Accumulation of wear and tear in archival and library collections. Part II: an epidemiological study

Cristina Duran-Casablancas* (1,2), Josep Grau-Bové (1), Tom Fearn (3), Matija Strlič (1)

(1) Institute for Sustainable Heritage, University College London, London, UK

(2) Amsterdam City Archives, Amsterdam, The Netherlands

(3) Department of Statistical Science, University College London, London, UK

*Corresponding author

Cristina Duran-Casablancas: cristina.casablancas.14@ucl.ac.uk, c.duran@amsterdam.nl

Josep Grau-Bové: josep.grau.bove@ucl.ac.uk

Tom Fearn: t.fearn@ucl.ac.uk

Matija Strlič: m.strlic@ucl.ac.uk

Abstract

This paper proposes a new approach to collection surveying based on epidemiology, the discipline that describes and explains disease patterns in populations. In epidemiology the focus of attention lies not only on the occurrence of a disease but also on the characteristics of the individuals which might play a role in the occurrence of the disease. To explore the applicability of epidemiology to heritage collections, we take as example the study of the occurrence and accumulation of wear and tear in archive collections, which so far has only been studied in controlled experimental studies. We designed an observational study (survey) in which the assessment of mechanical failure is understood as the outcome variable, and the factors that might affect the degree of failure are defined as exposure variables. To evaluate the relevance of the assessed factors in relation to the observed mechanical failure, exploratory data analyses were conducted by comparing groups of objects that differ regarding their level of exposure to different

factors. Although highly scattered data is not unusual in this type of studies and confounding has to be taken into account during the data analysis, this paper shows that through an epidemiological approach to surveys, the factors that have a greater effect on mechanical failure can be identified. Moreover, the rate of failure can also be determined for certain groups of objects. Also patterns of decay emerge which show the vulnerability of certain groups of objects. In this paper the practical aspects of the design and analysis of observational epidemiological studies for heritage collections are discussed. As a final note, the applicability and relevance of this approach to support collection management is briefly discussed.

Keywords

Wear and tear, mechanical degradation, archival and library collections, survey, reliability, epidemiology

1. Introduction

An intrinsic characteristic of library and archive collections is that they are physically used, which may lead to mechanical degradation, understood as wear and tear (1). The relation between chemical degradation and the deterioration of mechanical properties of the paper has been extensively explored in experimental studies (2,3). However, for a better understanding of the role of the chemical and mechanical properties of paper on the occurrence of wear and tear, these properties should not be studied in isolation but in conjunction with the frequency of physical use. Recently, an experimental study described how mechanical degradation accumulates at a given average frequency of use as a function of time (wear-out function), depending on the cellulose degree of polymerisation, DP (4). In addition, a damage function has been developed which combines the wear-out function with a dose response function based on input variables such as temperature, relative humidity and pH (5). This damage function, which takes

physical use into account, has been developed for paper collections. Damage functions for archive and library objects made of materials other than paper, such as bindings, have not been developed yet.

The wear-out function for paper collections is based on data obtained in controlled experiments, which only partly reproduce the conditions in which original objects are kept and used. Hence, the analysis of data collected in real case studies remains essential not only to validate the experimental model, but also to evaluate whether other factors should be taken into account for the full understanding of the mechanical degradation process. For instance, the classification of failure causes used in reliability engineering has been suggested as a method to list the factors contributing to mechanical degradation (6). This classification divides the causes of failure into factors prior to use (design and manufacture) and factors during use (maintenance and usage). In order to explore the relevance of these factors we propose a new study of heritage collections based on epidemiology.

Epidemiology is "the science and practice that describes and explains disease patterns in populations"(7). With this aim, characteristics of individuals are recorded and then analysed by comparing groups of individuals within a population. Observational studies are designed to explore associations between the outcome variable (the event of interest) and the exposure variables (the factors that might affect the degree of the outcome variable). As the researcher has no control over the characteristics of the individuals included in this type of study, the results of observational studies are usually used to develop hypotheses, which need to be further explored in experimental studies.

Epidemiology uses different study designs. Longitudinal studies, in which individuals are followed during a certain time interval, provide a higher level of evidence than studies in which the event of interest is assessed at a certain point in the time. However, longitudinal studies are less suitable for the study of wear and tear, as follow-up studies

need significant resources to assess objects repeatedly: the rate of the accumulation of wear and tear is expected to be low and therefore more difficult to detect in short time intervals.

