Consumer-driven e-commerce
A literature review, design framework, and research agenda on last-mile logistics models

Stanley Frederick W.T. Lim, Xin Jin and Jagjit Singh Srai
Institute for Manufacturing, Department of Engineering,
University of Cambridge, Cambridge, UK

Abstract

Purpose – The purpose of this paper is to re-examine the extant research on last-mile logistics (LML) models and consider LML’s diverse roots in city logistics, home delivery and business-to-consumer distribution, and more recent developments within the e-commerce digital supply chain context. The review offers a structured approach to what is currently a disparate and fractured field in logistics.

Design/methodology/approach – The systematic literature review examines the interface between e-commerce and LML. Following a protocol-driven methodology, combined with a “snowballing” technique, a total of 47 articles form the basis of the review.

Findings – The literature analysis conceptualises the relationship between a broad set of contingency variables and operational characteristics of LML configuration (push-centric, pull-centric, and hybrid system) via a set of structural variables, which are captured in the form of a design framework. The authors propose four future research areas reflecting likely digital supply chain evolutions.

Research limitations/implications – To circumvent subjective selection of articles for inclusion, all papers were assessed independently by two researchers and counterchecked with two independent logistics experts. Resulting classifications inform the development of future LML models.

Practical implications – The design framework of this study provides practitioners insights on key contingency and structural variables and their interrelationships, as well as viable configuration options within given boundary conditions. The reformulated knowledge allows these prescriptive models to inform practitioners in their design of last-mile distribution.

Social implications – Improved LML performance would have positive societal impacts in terms of service and resource efficiency.

Originality/value – This paper provides the first comprehensive review on LML models in the modern e-commerce context. It synthesises knowledge of LML models and provides insights on current trends and future research directions.

Keywords Literature review, E-commerce, Omnichannel, Digital supply chains, Last-mile

Introduction

Last-mile delivery has become a critical source for market differentiation, motivating retailers to invest in a myriad of consumer delivery innovations, such as buy-online-pickup-in-store, autonomous delivery solutions, lockers, and free delivery upon minimum purchase levels (Lim et al., 2017). Consumers care about last-mile delivery because it offers convenience and flexibility. For these reasons, same-day and on-demand delivery services are gaining traction for groceries (e.g. Deliv Fresh, Instacart), pre-prepared meals (e.g. Sun Basket), and retail purchases (e.g. Dropoff, Amazon Prime Now) (Lopez, 2017). To meet customer needs, parcel carriers are increasing investments into urban and automated distribution hubs (McKevitt, 2017). However, there is a lack of understanding as to how best to design last-mile delivery models with retailers turning to experimentations that, at times, attract scepticism from...
industry observers (e.g. Cassidy, 2017). For example, Sainsbury’s, Somerfield, and Asda established innovative pick centres, but closed them down within a few years (Fernie et al., 2010). eBay launched its eBay Now same-day delivery service in 2012, but in July 2015, it announced the closure of this programme. Google, likewise, opened and then closed its two delivery hubs for Google Express in 2013 and 2015, respectively (O’Brien, 2015).

The development of these experimental last-mile logistics (LML) models, not surprisingly, created uncertainty within increasingly complicated and fragmented distribution networks. Without sustainable delivery economics, last-mile service provision will struggle to survive (as was the experience of Sainsbury’s, Somerfield, Asda, eBay, Google, and Webvan) with retailers increasingly challenged to find an optimal balance between pricing, consumer expectations for innovative new channels, and service levels (Lopez, 2017; McKevitt, 2017).

Although several contributions have been made in the LML domain, the literature on LML models remains relatively fragmented, thus hindering a comprehensive and holistic understanding of the topic to direct research efforts. Hitherto, existing studies provide limited or no guidance on how contingency variables influence the selection of LML configurations (Agatz et al., 2008; Fernie et al., 2010; Mangiaracina et al., 2015; Lagorio et al., 2016; Savelsbergh and Van Woensel, 2016). Our paper addresses this knowledge deficiency by reviewing the disparate academic literature to capture key contingency and structural variables characterizing the different forms of last-mile distribution. We then theoretically establish the connection between these variables thereby providing a design framework for LML models. Our corpus is comprised of 47 papers published in 16 selected peer-reviewed journals during the period from 2000 to 2017. The review is performed from the standpoint of retailers operating LML. As such, some LML research streams are deliberately excluded, including issues related to public policy, urban traffic regulations, logistics infrastructure, urban sustainability, and environment.

This paper is structured as follows. First, we provide a working definition for LML and introduce relevant terms. Second, we set out the research methodology and conduct descriptive analyses of the corpus. The substantive part of the paper is an analysis of the literature on LML models and the development of a design framework for LML. The framework synthesises a set of structural and contingency variables and explicates their interrelationships, shedding light on how these interactions influence LML design. Finally, we highlight the key gaps in the extant literature and propose future research opportunities.

Defining LML
The term “last-mile” originated in the telecommunications industry and refers to the final leg of a network. Today, LML denotes the last segment of a delivery process, which is often regarded as the most expensive, least efficient aspect of a supply chain and with the most pressing environmental concerns (Gevaers et al., 2011). Early definitions of LML were narrowly stated as the “extension of supply chains directly to the end consumer”; that is, a home delivery service for consumers (Punakivi et al., 2001; Kull et al., 2007). Several synonyms, such as last-mile supply chain, last-mile, final-mile, home delivery, business-to-consumer distribution, and grocery delivery, have also been used.

Despite their nuances, existing LML definitions converge on a common understanding that refers to the last part of a delivery process. However, existing definitions (details available from the authors) appear incomplete in capturing the complexities driven by e-commerce, such as omission in defining an origin (Esper et al., 2003; Kull et al., 2007; Gevaers et al., 2011; Ehmke and Mattfeld, 2012; Dablanc et al., 2013; Harrington et al., 2016); exclusion of in-store order fulfilment processes as a fulfilment option (Hubner, Kuhn and Wollenburg, 2016); and/or non-specification of the destination (or end point), including failure to capture the collection delivery point (CDP) as a reception option (Esper et al., 2003;
Kull et al., 2007). Without a consistent and robust definition of LML, the design of LML models is problematic. For the purpose of this review, we examine existing terminology on last-mile delivery systems in order to create a working definition for LML. As part of this definition, we introduce the concept of an “order penetration point” (Fernie and Sparks, 2009) as a way of defining the origin of the last-mile. The order penetration point refers to an inventory location (e.g. fulfilment centre, manufacturer site, or retail store) where a fulfilment process is activated by a consumer order. After this point, products are uniquely assigned to the consumers who ordered them, making the order penetration point a natural starting point for LML. The destination point is commonly dictated by the consumer, hence we use “final consignee’s preferred destination point” as the terminology to indicate where an order is delivered. The choice of destination point could be a home/office, a reception box (RB), or a pre-designated CDP.

Synthesising these developments, we propose the following working definition for LML:

Last-mile logistics is the last stretch of a business-to-consumer (B2C) parcel delivery service. It takes place from the order penetration point to the final consignee’s preferred destination point.

Extending the above definition, we reference Bowersox et al.’s (2012) view of a supply chain as a series of “cycles”, with half the cycle being the product/order flow and the other, information flow. We also reference the Supply Chain Operations Reference model (Supply Chain Council, 2010) recognising that LML operates within a broader supply network. In particular, the LML cycle coincides with the consumer service cycle, interfacing direct-to-consumer-goods manufacturers, wholesalers, retailers with the end consumer (Bowersox et al., 2012). The process may be divided into three sub-processes, namely source, make, and deliver.

