho, M., Marger, M., Beart, J., Yip, I. & Shekelle, P.** 1997. Is the quality of diabetes care better in a diabetes clinic or in a general medicine clinic? *Diabetes Care*, 20(4), pp. 472-475.
[68] **Maclean, A. & Renwick, C.** 1993. Audit of pre-operative starvation. *Anaesthesia*, 48(2), pp. 164-166.
[69] **Kale, V.P., Joshi, G.S., Gohil, P.B., and Jain, M.R.** 2009. Effect of fasting duration on clinical pathology results in wistar rats. *Veterinary Clinical Pathology*, 38(3), pp. 361-366.
[70] **Grol, R., Rooijackers-Lemmers, N., Van Kaathoven, L., Wollersheim, H. & Mokkink, H.** 2003. Communication at the interface: Do better referral letters produce better consultant replies? *British Journal of General Practice*, 53(488), pp. 217-219.

[71] **Benjamin, R.M.** 2012. Medication adherence: Helping patients take their medicines as directed. *Public Health Reports*, 127(1), pp. 2-3.
[72] **Cutler, D.M. & Everett, W.** 2010. Thinking outside the pillbox: Medication adherence as a priority for health care reform. *New England Journal of Medicine* 362 (17), pp 1553–1555
[73] **Dimatteo, M.R., Giordani, P.J., Lepper, H.S. & Croghan, T.W.** 2002. Patient adherence and medical treatment outcomes: A meta-analysis, *Medical Care*, 40, pp. 794-811.
[74] **Ellenbecker, C.H., Frazier, S.C. & Verney, S.** 2004. Nurses' observations and experiences of problems and adverse effects of medication management in home care. *Geriatric Nursing*, 25(3), pp. 164-170.
[75] **Mccarthy, R.** 1998. The price you pay for the drug not taken. *Business and Health*, 16(10), pp. 27-33.
[76] **Meredith, J.O., Grove, A.L., Walley, P., Young, F. & Macintyre, M.B.** 2011. Are we operating effectively? A lean analysis of operating theatre changeovers. *Operations Management Research*, 4(3-4), pp. 89-98.
[77] **Burnier, M.** 2006. Medication adherence and persistence as the cornerstone of effective antihypertensive therapy. *American Journal of Hypertension*, 19(11), pp. 1190-1196.
[78] **Bedell, S.E., Jabbour, S., Goldberg, R., Glaser, H., Gobble, S., Young-Xu, Y., Graboys, T.B. & Ravid, S.** 2000. Discrepancies in the use of medications: Their extent and predictors in an outpatient practice. *Archives of Internal Medicine*, 160(14), pp. 2129-2134.
[79] **The Commonwealth Fund.** 2005. *Improving health care quality: International health policy survey of sicker adults*. New York: The Commonwealth Fund.
[80] **Mulhem, E., Lick, D., Varughese, J., Barton, E., Ripley, T. & Haveman, J.** 2013. Adherence to medications after hospital discharge in the elderly. *International Journal of Family Medicine, vol. 2013, 6 pages*
[81] **UN General Assembly.** 1948. *Universal declaration of human rights*. New York: UN General Assembly.
[82] **Backman, G., Hunt, P., Khosla, R., Jaramillo-Strouss, C., Fikre, B.M., Rumble, C., Pevalin, D., Páez, D.A., Pineda, M.A., Frisancho, A., Tarco, D., Motlagh, M., Farcasanu, D. & Vladescu, C.** 2008. Health systems and the right to health: An assessment of 194 countries. *The Lancet*, 372(9655), pp. 2047-2085.
[83] **Hurst, J.** 2002. Performance measurement and improvement in OECD health systems: Overview of issues and challenges. In OECD, *Measuring up: Improving health system performance in OECD countries*. Paris: OECD Publishers.