Another type of study, case-control, collects data at one single point in the time. In these studies exposure variables are compared between two groups: the cases, which have developed the disease, and the controls, which have not. Another type of design are cross-sectional studies, also known as surveys. Surveys are a tool frequently used in heritage institutions, usually as a descriptive tool to report the prevalence of certain damage types within a collection.

However, cross-sectional studies can also be analytical. This paper discusses the design and analysis of cross-sectional studies as an analytical tool to study heritage collections. As an example, we used the collections of the Amsterdam City Archive to demonstrate an observational study, designed to identify the most relevant factors involved in the mechanical degradation process in archive and library collections. To meet this goal, we assessed a set of properties of records. These consisted of physical and chemical properties as well as metadata. This set of properties was based on the classification of failure causes provided by reliability engineering. The observed mechanical degradation was also recorded. The epidemiological approach, comparing groups of objects according to their characteristics, then enabled us to evaluate the relevance of the assessed factors in relation to the observed mechanical failure in the collections surveyed.

The focus of this paper is mainly to discuss the methodology used in this observational study. Therefore, results are reported with the aim to support the discussion of the methodology. As a final note, we briefly discuss the relevance of the results obtained by this epidemiological approach to support collection management.

2. Methodology

2.1 Observational study design

The observational study was carried out on archive documents, loose sheets and bound records, including their bindings, dating from 1600 to 2000.

Throughout the study, data was collected and analysed on two levels. The first level, "structures", focused on the study of an inventory number as a whole. Inventory numbers can consist of a stack of loose records or a binding. The second level, referred to as "components", collected data of the characteristics of single paper sheets.

For the purpose of this study, and following the epidemiological approach, outcome variables were defined as those related to the assessment of mechanical failure. The factors that might affect the degree of failure were defined as the exposure variables. As shown in Table 1 the exposure variables were grouped according to the four causes of failure proposed by reliability engineering. The sources used to collect data for each variable are also reported in Table 1. This table and the terminology will be described in more detail in later sections.

Table 1

Summary of the observational study design, including the outcome and exposure variables assessed according to two levels, structure and components, and the data source.

<table>
<tr>
<td colspan="2" align="center">VARIABLE</td>
<td colspan="2" align="center">LEVEL</td>
<td rowspan="2"></td>
</tr>
<tr>
<td colspan="2"></td>
<td colspan="2" align="center">1. Structure
(inventory number)</td>
<td align="center">2. Component
(single sheet)</td>
</tr>
<tr>
<td colspan="2"></td>
<td align="center">A. Stack
loose records</td>
<td align="center">B. Bindings</td>
<td></td>
<td></td>
</tr>
<tr>
<td rowspan="7">Outcome</td>
<td>Number failed sheets (total)</td>
<td>Visual observation (counting)</td>
<td>Visual observation</td>
<td>---</td>
<td rowspan="7">Failure</td>
</tr>
<tr>
<td>Number failed sheets (failure 1)</td>
<td>Visual observation (counting)</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Number failed sheets (failure 2)</td>
<td>Visual observation (counting)</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Number failed sheets (failure 3)</td>
<td>Visual observation (counting)</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Failed yes/no</td>
<td>---</td>
<td>Visual observation</td>
<td>---</td>
</tr>
<tr>
<td>Failed part</td>
<td>---</td>
<td>Visual observation</td>
<td>---</td>
</tr>
<tr>
<td>Failure category (0-3)</td>
<td>---</td>
<td>---</td>
<td>Visual observation</td>
</tr>
</table>