These three sub-processes set the focus for this review and they align with the delivery order phase in LML, namely picking, packing, and delivery. This model is consistent with Campbell and Savelsbergh’s (2005) view of the business-to-consumer process comprising order sourcing, order assembly, and order delivery. Accordingly, this review focuses on the examination of LML models: LML distribution structures and the contingency variables associated with these structures. The term “distribution structure” covers the stages from order fulfilment to delivery to the final consignee’s preferred destination point. It includes the modes of picking (e.g. warehouse or store-based), transportation (e.g. direct delivery by the retailer’s own fleet), and reception (e.g. consumer-pickup) (Kämäräinen et al., 2001). The associated contingency variables provide guidance for decision support by highlighting the key characteristics of each distribution structure for the design and selection, matching product, and consumer attributes (Boyer and Hult, 2005).

Research methodology

A structured literature review aims to identify the conceptual content of a rapidly growing field of knowledge, as well as to provide guidance on theory development and future research direction (Meredith, 1993; Easterby-Smith et al., 2002; Rousseau et al., 2008). Structured reviews differ from more narrative-based reviews because of the requirement to provide a detailed description of the review procedure in order to reduce bias; this requirement thereby increases transparency and replicability (Tranfield et al., 2003). Therefore, undertaking a structured review ensures the fidelity, completeness, and rigour of the review itself (Greenhalgh and Peacock, 2005).

Our review provides a snapshot of the diversity of theoretical approaches present in LML literature. It does not pretend to cover the entirety of the literature, but rather offers an informative and focused evaluation of purposefully selected literature to answer specific research questions. In the following sections, we discuss the execution of the four main steps (planning, searching, screening, and extraction/synthesis/reporting) as outlined by
Tranfield et al. (2003). We also incorporate literature review guidelines suggested by Saenz and Koufteros (2015). Our study uses key research questions identified by an expert panel and we reference the Association of Business Schools journal ranking 2015 to decide which journals to include in this scholarship (Cremer et al., 2015). Our review includes the classification of contributions across methodological domains. In later sections, we utilise insights from the literature review to develop an LML design framework that captures the relationships between distribution structures via a set of structural and associated contingency variables.

Planning
At the outset of this study, a panel was formed comprising academic scholars and industry practitioners who are experts in the field of LML and retail operations. These experts participated in structured meetings in which the critical research needs related to LML design were discussed. The following research questions were then formulated based upon the emergent themes:

RQ1. What is the current state of research and practice on LML distribution in the e-commerce context?

RQ2. What are the associated contingency variables that can influence the selection of LML distribution structures?

RQ3. How can the contingency variables identified in RQ2 be used to inform the selection of LML distribution structures?

Searching
The academic material in this study covers the period from 2000 to 2017. This period coincides with critical industry events, such as the emergence and subsequent demise of the online grocer, Webvan. The review is limited to peer-reviewed publications to ensure the quality of the corpus (Saenz and Koufteros, 2015) and considers 16 journals, including one practitioner journal (MIT Sloan Management Review), to capture theoretical perspectives on industry best practices. Only articles from the selected journals have been included in this review, with one exception, where we included the article by Wang et al. (2014), published in Mathematical Problems in Engineering. The article was deemed critical as it represents the only piece of work to date that connects and extends prior research on the evaluation of CDPs.

The 16 journals were selected based on their primary focus on empirical and conceptual development, rather than on their discussion of analytical modelling. Although we appreciate that there are significant research studies in this area (e.g. operations research), the focus of this review led us to primarily consider how scholars conceptualise LML distribution structures and apply theoretical variables to LML design through quantitative, qualitative, or conceptual approaches, rather than through mathematically based models. The mathematically based model literature focuses on the development of stylised and optimisation models in areas of multi-echelon distribution systems, vehicle routing problems (Savelsbergh and Van Woensel, 2016), buy-online-pick-up-in-store services (Gao and Su, 2017), pricing and delivery choice, inventory-pricing, delivery service levels, discrete location-allocation, channel design, and optimal order quantities via newsvendor formulation for different fulfilment options (Agatz et al., 2008), amongst others. These studies typically employ a series of assumptions to simplify real-world operations in order to provide closed-form or heuristic-based prescriptive solutions (Agatz et al., 2008; Savelsbergh and Van Woensel, 2016). Therefore, this review excluded journals with a primarily mathematical modelling or operations research focus. However, it included relevant mathematical modelling articles – published in any of the 16 selected journals – as...
long as they explicated types of LML distribution structure and/or the associated contingency variables. Finally, this study also excluded general management journals in order to fit the operational focus of this research.

The literature search was conducted on the following databases: ISI Web of Science, Science Direct, Scopus, and ABI/Inform Global. Two search rounds were undertaken to maximise inclusion of all relevant articles. The first literature probe was performed using the following search terms: “urban logistics” OR “city logistics” OR “last-mile logistics” OR “last mile logistics”. To extend the corpus, we incorporated the “snowballing” technique of tracing citations backward and forward to locate leads to other related articles; this study used this process in the second round to supplement a protocol-driven methodology. This approach resulted in new search terms and scholar identification to refine the search strategy as the study unfolded (Greenhalgh and Peacock, 2005). The following new search terms were identified: “home delivery”, “B2C distribution”, “extended supply chain”, “final mile”, “distribution network”, “distribution structure”, and “grocery delivery”. These new keywords were then used to create additional search strings with Boolean connectors (AND, OR, AND NOT). Finally, in order to cross-check the searches, we consulted with a supply chain professor from Arizona State University and one from the University of Cambridge. It is therefore posited that the review coverage is reasonably comprehensive.

**Screening**

For the purpose of this review, the abstract of each paper selected was studied in order to assess its relevance to LML in the e-commerce context. When an abstract proved inconclusive or unavailable, the full paper was examined to determine the relevant contribution to the LML topic. We evaluated article relevance using the following criteria in order to ensure fidelity and objectivity:

1. **Exclusion criteria:** paper titles bearing the terms “urban logistics”, “city logistics”, “last-mile logistics”, or “last-mile” but with limited coverage on distribution structures and the associated design variables were excluded (e.g. public policy, urban traffic regulations, logistics infrastructure, urban sustainability, environment), as were editorial opinions, conference proceedings, textbooks, book reviews, dissertations, and unpublished working papers.

2. **Inclusion criteria:** papers with coverage of distribution structures and design variables in the e-commerce context were included, regardless of their actual study focus. We included multiple research methods to have both established findings as well as emergent theorising.

During the search phase, we identified 425 articles referencing our subject terms. We eliminated duplicates based on titles and name of authors and rejected articles matching the exclusion criteria. For example, while the paper by Gary et al. (2015) holds the keyword “urban logistics” in the title, it focuses on logistics prototyping, rather than LML models, as a method to engage stakeholders. This paper, therefore, was excluded. The elimination stage resulted in 100 articles being considered relevant for further review. Results were exported to reference management software, EndNote version X8, for further review and to facilitate data management. We then adopted the inclusion criteria to select the final articles. Finally, we grouped the articles into two categories: LML distribution structures and the associated contingency variables. Ultimately, a total of 47 journal articles form the corpus of this review.

**Extraction, synthesis, and reporting**

Following an initial review of the 47 articles, a summary of each article was prepared using a spreadsheet format organised under descriptive (year, journal, title), methodology (article type,
theoretical lens, sampling protocol), and thematic categories (article purpose, context, LML distribution structures, design variables, others) as adapted from Pilbeam et al. (2012).