[84] **Sarnak, D.O., Squires, D., Kuzmak, G. & Bishop, S.** 2017. Paying for prescription drugs around the world: Why is the US an outlier? *Issue Brief (The Commonwealth Fund)*, October, pp. 1-14.
[85] **Osborn, R., Squires, D., Doty, M.M., Sarnak, D.O. & Schneider, E.C.** 2016. In new survey of eleven countries, US adults still struggle with access to and affordability of health care. *Health Affairs*, 35(12), pp. 2327-2336.
[86] **Affordable Care Act.** 2010. *Public Law 111-148*, in *Title IV, x4207, USC HR*.
[87] **Kaplan, T. & Pear, R.** 2017. Unity is elusive as G.O.P. presses health overhaul. *New York Times*, Arthur Ochs Sulzberger Jr. (ed)
[88] **Blumenthal, D. & Hsiao, W.** 2005. Privatization and its discontents: The evolving Chinese health care system. *New England Journal of Medicine*, 353, pp. 1165-1170.
[89] **Hsiao, W.C.** 1995. The Chinese health care system: Lessons for other nations. *Social Science & Medicine*, 41(8), pp. 1047-1055.
[90] **Hall, V., Thomsen, R.W., Henriksen, O. & Lohse, N.** 2011. Diabetes in sub Saharan Africa 1999-2011: Epidemiology and public health implications. A systematic review. *BMC Public Health*, 11(564) pp 1-12.
[91] **Leive, A. & Xu, K.** 2008. Coping with out-of-pocket health payments: Empirical evidence from 15 African countries. *Bulletin of the World Health Organization*, 86, pp. 849-856C.
[92] **Su, T.T., Kouyaté, B. & Flessa, S.** 2006. Catastrophic household expenditure for health care in a low-income society: A study from Nouna district, Burkina Faso. *Bulletin of the World Health Organization*, 84, pp. 21-27.
[93] **Silvester, K., Lendon, R., Bevan, H., Steyn, R. & Walley, P.** 2004. Reducing waiting times in the NHS: Is lack of capacity the problem? *Clinician in Management*, 12(3), pp. 105-109.
[94] **Broyles, J.R.** 2007. Estimating business loss to a hospital emergency department from patient reneging by queuing-based regression. *IIE Annual Conference Proceedings*. Institute of Industrial and Systems Engineers (IISE), pp. 613 - 613.
[95] **Balderston, D., Gonzalez, M. & López, A.M.** 2000. *Encyclopedia of contemporary Latin American and Caribbean cultures*. Abingdon: Routledge.
[96] **Van Oosterhout, J.J., Bodasing, N., Kumwenda, J.J., Nyirenda, C., Mallewa, J., Cleary, P.R., De Baar, M.P., Schuurman, R., Burger, D.M. & Zijlstra, E.E.** 2005. Evaluation of antiretroviral therapy results in a resource-poor setting in Blantyre, Malawi. *Tropical Medicine & International Health*, 10(5), pp. 464-470.
[97] **Hartmann, D. & Sunjka, B.P.** 2013. Private theatre utilisation in South Africa: A case study. *South African Medical Journal*, 103(5), pp. 285-287.
[98] **Hannan, E.L., Giglio, R.J. & Sadowski, R.S.** 1974. A simulation analysis of a hospital emergency department. *Proceedings of the 7th Conference on Winter Simulation, Volume 1*. ACM. pp. 379-388.
[99] **Paul, S.A., Reddy, M.C. & Deflitch, C.J.** 2010. A systematic review of simulation studies investigating emergency department overcrowding. *Simulation*, 86(8-9), pp. 559-571.
[100] **Ghobbar, A.A. & Friend, C.H.** 2002. Sources of intermittent demand for aircraft spare parts within airline operations. *Journal of Air Transport Management*, 8(4), pp. 221-231.
[101] **Moerman, D.E.** 2002. *Meaning, medicine, and the 'placebo effect'*. Cambridge: Cambridge University Press.