Exposure	Failure on fold	---	---	Visual observation	**Design**
	Thickness (mm)	Physical measurement	Physical measurement	Physical measurement	
	Weight (g)	Physical measurement	Physical measurement	---	
	Height (mm)	Physical measurement	Physical measurement	---	
	Width (mm)	Physical measurement	Physical measurement	---	
	Number of sheets (total)	Visual observation (counting)	---	---	
	Type of binding	---	Visual observation	---	
	Binding material	---	Visual observation	---	
	Protection	Visual observation	Visual observation	---	**Maintenancee**
	Prior owner	Collection management system	Collection management system	Collection management system	
	Presence of fold	---	---	Visual observation	
	Number requisitions in reading room since 1998	Collection management system	Collection management system	Collection management system	**Usage**
	Number years inventory available to visitors	Collection management system	Collection management system	---	
	Date (age)	Collection management system	Collection management system	Collection management system	
	Paper type	---	---	Near infrared spectrometry (SurveNIR)	**Manufacture**
	pH	---	---		
	Degree of polymerization	---	---		
	Tensile strength (N)	---	---		
	Tensile strength folded (N)	---	---		
	Protein content (%)	---	---		
	Rosin content (mg/g)	---	---		
	Lignin content (mg/g)	---	---		

2.2 Sampling

The collections of the Amsterdam City Archives, which contain archival records on ca. 50 km of shelving, were used for this case study. As the purpose of this research was not to report the prevalence of damage within the collections of the Amsterdam City Archives, the sample was not required to be representative of the whole of the population. The sampling method was designed to obtain an equal representation of groups of interest within certain variables.

2.2.1 Structure level sampling

Based on the results of a pilot conducted on 100 inventory numbers from four different collections and the information on the characteristics of the records already available within the archive, we selected four exposure variables: the number of requests in the

reading room since 1998, the age of the documents, prior owner of the collection before it entered the archive, and the number of sheets within an inventory number.

Inventory numbers were randomly selected from two groups: those requested more than 10 times since 1998 and the group with fewer than 10 requisitions. The selection of the objects in the two groups (high and low frequency of use) continued until an equal representation of all the relevant groups (century, more/less 100 sheets, private/public owner) was obtained within each of the two groups.

2.2.2 Components level sampling

Within an inventory number with loose records, one single sheet for each failure category (see failure assessment section below) was randomly selected. This means that, for each inventory number, one to a maximum of three sheets were selected depending on whether or not a certain failure category was present within the inventory number. In addition, one sheet with no damage was selected from each inventory number in order to be compared with the damaged ones.

Research findings have indicated that the chemical characteristics of bound paper can vary depending on where the measurement is taken within the book block (8). Therefore, in order to reduce possible differences in the measurements due to the diversity of sample locations, one page in the middle of the text block was consistently selected. Measurements on loose and bound records were taken on the upper right-hand corner of the pages, which is expected to be less soiled as it not touched when turning the pages.

The resulting sample, including the pilot, consisted of 360 inventory numbers from 80 different collections: 186 inventory numbers with loose records and 174 bindings. At the "components" level, 530 sheets were assessed.

2.3 Failure assessment: outcome variable

Four categories were used to assess mechanical failure in paper sheets based on the visual assessment of the number of tears or missing pieces per sheet:

- No failure (0): no tears or missing pieces

- Failure 1: up to 5 tears, smaller than 2 cm

- Failure 2: more than 5 tears smaller than 2 cm and/or tear(s) larger than 2 cm

- Failure 3: when a piece was missing, regardless of the number of tears

These categories differ from the ones used in the experimental study conducted by Strlič et al., designed to develop the wear-out function for historic paper (4). In this study, tears (categorized as small, medium and large) and missing pieces were recorded individually for each sheet. However, due to the notably lager sample size used in this survey and taking into account that a single inventory number can comprise more than 500 sheets, this level of detail was not achievable in this observational study. Instead, we chose an ordered classification, where a tiny tear (failure 1) can progress into a larger tear (failure 2) and eventually become a missing part (failure 3).

This ordered classification in paper-based objects is less clear with regards to bindings. Hence, a binary response was recorded depending on whether parts of bindings were torn or broken. The failed part of the binding was also recorded separately. Bindings which had been restored were included in the category "failed", as the interest was whether or not failure had occurred rather than the actual usability of the objects.

2.4 Properties assessment: exposure variables

2.4.1 Design

Design can be understood as the specifications of an object or, in other words, the physical description of the object: the size, type of binding, thickness of paper, etc. These features are important aspects when analysing the reliability of certain groups of objects, as these features are frequently related to the end-use that the objects are meant to fulfil. Therefore, the physical description of the object was recorded. For each inventory number, the following variables were assessed: weight (g), height, width and

thickness (mm). In addition, the thickness of single sheets of paper (mm) was measured using a micrometer (standard DIN-863-1).