Accordingly, we conducted three analyses: descriptive, methodological, and thematic (Richard and Beverly, 2014). The descriptive analysis summarises the research development over the period of interest, and the distribution statistics of the journals. The methodological analysis highlights the research methods employed in the domain, while the thematic analysis synthesises the main outcomes from the extracted literature and provides an overview of the review structure. Reporting structures were organised in a manner that sequentially responds to the research questions posed earlier.

Descriptive analysis

Table I provides summary statistics of the papers reviewed, author affiliations, identifying contributions, as well as those journals where surprisingly contributions have yet to be made.

Methodological analysis

While research methods used in this field vary, the articles studied primarily used analytical modelling and simulation (19 articles (40.43 per cent)), statistical methods (6 articles (12.77 per cent)), and survey (7 articles (14.89 per cent)). Remaining articles comprised case research, including mixed-method scholarship (6 articles (12.77 per cent)), conceptual development (5 articles (10.64 per cent)), and literature reviews (4 articles (8.51 per cent)). We framed these research methods in our study as typology-oriented provision, literature review and conceptual studies, empirical studies, and mathematical modelling. We briefly cover representative articles for each type below:

- Typology-oriented provision: owing to the recent proliferation of LML models, a typology-oriented approach was particularly conducive for understanding LML practices. Lee and Whang (2001), Chopra (2003), Boyer and Hult (2005), and

<table>
<thead>
<tr>
<th>Journal</th>
<th>Number of articles</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>International Journal of Physical Distribution and Logistics Management</td>
<td>9^a</td>
<td>19.15</td>
</tr>
<tr>
<td>International Journal of Retail and Distribution Management</td>
<td>7^b</td>
<td>14.89</td>
</tr>
<tr>
<td>Management Science</td>
<td>4</td>
<td>8.51</td>
</tr>
<tr>
<td>Journal of Operations Management</td>
<td>4</td>
<td>8.51</td>
</tr>
<tr>
<td>Transportation Science</td>
<td>4</td>
<td>8.51</td>
</tr>
<tr>
<td>European Journal of Operational Research</td>
<td>4</td>
<td>8.51</td>
</tr>
<tr>
<td>International Journal of Logistics Research and Applications</td>
<td>3</td>
<td>6.38</td>
</tr>
<tr>
<td>Transportation Research Part E</td>
<td>3</td>
<td>6.38</td>
</tr>
<tr>
<td>Journal of Business Logistics</td>
<td>2</td>
<td>4.26</td>
</tr>
<tr>
<td>MIT Sloan Management Review</td>
<td>2</td>
<td>4.26</td>
</tr>
<tr>
<td>Supply Chain Management: An International Journal</td>
<td>1</td>
<td>2.13</td>
</tr>
<tr>
<td>International Journal of Operations and Production Management</td>
<td>1</td>
<td>2.13</td>
</tr>
<tr>
<td>International Journal of Production Research</td>
<td>1</td>
<td>2.13</td>
</tr>
<tr>
<td>Planning and Production Control</td>
<td>1</td>
<td>2.13</td>
</tr>
<tr>
<td>Production and Operations Management Journal</td>
<td>0</td>
<td>0.00</td>
</tr>
<tr>
<td>Journal of Supply Chain Management</td>
<td>0</td>
<td>0.00</td>
</tr>
<tr>
<td>Mathematical Problems in Engineering</td>
<td>1</td>
<td>2.13</td>
</tr>
<tr>
<td>Total</td>
<td>47</td>
<td>100</td>
</tr>
</tbody>
</table>

Notes: ^aTwo articles from special issue on retail logistics (2016); ^bone article from special issue on post crisis in Europe (2016); ^cEurope- and US-based researchers (based on the first author’s affiliation) contributed 44.68 per cent (21 articles) and 38.30 per cent (18 articles) of the papers, respectively. The remaining articles were published by researchers based in China, Canada, Turkey, Australia, Brazil, and Singapore.
Vanelslander et al. (2013) each developed LML structural types to assist design under different consumer and product attributes. These studies mostly captured the linearly “chain-centric” LML models prevalent in the pre-digital era.

- Literature review and conceptual studies: several reviews have contributed in this domain. Some papers focused on specific areas, such as the evolution of British retailing (Fernie et al., 2010) and distribution network design (Mangiaracina et al., 2015), whereas others discussed several topics at once (Agatz et al., 2008; Lagorio et al., 2016; Savelsbergh and Van Woensel, 2016). Narrowly focused papers identified limited LML structural types or variables influencing distribution network design, while more broadly focused papers examined wider issues in urban, city, or multichannel logistics. Conceptual studies typically provided guidance on the selection of LML “types” based on certain performance criteria (e.g. Punakivi and Saranen, 2001; Chopra, 2003), or logistics service quality (e.g. Yuan and David, 2006).

- Empirical studies: these studies mainly compared LML types or demonstrated the impact of particular variables upon LML. Studies undertaking the former research purpose (contrasting types) employed simulations, field/mail surveys, and econometrics to examine performance or CO₂ emissions (e.g. Punakivi et al., 2001). One paper employed a mixed-method approach (case research and modelling) to understand the organisation of the physical distribution processes in omnichannel supply networks (Ishfaq et al., 2016). Empirical studies aiming at the latter research purpose (evidencing impact) used field and laboratory experiments and statistical methods on survey data to examine the interplay between operational strategies and consumer behaviour (e.g. Esper et al., 2003; Boyer and Hult, 2005; Kull et al., 2007). These studies also employed econometrics to examine the effects of cross-channel interventions (e.g. Forman et al., 2009; Gallino and Moreno, 2014). Additionally, a few studies used case research to provide operational guidance via framework development, such as last-mile order fulfilment (Hübner, Kuhn and Wollenburg, 2016) and LML design, to capture the interests of various stakeholders (Harrington et al., 2016).

- Mathematical modelling: studies also employed a variety of mathematical tools and techniques to formulate LML problems and find optimum solutions, mostly for vehicle routing problems (Campbell and Savelsbergh, 2005; McLeod et al., 2006; Aksen and Altinkemer, 2008; Crainic et al., 2009; Wang et al., 2014). In their work, Campbell and Savelsbergh (2006) combined optimisation modelling with simulation to demonstrate the value of incentives. Other studies focused on identifying optimum distribution strategies (e.g. Netessine and Rudi, 2006; Li et al., 2015), inventory rationing policy (Ayanso et al., 2006), delivery time slot pricing (Yang and Strauss, 2017), and formulating new models to capture emerging practices, such as crowd-sourced delivery (Wang et al., 2016).

**Thematic analysis**
The grounded theory approach (Glasser and Strauss, 1967) was used to code and classify emerging repeated concepts and terminologies via the qualitative data analysis software, MAXQDA version 12. The classification was based on the two categories defining LML models: LML distribution structures and their associated contingency variables. Coding of the data was conducted independently by two authors. The distinguishing terms and concepts were documented in a codebook; where their views differed, the issues were discussed until consensus was reached. Terminologies relating to each classification level were either derived from the extant literature or introduced in this paper to unify key concepts.
For the first category, the types of LML distribution structure are classified based on different levels of effort required by vendor and consumer: push-centric, pull-centric, and hybrid. A push-centric system requires the vendor to wholly undertake the distribution functions required to deliver the ordered product(s) to the consumer’s doorstep; a pull-centric system requires the consumer to wholly undertake the collection and transporting function; and a hybrid system requires some effort on the parts of both the vendor and consumer and is varied by the location of the decoupling point. A further breakdown divided the push-centric distribution system into modes of picking (manufacturer-based, distribution centre (DC)-based, and local brick-and-mortar (B&M) store-based); the pull-centric distribution system was divided into modes of collection from fulfilment point (local B&M store and information store); and the hybrid distribution system was divided into modes of CDP (attended collection delivery point (CDP-A) and unattended collection delivery point (CDP-U)).

The second category captures the associated contingency variables commonly used in existing studies. This study created a list of 13 variables that influence the structural forms of last-mile distribution: consumer geographical density, consumer physical convenience, consumer time convenience, demand volume, order response time, order visibility, product availability, product variety, product customisability, product freshness, product margin, product returnability, and service capacity. These variables determine the manner in which, or the efficiency with which, a distribution structure fulfils consumer needs while relating to the idiosyncrasies of product types.

These classifications facilitate the understanding of LML models and enabled future structural variables to be consistently categorised. Figure 1 serves to present a structural overview of the LML models reported in the literature.

**Review of LML distribution structures**

From the preceding section, an LML distribution structure involves the linear movement of a product from source to consumer. Technically, this can only have three basic forms:

1. **push**: product “sent” to consumer’s postcode by someone other than the consumer;
2. **pull**: product “fetched” from product source by the consumer; and
3. **hybrid**: product “sent” to an intermediate site, from which the product is “fetched” by the consumer.

Table II summarises the corpus on LML distribution structures.

**Push-centric system: n-tier direct to home**

This study found that the push-centric system is the most commonly adopted distribution form. It typically comprises a number of intermediate stages (n-tier) between the source and destination in order to create distribution efficiencies. The literature classifies three picking variants according to fulfilment (inventory) location: manufacturer-based (or “drop-shipping”), DC-based, or local B&M store-based (i.e. retailer’s intermediate warehouse or store). The destination can either be consumers’ homes or, increasingly, their workplaces (McKinnon and Tallam, 2003). The mode of delivery can be in-sourced (using retailer’s own vehicle fleet), outsourced to a third-party logistics provider (3PL) (Boyer and Hult, 2005), or crowd-sourced using independent contractors (Wang et al., 2016).

When selecting a distribution channel, retailers need to trade-off between fulfilment capabilities, inventory levels (Netessine and Rudi, 2006), product availability and variety (Agatz et al., 2008), transportation cost (Rabinovich et al., 2008), and responsiveness (Chopra, 2003). The nearer the picking site is to the consumer segment, the more responsive is the channel. However, this responsiveness comes at the expense of lower-level inventory aggregation and higher risks associated with stock-outs (Netessine and Rudi, 2006).
Figure 1. Classification of literature review on LML models

- **Structural Variables**
  - Product source
  - Geographical scope
  - Mode of distribution
  - Number of nodes

- **LML Model**
  - LML distribution structures

- **Contingency variables**
  - Consumer geographical density
  - Consumer physical convenience
  - Consumer time convenience
  - Demand volume
  - Order response time
  - Order visibility
  - Product availability
  - Product variety
  - Product customisability
  - Product freshness
  - Product margin
  - Product returnability
  - Service capacity

- **Push-centric system**
  - Manufacturer-based pick
  - Distribution centre (DC)-based pick
  - Local brick-and-mortar (B&M) store-based pick

- **Pull-centric system**
  - Local B&M store
  - Information store

- **Hybrid system**
  - Collection delivery point (CDP) (Attended)
  - Collection delivery point (CDP) (Unattended)
Pull-centric system: consumer self-help
The literature also discussed two variants of the pull-centric system. Both variants require consumers to participate (or self-help) throughout the transaction process, from order fulfilment to order transportation. The first variant represents the traditional way of shopping at a local B&M store, with consumers performing the last-mile “delivery”. The second “information store” variant adopts a concept known as “dematerialisation”, substituting information flow for material flow (Lee and Whang, 2001). This variant recognises that material or physical flows are typically more expensive than information flows due to the costs of (un)loading, handling, warehousing, shipping, and product returns.

This study found that despite the popularity of online shopping, there are still occasions where consumers favour traditional offline shopping. Perceived or actual difficulties with inspecting non-digital products, the product returns process, or slow and expensive shipping can deter consumers from online shopping (Forman et al., 2009). This study also demonstrates other benefits of a pull-centric system, including lower capital investments and possible carry-over (or halo) effects into in-store sales (Johnson and Whang, 2002).

### Table II. Types of LML distribution structure

<table>
<thead>
<tr>
<th>Level 1 classification</th>
<th>Level 2 classification</th>
<th>References</th>
</tr>
</thead>
</table>

Notes: GRP, goods reception point; RB, reception box (independent/with docking mechanism/shared); ULB, urban logistics boxes
Hybrid system: n-tier to consumer self-help location

The rich literature here mainly compared different modes of reception. Variants typically entailed a part-push and part-pull configuration. For instance, the problem associated with “not-at-home” responses within attended home delivery (AHD) can be mitigated by delivering the product to a CDP for consumers to pick up. The literature discussed two CDP variants: CDP-A and CDP-U. It found that retailers establish CDP-A through developing new infrastructure development, through utilising existing facilities, or establishing partnerships with a third party (Wang et al., 2014). Other terminologies associated with CDP-A include “click-and-collect”, “pickup centre”, “click-and-mortar”, and “buy-online-pickup-in-store”. The literature showed that retailers establish CDP-U (or unattended reception) through independent RBs equipped with a docking mechanism, or shared RBs, whose locations range from private homes to public sites (e.g. petrol kiosks and train stations) accessible by multiple users (McLeod et al., 2006).

These CDP-A and CDP-U strategies are commonly adopted by multi/omnichannel retailers to exploit their existing store networks, to provide convenience to consumers through ancillary delivery services, and to expedite returns handling (Yrjolä, 2001). Moreover, the research showed that integrating online technologies with physical infrastructures enables retailers to achieve synergies in cost savings, improved brand differentiation, enhanced consumer trust, and market extension (Fernie et al., 2010). Studies have also investigated the cost advantage and operational efficiencies of using CDP-U over AHD and CDP-A (e.g. Wang et al., 2014). CDP-U reduces home delivery costs by up to 60 per cent (Punakivi et al., 2001), primarily by exploiting time window benefit (Kämäräinen et al., 2001).

Development of LML design framework

This section addresses the second and third research questions by developing a framework that contributes to LML design practice. The development process is governed by contingency theory (Lawrence and Lorsch, 1967), in which “fit” is a central concept. The contingency theory maintains that structural, contextual, and environmental variables should fit with one another to produce organisational effectiveness. The management literature conceptualises fit as profile deviation (e.g. Jauch and Osborn, 1981; Doty et al., 1993) in terms of the degree of consistency across multiple dimensions of organisational design and context. The probability of organisational effectiveness increases as the fit between the different types of variables increases (Jauch and Osborn, 1981; Doty et al., 1993). In this paper, the environmental and contextual variables are jointly branded as contingency variables since the object was to examine how these variables impact the structural form of LML distribution.

We developed the LML design framework in two steps. First, we synthesised a set of LML structural and contingency variables and established the relationship between these through a review of the LML literature. Second, we reformulated the descriptive (i.e. science-mode) knowledge obtained via the first step into prescriptive (i.e. design-mode) knowledge. We adopted the contingency perspective in combination with Romme’s (2003) approach to inform knowledge reformulation.