In the case of loose documents, the total number of sheets in an inventory number was counted. In the case of bindings, the binding material (parchment, leather, linen) and the type of binding (hollow, tight or spring-back, among others) were recorded.

2.4.2 Maintenance

The use of archive boxes as a protective measure is well-embedded in preservation policies in archives and libraries. In the repositories of the Amsterdam City Archives, two main types of protection can be found: portfolios and boxes. In portfolios, documents are kept vertically on the lower short edge, between two boards tied together with twill tape. In boxes, documents are kept vertically on the long edge or flat in the box. These two categories - portfolios and boxes - were recorded.

Portfolios and boxes are the result of maintenance applied to the objects when they enter an archive. However, prior ownership could also be a factor to take into consideration when evaluating the differences in maintenance of the collections. Different use and maintenance can be expected for the more intensively-used public collections and for the private ones (such as a church, a university or a family collection). Therefore, one variable was whether the prior owner was a public body or a private owner.

In the past, loose records were frequently kept folded. The actual collections show evidence of this practice because the folds remain visible or the records remain folded. In order to determine the relevance of a fold related to the chance of it evolving into a tear, the presence of folds was recorded, as well as whether a tear was present in the fold.

2.4.3 Usage

In the case of the Amsterdam City Archives, reliable data is available on the number of times that an object has been accessed in the reading room since the introduction of computers in the archives in 1998.

One way to obtain information on the use of collections before the use of computerised collection management systems is to take into account when the inventory was made available to visitors, because a collection cannot be accessed publicly, i.e. on a large scale, until its content is described. Therefore, the year when the inventory was made can be considered as a proxy for the potential exposure in the reading room.

2.4.4 Manufacture

In the hystory of paper, leather and parchment production share the developments in the production were driven by an increasing demand for these products. Increasing production and a reduction of the production time negatively affected the chemical and mechanical properties of the end products, which plunged dramatically after 1850 (9,10). To understand the relevance of the use of certain raw materials and manufacturing processes with regard to the mechanical failure of archive documents, near-infrared-spectroscopy was used to obtain, non-invasively, information on the chemical and mechanical properties of single sheets of paper. The data was collected with the SurveNIR instrument (Lichtblau e.K., Germany) which uses near-infrared-spectroscopic technology and correlates the spectra with analytical data obtained on a reference collection using the traditional analytical method (11). SurveNIR provided data on the following properties: type of paper, pH, degree of polymerisation, tensile strength (N), tensile strength folded (N), protein (%), rosin and lignin content (mg/g). One limitation of SurveNIR is that it cannot determine the DP of ground wood paper, which is the type of paper of most concern in terms of preservation, as a low DP is expected.

As there is no tool comparable to SurveNIR for obtaining information on the properties of other materials of interest (parchment or leather), the focus was on paper-based documents.

2.5 Data analysis

In statistical terms, the aim of this work was to identify the exposure variables that have a larger effect on the outcome variable, understood as mechanical failure. To compare groups of interest, data were grouped. For example, for most of the analyses the discrete variable of *failure* was grouped as a binary outcome, *failed* and *not failed*. The group *failed* included *failure 1, 2* and *3*.

To analyse the relevance of an exposure variable in relation to the event of interest, binary logistic regression was conducted. The event of failure is introduced as a binary response (failed or not failed) and the exposure variables can be added or removed from the regression to study whether their contribution to the event of failure is statistically significant.

Scatter plots with linear regression were used to analyse the correlation between pairs of variables. To show the correlations between more than two variables, we conducted principal component analysis (PCA), a mathematical procedure to reduce the number of variables to a few linear combinations of the data, called principal components, but still containing maximum variance of the data set. In the loading plot, an angle of 90^0 between two loadings can be interpreted as lack of correlation between the respective variables, while angles of 0^0 or 180^0 indicate positive or negative correlations between the associated variables, respectively.

The size effect of an exposure variable on the failure was also analysed. In the case of a numerical outcome, the mean outcome in two exposure groups was calculated and then we tested whether the mean difference was statistically significant (two-sample *t*-test, assuming equal variance).

In the case of the categorical variables assessed on the bindings, 2×2 or larger contingency tables were formed and chi-squared tests were conducted. In the 2×2 table, the observations were classified according to whether they had experienced the failure outcome and also according to the exposure group.