Synthesising LML structural variables

The three dominant system dynamics identified (push-centric, pull-centric, and hybrid), reflect variations in LML distribution structures. By comparing these structures, we conceptualised a set of structural variables, whose variations can explain the structural forms of last-mile distribution. These structural variables consist of product source, geographical scope, mode of distribution, and number of nodes. These variables are expanded below:

- Product source refers to the location where products are stored when an order is accepted; it coincides with the start point of an LML network. It can be contextualised
as a supply network member entity (manufacturer, distributor, or retailer). To illustrate, the computer manufacturer Dell (customisation services), online grocer Ocado (home delivery services), and the UK’s leading supermarket chain Tesco (click-and-collect services) source their products from manufacturer, distributor, and retailer sites, respectively.

- Geographical scope concerns the distance separating the start point (product source) and the end point (final consignee’s preferred destination point) of an LML network. An LML network can be classified as centrally based (e.g. Dell Services) or locally based (e.g. Tesco’s click-and-collect).

- Mode of distribution describes the delivery mode from the point where an order is fully fulfilled to the end point; it can be classified into three types: self-delivery (e.g. Tesco’s self-owned fleet for home deliveries), 3PL delivery including crowdsourcing (e.g. Dell Services), and consumer-pickup (e.g. Tesco’s click-and-collect services).

- Number of nodes concerns the operations in which products are “stationary”, residing in a facility for processing or storage. As opposed to nodes, links represent movements between nodes. There are two variations in respect to this variable: two-node and multiple-node. For example, a two-node structure can be found in Dell’s direct-to-consumer distribution channel, where computers are assembled and orders fulfilled at the factory prior to direct home delivery. In contrast, multiple-node structures are reflected in “in-transit merge” structure where an order comprising components sourced from multiple locations are assembled at a common node. As a case in point, when consumer order a computer processing unit (CPU) from Dell along with a Sony monitor, a parcel carrier would pick up the CPU from a Dell factory and the monitor from a Sony factory, then would merge the two into a single shipment at a hub prior to delivery (Chopra, 2003).

**Synthesising LML contingency variables**
The contingency variables exist in two modes: systematic development of LML contingency variables; and empirical analysis of particular variables or their interrelationships with other variables. To illustrate the former, Chopra (2003), Boyer and Hult (2005), and Gevaers et al. (2011) each produced their own lists of contingency variables. Regarding the latter, Boyer et al. (2009), for example, performed a simulation analysis of two independent variables (consumer density and delivery window size) and one dependent variable (delivery cost). The most common contingency-reported variables to have a direct impact on the structural form of LML distribution are listed below and commentary on them are summarised in Table III:

- Consumer geographical density: the number of consumers per unit area (Boyer and Hult, 2005; Boyer et al., 2009; Mangiaracina et al., 2015).

- Consumer physical convenience: the effort consumers exert to receive orders (Chopra, 2003; Harrington et al., 2016).

- Consumer time convenience: the time committed by consumers for the reception of orders. This variable fluctuates according to the structural form of last-mile distribution (Rabinovich and Bailey, 2004; Yuan and David, 2006).

- Demand volume: the number of products ordered by consumers relative to the distribution structure (Chopra, 2003; Boyer and Hult, 2005).

- Order response time: the time difference between order placement and order delivery (Kämäräinen et al., 2001; Mangiaracina et al., 2015).
Consumer geographical density is closely related to modes of delivery. When consumer geographical density is high, a merchant's own fleet can be effectively utilised; when consumer geographical density is low, companies use 3PLs to consolidate orders to increase fleet utilisation. Consumer physical convenience is higher for LML models using direct-to-home delivery since almost no consumer physical movement is needed. By contrast, consumer physical convenience is lower for models featuring consumer-pickup, since these models require consumers to physically travel to the pickup sites. However, physical inconvenience for order pickup may be offset under situations where a consumer's commute route overlaps with the pickup route. Consumer-pickup models are most suited for consumers with higher availability of time for order pickup; with direct-to-home delivery preferred for time-poor consumers; models offering longer time windows for AHD are suitable for consumers with time flexibility. High demand products are typically sourced locally and in closer proximity to the consumer to reduce transport costs and fast response times. By contrast, low demand products are stored centrally to gain benefits of aggregation and concomitantly, to allow for extension of geographical coverage. It is possible for low demand and high margin products to be sourced locally, since the benefit gained through higher margins may potentially offset the loss incurred through localised storage structures. Short response time can be achieved by reducing transport time by localising operations. Furthermore, response time increases as the number of additional nodes in an LML network increases. As products are being processed at a node, their movement discontinues. Models based on in-transit merge lead to a longer response time in comparison with, for example, a model that features uninterrupted drop-shipping. Visibility is easiest when an LML model is simplified as merely two connected nodes, with one being the product source point and the other being the final consignee's preferred destination point. For instance, an LML model featured with distributor source and 3PL delivery can be simplified as a two-node network comprised of a distributor and a consumer's home. This two-node network has the potential to provide the highest order visibility since it is necessary for information systems of all nodes to be integrated. In a multiple-node with more complex network, order visibility decreases. Moreover, for LML models with consumer-pickup to be operational, high visibility via advance consumer notification must be provided. These variables share a single variation pattern. A manufacturer holds an inventory of greater variety and quantity than a distributor does, since the former occupies a more central position in an LML network and serves a wider consumer base. Correspondingly, a distributor has a greater variety of inventory than a retailer. Therefore, manufacturers, distributors, and retailers, respectively, provide product variety and product availability in descending order. Moreover, LML models based on 3PL delivery provide greater variety given the specialist role of 3PL, which is capable of carrying more differentiated products. Because customisation processes typically involve physical alteration of product and/or packaging, the product manufacturer, by default, has the highest control over customisation; with the distributor and retailer having decreasing amounts of customisation control (e.g. Dell’s ability to customise-to-order). This variable is governed by two factors: first, the shipping distance separating the manufacturer and consumer’s home, and second, the shipping “directness” (whether the product is shipped to the destination point directly or through one or more “stops” for additional processing). The longer the shipping distance or the higher the shipping indirectness, the lower the level of product freshness. LML models with a central product sourcing point, such as the in-transit merge and drop-shipping models, are more compatible with higher margin products because the benefits of inventory aggregation are high and consumers are more willing to tolerate a longer response time. By comparison, LML models with local product sourcing points

<table>
<thead>
<tr>
<th>Variable</th>
<th>Commentary and discussion on contingency variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consumer geographical density</td>
<td>Consumer geographical density is closely related to modes of delivery. When consumer geographical density is high, a merchant's own fleet can be effectively utilised; when consumer geographical density is low, companies use 3PLs to consolidate orders to increase fleet utilisation.</td>
</tr>
<tr>
<td>Consumer physical convenience</td>
<td>Consumer physical convenience is higher for LML models using direct-to-home delivery since almost no consumer physical movement is needed. By contrast, consumer physical convenience is lower for models featuring consumer-pickup, since these models require consumers to physically travel to the pickup sites. However, physical inconvenience for order pickup may be offset under situations where a consumer's commute route overlaps with the pickup route.</td>
</tr>
<tr>
<td>Consumer time convenience</td>
<td>Consumer-pickup models are most suited for consumers with higher availability of time for order pickup; with direct-to-home delivery preferred for time-poor consumers; models offering longer time windows for AHD are suitable for consumers with time flexibility.</td>
</tr>
<tr>
<td>Demand volume</td>
<td>High demand products are typically sourced locally and in closer proximity to the consumer to reduce transport costs and fast response times. By contrast, low demand products are stored centrally to gain benefits of aggregation and concomitantly, to allow for extension of geographical coverage. It is possible for low demand and high margin products to be sourced locally, since the benefit gained through higher margins may potentially offset the loss incurred through localised storage structures.</td>
</tr>
<tr>
<td>Order response time</td>
<td>Short response time can be achieved by reducing transport time by localising operations. Furthermore, response time increases as the number of additional nodes in an LML network increases. As products are being processed at a node, their movement discontinues. Models based on in-transit merge lead to a longer response time in comparison with, for example, a model that features uninterrupted drop-shipping.</td>
</tr>
<tr>
<td>Order visibility</td>
<td>Visibility is easiest when an LML model is simplified as merely two connected nodes, with one being the product source point and the other being the final consignee's preferred destination point. For instance, an LML model featured with distributor source and 3PL delivery can be simplified as a two-node network comprised of a distributor and a consumer's home. This two-node network has the potential to provide the highest order visibility since it is necessary for information systems of all nodes to be integrated. In a multiple-node with more complex network, order visibility decreases. Moreover, for LML models with consumer-pickup to be operational, high visibility via advance consumer notification must be provided.</td>
</tr>
<tr>
<td>Product availability and product variety</td>
<td>These variables share a single variation pattern. A manufacturer holds an inventory of greater variety and quantity than a distributor does, since the former occupies a more central position in an LML network and serves a wider consumer base. Correspondingly, a distributor has a greater variety of inventory than a retailer. Therefore, manufacturers, distributors, and retailers, respectively, provide product variety and product availability in descending order. Moreover, LML models based on 3PL delivery provide greater variety given the specialist role of 3PL, which is capable of carrying more differentiated products.</td>
</tr>
<tr>
<td>Product customisability</td>
<td>Because customisation processes typically involve physical alteration of product and/or packaging, the product manufacturer, by default, has the highest control over customisation; with the distributor and retailer having decreasing amounts of customisation control (e.g. Dell’s ability to customise-to-order).</td>
</tr>
<tr>
<td>Product freshness</td>
<td>This variable is governed by two factors: first, the shipping distance separating the manufacturer and consumer’s home, and second, the shipping “directness” (whether the product is shipped to the destination point directly or through one or more “stops” for additional processing). The longer the shipping distance or the higher the shipping indirectness, the lower the level of product freshness.</td>
</tr>
<tr>
<td>Product margin</td>
<td>LML models with a central product sourcing point, such as the in-transit merge and drop-shipping models, are more compatible with higher margin products because the benefits of inventory aggregation are high and consumers are more willing to tolerate a longer response time. By comparison, LML models with local product sourcing points</td>
</tr>
</tbody>
</table>
Order visibility: the ability of consumers to track their order from placement to delivery (Chopra, 2003; Harrington et al., 2016).

Product availability and product variety: product availability is the probability of having products in stock when a consumer order arrives (Chopra, 2003; Yuan and David, 2006).

Product variety is the number of unique products (or stock keeping units) offered to consumers (Punakivi et al., 2001; Punakivi and Saranen, 2001).

Product customisability: the ability for products to be adapted to consumer specifications (Boyer and Hult, 2005).

Product freshness: the time elapsed from the moment a product is fully manufactured to the moment when it arrives at the consumption point (Boyer and Hult, 2005).

Product margin: the net income divided by revenue (Boyer and Hult, 2005; Campbell and Savelsbergh, 2005).

Product returnability: the ease with which consumers can return unsatisfactory products (Chopra, 2003; Yuan and David, 2006).

Service capacity: the ability of an LML system to provide the intended delivery service and to match consumer demand at any given point in time (Rabinovich and Bailey, 2004; Yuan and David, 2006).

### Synthesising the relationship between LML structural and contingency variables
Following the identification of the structural and contingency variables, we draw a connection between structural and contingency variables and the relationships between them. This connection is achieved by capturing the differences between these variables and establishing the correlations between particular variations of structural variables and those of contingency variables. For example, Chopra (2003, p. 125) stated:

Firms that target customers who can tolerate a large response time require few locations that may be far from the customer and can focus on increasing the capacity of each location. On the other hand, firms that target customers who value short response times need to locate close to them.

This statement identifies the association between a structural variable, namely “geographical scope”, and a contingency variable, namely “order response time”. Within the literature,
two variations emerged for each variable: centralised vs localised network for geographical scope and long vs short delivery period for order response time; i.e. centralised geographical scope corresponds to long response time, while localised scope is more responsive. As such, the findings demonstrate that by identifying connecting rationales and the variations at different levels for each variable, we can capture correlations between two sets of variables (i.e. structural and contingency). Continuing this procedure across relevant statements found in our corpus, Table IV summarises the outputs.

**Reformulation from science-mode into design-mode knowledge**

We adopted the approach by Romme (2003) to reformulate the descriptive knowledge (i.e. science-mode, developed in the previous section) into prescriptive (i.e. design-mode) knowledge so that the latter becomes more accessible to guide practitioners in their LML design thinking. This approach has previously been used to contextualise various design scenarios (e.g. Zott and Amit, 2007; Holloway et al., 2016; Busse et al., 2017). For example, Busse et al. (2017) employed a variant of the approach to investigate how buying firms facing low supply chain visibility can utilise their stakeholder network to identify salient supply chain sustainability risks.

<table>
<thead>
<tr>
<th>Structural variable</th>
<th>Geographical scope</th>
</tr>
</thead>
<tbody>
<tr>
<td>Product source</td>
<td>Locally based</td>
</tr>
<tr>
<td>Manufacturer</td>
<td>n/a</td>
</tr>
<tr>
<td>Distributor</td>
<td>n/a</td>
</tr>
<tr>
<td>Retailer</td>
<td>n/a</td>
</tr>
<tr>
<td>Centrally based</td>
<td>n/a</td>
</tr>
<tr>
<td>Consumer geographical density</td>
<td>n/a</td>
</tr>
<tr>
<td>Consumer physical convenience</td>
<td>n/a</td>
</tr>
<tr>
<td>Consumer time convenience</td>
<td>n/a</td>
</tr>
<tr>
<td>Demand volume</td>
<td>Low</td>
</tr>
<tr>
<td>Order response time</td>
<td>n/a</td>
</tr>
<tr>
<td>Order visibility</td>
<td>n/a</td>
</tr>
<tr>
<td>Product availability</td>
<td>High</td>
</tr>
<tr>
<td>Product variety</td>
<td>High</td>
</tr>
<tr>
<td>Product customisability</td>
<td>High</td>
</tr>
<tr>
<td>Product freshness</td>
<td>n/a</td>
</tr>
<tr>
<td>Product margin</td>
<td>High</td>
</tr>
<tr>
<td>Product returnability</td>
<td>n/a</td>
</tr>
<tr>
<td>Service capacity</td>
<td>n/a</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Contingency variable</th>
<th>Product source</th>
<th>Mode of distribution</th>
<th>Number of nodes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Structural variable</td>
<td>3PL delivery</td>
<td>Two-node</td>
</tr>
<tr>
<td>Consumer geographical density</td>
<td>High</td>
<td>n/a</td>
<td>Low</td>
</tr>
<tr>
<td>Consumer physical convenience</td>
<td>High</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Consumer time convenience</td>
<td>High</td>
<td>Low</td>
<td>High</td>
</tr>
<tr>
<td>Demand volume</td>
<td>n/a</td>
<td>n/a</td>
<td>–</td>
</tr>
<tr>
<td>Order response time</td>
<td>n/a</td>
<td>n/a</td>
<td>–</td>
</tr>
<tr>
<td>Order visibility</td>
<td>n/a</td>
<td>n/a</td>
<td>–</td>
</tr>
<tr>
<td>Product availability</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>Product variety</td>
<td>n/a</td>
<td>n/a</td>
<td>–</td>
</tr>
<tr>
<td>Product customisability</td>
<td>n/a</td>
<td>n/a</td>
<td>–</td>
</tr>
<tr>
<td>Product freshness</td>
<td>n/a</td>
<td>n/a</td>
<td>–</td>
</tr>
<tr>
<td>Product margin</td>
<td>Low</td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>Service capacity</td>
<td>Full</td>
<td>n/a</td>
<td>Fluctuated</td>
</tr>
<tr>
<td>Consumer time convenience</td>
<td>High</td>
<td>Low</td>
<td>High</td>
</tr>
</tbody>
</table>