The Minitab-17 program (Minitab Inc., PA, USA) was used to plot and analyse the data.

3. Results and discussion

3.1 Associations

One of the challenges of cross-sectional studies is coping with the large amount of data collected during the survey. In order to single out the most relevant exposure variables in relation to the outcome variable, multivariate data analyses can be conducted as a first approach to reveal associations between the different variables. For example, we introduced the exposure variables for inventory numbers with unbound records in a binary logistic regression. This type of regression permitted us to introduce whether a single sheet had failed or not as an event, in order to evaluate which exposure variables are associated with this event. The regression indicated that the factors statistically significant to the event of failure are the number of times that an inventory has been requested in the reading room, the document age, the thickness of the inventory number, the type of protective enclosure (p <0.001) and the prior owner (p=0.047). Regression showed that the year when the documents were made available in the reading room (p=0.155) is not significant to explain the occurrence of failure.

In order to visualize how strong the associations are, PCA was conducted including the variables showing a strong evidence of a real effect (p <0.001). By conducting this type of multivariate analyses, the exposure variables that are correlated to the event of interest, mechanical failure in our study, are identified. The relevant data are singled out and the amount of data for further analysis is reduced.

The PCA showed that the percentage of failed sheets within an inventory number with unbound records is strongly correlated to the number of times that the inventory number has been requested in the reading room (Fig. 1)[1]. This result might seem obvious as

mechanical degradation is related to handling; the higher the number of requests, the higher the percentage of mechanical degradation. But, at the same time, it also implies that most of the damage that we now observe happened since the objects entered the archive and started being used intensively, and that other factors, for instance, maintenance and use before entering the archive, play a smaller role.
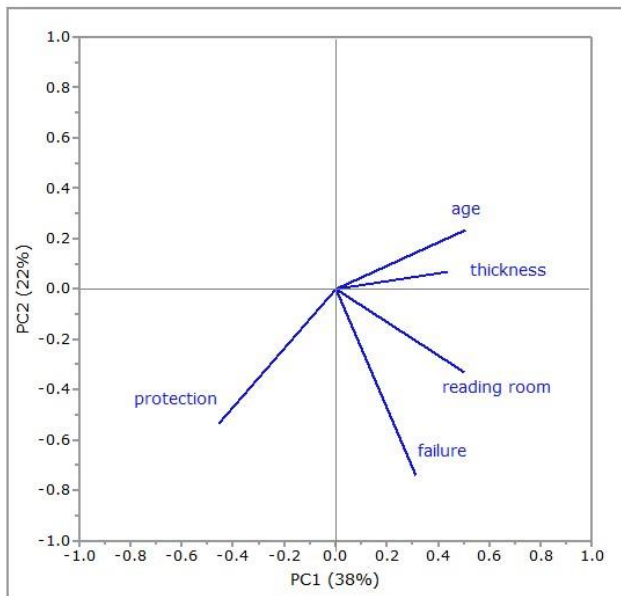
Fig. 1

Loadings of the PCA, showing correlations between failure (percentage of failed sheets within an inventory number), number of requests in the reading room, type of protective enclosure, age and thickness of inventory numbers with loose sheets. The % variance explained by the principal components is in parentheses; n= 182.

3.2 Effect size

When an observational study is designed, the expected effect size is used to determine the sample size. As this research is the first study which focusses on mechanical degradation in the actual setting of the collections, the expected effect size has not been quantified yet. However, in behavioural science, where highly scattered data is not uncommon, sample sizes have been suggested depending on which effect size we want

to detect. In the case of two-sample *t*-test, for instance, to detect a medium difference with 0.80 power and a significance level of 0.05, Cohen (12) suggests 64 observations for each group while, for a small difference, 393 are needed. A large effect size can be detected with only 26 observations for each group. According to Cohen, small, medium and large effect size are defined as *d*= 0.20, 0.50 and 0.80, where *d* is the difference between independent means, divided by the within-population standard deviation (12).