Table IV. LML design framework

**Notes:** These four cells are interrelated. The two combinations, namely “high margin and low volume” and “low margin and high volume”, can both lead to sufficient incentives for a localised operation and a retailer-sourcing strategy. Scale of rating – fast, medium, slow; low, medium, high; and full, fluctuated.
Romme (2003) made a distinction between science and design modes of engaging in organisational research. The former is representational, descriptive, and analytic and it concerns the study of variances among variables over time and/or space. The latter is pragmatic, actionable, and exists in a more usable form in order to facilitate changing incumbent organisational systems and situations into the desired states. Furthermore, Romme (2003) proposed a framework exploiting synergies between science and design modes. He argued, suppose there is a science-mode proposition “if \( x \), then \( y \)” and a design-mode proposition “to achieve \( O \), do \( A \)”, to connect both modes, the former can be reformulated to the latter as follows:

1. if necessary, redefine descriptive (properties of) variables into imperative ones (e.g. actions to be taken);
2. redefine the probabilistic nature of a hypothesis into an action-oriented design proposition;
3. add any missing context-specific conditions and variables (drawing on other research findings obtained in science- or design-mode); and
4. in case of any interdependencies between hypotheses/propositions, formulate a set of propositions.

As an example, take the following science-mode statement:

“If order response time delivered by an LML network is short, then the geographical scope of the LML network should be localised.”

We can reformulate this statement into the following design-mode statement:

“For an LML network to achieve short order response time, localise the geographical scope.”

Following similar procedures, the science-mode knowledge describing the relationships between structural and contingency variables can be reformulated to the design-mode shown in Table V. Collectively, the resulting design-mode knowledge constitutes a set of design guidelines for LML practitioners.

**Main research issues, gaps, and future lines of research**

Although the literature covered in this study thoroughly addresses LML structures, the extant literature has limitations. Based on this study’s findings, there are four main areas that require future study.

**Operational challenges in executing last-mile operations**

The extant literature has focused on the planning aspect of LML, rather than exploring operational challenges. Consequently, research often takes a simplistic chain-level perspective of LML in order to develop simplistic design prescriptions for practitioners. While this approach seems suitable in the pre-digital era, it is inadequate to capture the complexities of last-mile operations in the omnichannel environment (Lim et al., 2017). The focus on LML nodes as solely unifunctional is also inadequate (Vanelslander et al., 2013). Not acknowledging the multi-functionality of individual nodes limits understanding of how this variant works.

To address the limitations of extant research, we propose extending the current research from adding linear point-to-point LML “chains” (e.g. Chopra, 2003; Boyer and Hult, 2005) to also addressing the “networks” context, where multiple chains are intertwined and more widely practised in the industry. A study of LML systems using 3PL shared by multiple companies is an example of necessary future research. We also recommend future research to address the multi-functionality of individual nodes in an LML system. A study that addresses the ability of an LML node to simultaneously be a manufacturer and a distributor introduces more structural variance and needs to be theoretically addressed.
<table>
<thead>
<tr>
<th>Structural variable</th>
<th>Science-mode knowledge: <em>if x, then y</em></th>
<th>Design mode knowledge: <em>to achieve O, do A</em></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Product source</strong></td>
<td>If (a particular conceptual variable) delivered by an LML network is (a particular contingency variation), then the product of the LML network should be sourced from (a particular structural variation)</td>
<td>If (product customizability) delivered by an LML network is (high), then the product of the LML network should be sourced from (manufacturer)</td>
</tr>
<tr>
<td></td>
<td>General formula</td>
<td>Example</td>
</tr>
<tr>
<td><strong>Geographical scope</strong></td>
<td>If (a particular contingency variable) delivered by an LML network is (a particular contingency variation), then the geographical scope of the LML network should be (a particular structural variation)</td>
<td>If (order response time) delivered by an LML network is (short), then the geographical scope of the LML network should be (localised)</td>
</tr>
<tr>
<td><strong>Mode of distribution</strong></td>
<td>If (a particular conceptual variable) delivered by an LML network is (a particular variation of the contingency variable), then the mode of distribution should be (a particular variation of the structural variable)</td>
<td>If (product variety) delivered by an LML network is (high), then the mode of distribution should be (3PL delivery)</td>
</tr>
<tr>
<td><strong>Number of nodes</strong></td>
<td>If (a particular conceptual variable) delivered by an LML network is (a particular contingency variation), then the number of nodes should be (a particular structural variation)</td>
<td>If (order response time) delivered by an LML network is (short), then the number of nodes present in the LML network should be (two)</td>
</tr>
</tbody>
</table>
Additionally, existing literature typically focuses on comparing structural variants’ performance outcomes and their corresponding consumer and product attributes. However, we argue that such focus limits our understanding of how LML distribution structures interact as part of the broader omnichannel system. Accordingly, an avenue for future research would employ configuration perspectives (Miller, 1986; Lim and Srai, 2018) to complement the traditional reductionist approaches (e.g. Boyer et al., 2009) in order to more holistically examine LML models. Future studies could consider the structural interactions with relational governance of supply network entities, in order to promote information sharing and enhancing visibility, which are critical in omnichannel retailing (Lim et al., 2016).

Finally, while recent articles have started to examine the effects of online and offline channel integrations (e.g. Gallino and Moreno, 2014), limited contributions have been made to date to understand how retailers integrate their online and offline operations and resources to deliver a seamless experience for consumers (Piotrowicz and Cuthbertson, 2014; Hübner, Kuhn and Wollenburg, 2016). We propose revisiting the pull-centric system variants in the context of active consumer participation to understand the approaches retailers can use to attract consumers to their stores. In this regard, the subject can benefit from insightful case studies to advance our understanding of the challenges retailers face, as well as the operational processes retailers adopt to meet these challenges.

Intersection between last-mile operations and “sharing economy” models
With the exception of one paper (Wang et al., 2016), the majority of the extant literature discusses conventional LML models. Given the rapidly growing sharing economy that generates innovative business models (e.g. Airbnb, Uber, Amazon Prime Now) in several sectors (e.g. housing, transportation, and logistics, respectively) and exploits collaborative consumption (Hamari et al., 2016, p. 2047) and logistics (Savelsbergh and Van Woensel, 2016), there is an immense research scope at the intersection between LML and sharing economy models. First, we propose empirical studies to examine how retailers can effectively employ crowdsourcing models for the last-mile and to show how they can effectively integrate these models into their existing last-mile operations, such as combining in-store fulfilment through delivery using “Uber-type” solutions. This type of study is critical for understanding the impact of crowdsourcing models on retail operations and for promoting their adoption. Second, papers addressing omnichannel issues (Hübner, Kuhn and Wollenburg, 2016; Hübner, Wollenburg and Holzapfel, 2016; Ishfaq et al., 2016) are emerging. The emergence of new omnichannel distribution models demands theoretical development and the identification of new design variables. These models include on-demand delivery model (e.g. Instacart), distribution-as-a-service (e.g. Amazon, Ocado), “showroom” concept stores (e.g. Bonobos.com, Warby Parker), in-store digital walls (e.g. Adidas U.S. adiVerse), unmanned delivery (e.g. drones, ground robots), and additive printing (e.g. The UPS store 3D print). Increasingly, we also observe the growing convergence of roles and functions between online and traditional B&M retailers, which suggests new integrated LML models. These new roles and functions demand future research. Finally, while collaborative logistics enable the sharing of assets and capacities in order to increase utilisation and reduce freight, its success rests on developing a logistics ecosystem of relevant stakeholders (including institutions). Consequently, exciting research opportunities exist to explore new design variables that capture key stakeholders’ interests at various levels (Harrington et al., 2016).

Data harmonisation and analytics: collection and sharing platforms
The literature review revealed that, to date, there has been a tendency towards geographical-based studies and the use of simulated data. For example, this review
reports studies based in Finland (Punakivi and Saranen, 2001), Scotland (McKinnon and Tallam, 2003), the USA (Boyer et al., 2009), England (McLeod et al., 2006), Germany (Wollenburg et al., 2017), and Brazil (Wanke, 2012), amongst others. While these studies contribute to generating a useful library of contexts, they are difficult to compare, given differences in geography and geographically based data collection and analysis methods. Moreover, the majority of the studies in this review (41.30 per cent) were based on modelling and simulated data with limited application to real-world data sets, which might suggest a lack of quality data sets. Simulated data limit the advancement of domain knowledge, thus the development of real-world data sets could significantly fuel progress. As such, more attention should be focused on developing data sets, e.g. through the use of transaction and consumer-level data, to gain insights into last-mile behaviours and to design more effective LML models.

Additionally, future studies should standardise data collection in order to address current trends in urbanisation and omnichannel retailing, which are changing retail landscapes and consumer shopping behaviours. This study recommends establishing a data collection framework to guide scholars in LML design, with scholars developing new competences in data mining analytics to exploit large-scale data sets.

Moving from prescriptive to predictive last-mile distribution design

Extant studies have derived correlations between variation of independent variables (e.g. order response time) and variation of dependent variables (e.g. degree of centralisation) to provide prescriptive solutions to the design of last-mile distribution structures. However, these relationships (both linear and non-linear) are often confounded by other factors due to the real-world complexities and they inherently face multicollinearity and endogeneity issues, including the omitted variable bias problem, which leads to biased conclusions. Moreover, model complexity increases as more variables are included, potentially causing overfitting. Given these complexities, researchers usually find immense challenges in untangling these relationships. In this regard, we offer several valuable future lines of research leveraging more advanced techniques for the design of last-mile distribution.

First, our review captured 13 contingency variables that influence the design of last-mile distribution. Future research could discuss other contingency variables and investigate the use of statistical machine and deep learning techniques to identify the most critical contingency variables and uncover hidden relationships to develop predictive models. Second, as urbanisation trends continue, more institutional attention is required on urban logistics focused on negative externalities (congestion and carbon emissions) driven by the intensification of urban freight. According to our review, there is insufficient attention paid to urban freight delivery, and we propose exploring archetyping of urban areas for the development of predictive models to guide the design of urban last-mile distribution systems.

Third, the developed design framework is based on the assumption that only one last-mile distribution structure may be adopted for a given scenario. As we observed in the omnichannel setting, it is common for retailers to concurrently operate multiple distribution structures. The interrelationships between the various structural combinations under the management of a single LML operator also present a potential future research direction.

Last, there is room for a combination of methods to more appropriately tackle the increasingly complicated and fragmented distribution networks in the omnichannel environment. Indeed, this research revealed only two papers in the corpus that have employed a mixed-method approach. Ishfaq et al. (2016) used case research and classification-tree analysis to understand the organisation of distribution processes in
omnichannel supply networks, while Campbell and Savelsbergh (2006) combined analytical modelling with simulation to demonstrate the value of incentives in influencing consumer behaviour to reduce delivery costs.

**Conclusions**

This paper offers the first comprehensive review and analysis of literature regarding e-commerce LML distribution structures and their associated contingency variables. Specifically, the study offers value by using a design framework to explicate the relationship between a broad set of contingency variables and the operational characteristics of LML configuration via a set of structural variables with clearly defined boundaries. The connection between contingency variables and structural variables is critical for understanding LML configuration choices; without understanding this connection, extant knowledge is non-actionable, leaving practitioners with an overwhelming number of seemingly relevant variables that have vague relationships with the structural forms of last-mile distribution.

From a theoretical contribution perspective, this paper identifies attributes of delivery performance linked to product-market segments and the system dynamics that underpin them. This understanding of the interrelationships between LML dimensions enables us to classify prior work, which is somewhat fragmented, to provide insights on emerging business models. The reclassification of LML structures helps practitioners understand the three dominant system dynamics (push-centric, pull-centric, and hybrid) and their related contingency variables. Synthesising structural and contingency variables, the network design framework (Table IV) sets out the connections, which when reformulated (Table V), provide practitioners design prescriptions under varying LML contexts.

Accordingly, the literature review demonstrates that push-centric LML models driven by order visibility performance are ideally suited to variety-seeking market segments where consumers prioritise time convenience over physical convenience. Conversely, it shows that pull-centric LML models favour order response time, order visibility, and product returnability performance, which are widely observed in markets where consumers desire high physical convenience, low product customisability, and high product variety. Most interestingly of all, this study explains the emergent hybrid systems, where service capacity performance excellence is delivered through multiple clusters of contingency variables, which suits availability-sensitive markets and markets where consumers prioritise physical (over time) convenience.

This paper identifies four areas for further research: operational challenges in executing last-mile operations; intersection between last-mile operations and sharing economy models; data harmonisation and analytics; and moving from prescriptive to predictive last-mile distribution design. Research in these areas could contribute to consolidating the body of knowledge on LML models while maintaining the essential multidisciplinary character. We hope that this review will serve as a foundation to current research efforts, stimulate suggested lines of future research, and assist practitioners to design enhanced LML models in a changing digital e-commerce landscape.

**Acknowledgements**

The authors would like to acknowledge the support from industry collaborators in the validation activities regarding the design framework developed in this paper. Research programmes funded by the UK Engineering and Physical Sciences Research Council (Grant: EP/K02888X/1 – Engineering Driven Sustainable Supply Networks), UK Advanced Manufacturing Supply Chain Initiative (Grant: 35708-233529, ReMediES – Reconfiguring Medicines End-to-end Supply), and A*STAR Singapore have supported these collaborative
multinational partner engagements. The authors specifically acknowledge those programmes involving fast-moving consumer goods multinationals and their retailer networks, as well as those with pharmaceutical manufacturers, their 3PLs, and their final-delivery pharmacy chains.

References


Consumer-driven e-commerce


**Corresponding author**
Stanley Frederick W.T. Lim can be contacted at: wtsfl2@cam.ac.uk

For instructions on how to order reprints of this article, please visit our website: www.emeraldgrouppublishing.com/licensing/reprints.htm
Or contact us for further details: permissions@emeraldinsight.com