As an example, Table 2 summarizes the results of the two-sample *t*-test regarding two exposure variables: the use in the reading room (which has been identified as significant by the logistic regression), and the prior owner (identified as not significant for the occurrence of failure). The results show that a high frequency of use does have an effect on the percentage of failed sheets compared to low accessed inventory numbers (*t*=-4.58, df=100, *p* <0.001). The same test, conducted to compare types of former owners, shows that the mean failed sheets in inventory numbers owned by public bodies is slightly higher than those with a former private owner. However, this difference is not statistically significant (*t*=-1.37, df=176, *p* 0.173). If we take into account the sample size in this study, then we can conclude that as a variable, the reading room has a large effect size on failure. In the pilot study the same results were observed with only 37 observations for each group. The effect size of 'prior owner' seems not to be significant. But, to be able to conclude that a size effect is entirely absent, the sample size should be increased. By increasing the sample size, the confidence intervals (CI) for the mean difference decrease and smaller differences between groups can be statistically detected.

Table 2. Results of *t*-tests and descriptive statistics on the percentage of failed sheets within an inventory number grouped by number of requisitions in the reading room since 1998 and prior owner.

Exposure variable	Outcome variable: Failed sheets (%)							
	M	SD	SE	n	95% CI for Mean Difference	*t*	df	*p*

	M	SD	SE	n	CI	t	df	p
Reading room <8x	18.1	25.6	2.2	136				
					-25.1, -9.9	-4.58	100	<0.001
Reading room ≥8x	35.6	22.2	3.1	50				
Owner: private	25.3	28.6	3.0	94				
					-2.3, 12.6	-1.37	176	0.173
Owner: government	20.2	22.6	2.4	92				

Note. M=mean, SD=standard deviation, SE=standard error of the mean, n=number of observations, CI=confidence interval, *t*= *t*-test, df=degrees of freedom, *p*= P-value

3.3 Failure rate

In cross-sectional studies it is not known when failure occurred, as the data is collected at one single point in the time. Only studies in which the individuals are followed during a certain time interval, can provide the data needed to conduct survival analysis to calculate the failure rate. However, cross-sectional studies can still provide information on this matter based on other metrics, such as the number of requests in the reading room or the age of the documents.

In Fig. 2, despite the data scatter, the linear regression shows a good correlation between the percentage of failed sheets in an inventory number and the number of times that the documents have been requested in the reading room since 1998 (R^2=.541, *p* <0.001, n=90). As the percentage of failed sheets increases linearly with the number of requests, the failure rate for inventory numbers with loose sheets can be determined as a function of the number of times accessed. These data suggest that the higher percentage of failed sheets is due to the intensive use of the records since their accession. Those records with a low frequency of use show a percentage of failed sheets lower than 10%. As most of these records have not been requested, i.e. accessed, this percentage of failure below 10% seems to be representative of the damage occurred before their accession to the archive.
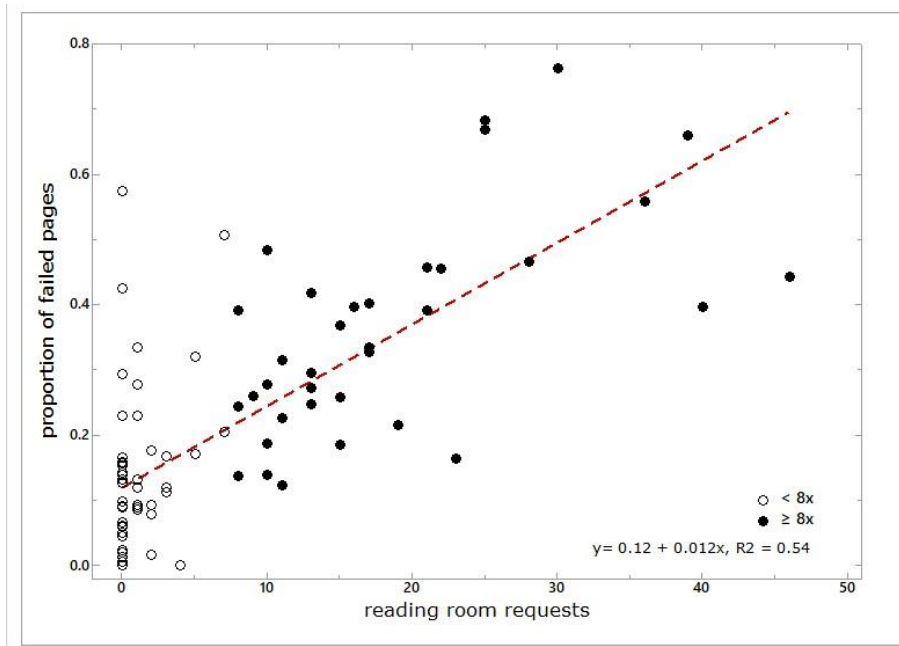
Fig. 2

Increase in percentage of failed sheets within an inventory number by the number of times that the inventory number has been accessed in the reading room since 1998. The plot shows inventory numbers containing more than 100 sheets. The regression line is drawn using all data points.

In the case of the Amsterdam City Archives, reliable data on the number of requests is available since 1998, the year when computers were introduced to record the number of requests. Prior to 1998, requests were also recorded but using another type of counting method which is not comparable to digital recording. However, as there is a clear association between the number of visitors and the number of requested records through the years (Fig. 3), the number of visitors can be seen as a good indicator of the use of the collections in the reading room. Since the mid-1970's, there has been growing interest in archives (Fig. 3), resulting in about 2/3 of the total use of the archives since 1983 and 1/3 since 1998 (397.452 of the 1.205.344 visitors since 1901). Therefore, the data available for the reading room since 1998 are not only representative of the last 18 years, but also for 1/3 of the total use of the collection within the archive. Although the

failure rate cannot be determined, the data give an indication of at what rate mechanical damage has occurred.
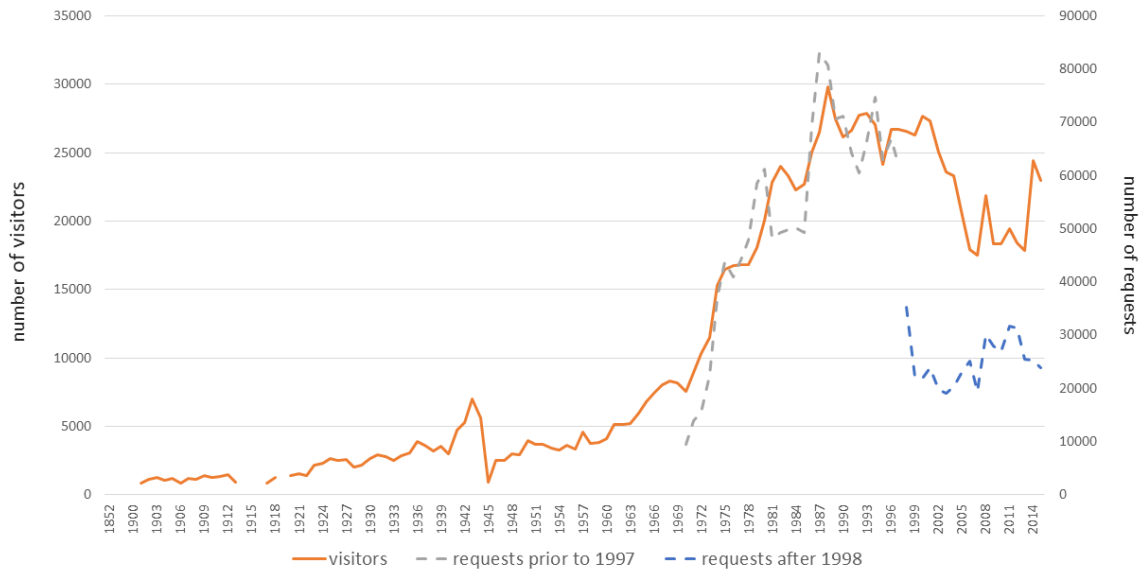
Fig. 3

Plot of number of visitors in the reading room per year between 1901 and 2016 at the Amsterdam City Archives and number of requests according to the counting method used prior to 1997 and the digital counting after 1998. Between 1852, the year the archive was founded, and 1900, no record of the number of visitors was kept.

Another example of how failure rate can be obtained using epidemiological data is shown in Fig. 4. Inventory numbers dating after 1850 show a clear difference in the percentage of mechanical failure between low and highly accessed documents. The group of inventory numbers with a low frequency of use shows a constant rate of deterioration (R^2= 0.505, p <0.001, n=42) (Fig. 4b). Therefore, based on this evidence, it is possible to quantify: (i) the effect of intensive use of records; and (ii) the accumulation of wear and tear through time for collections dating from 1850 with a low frequency of use since they entered the archive.
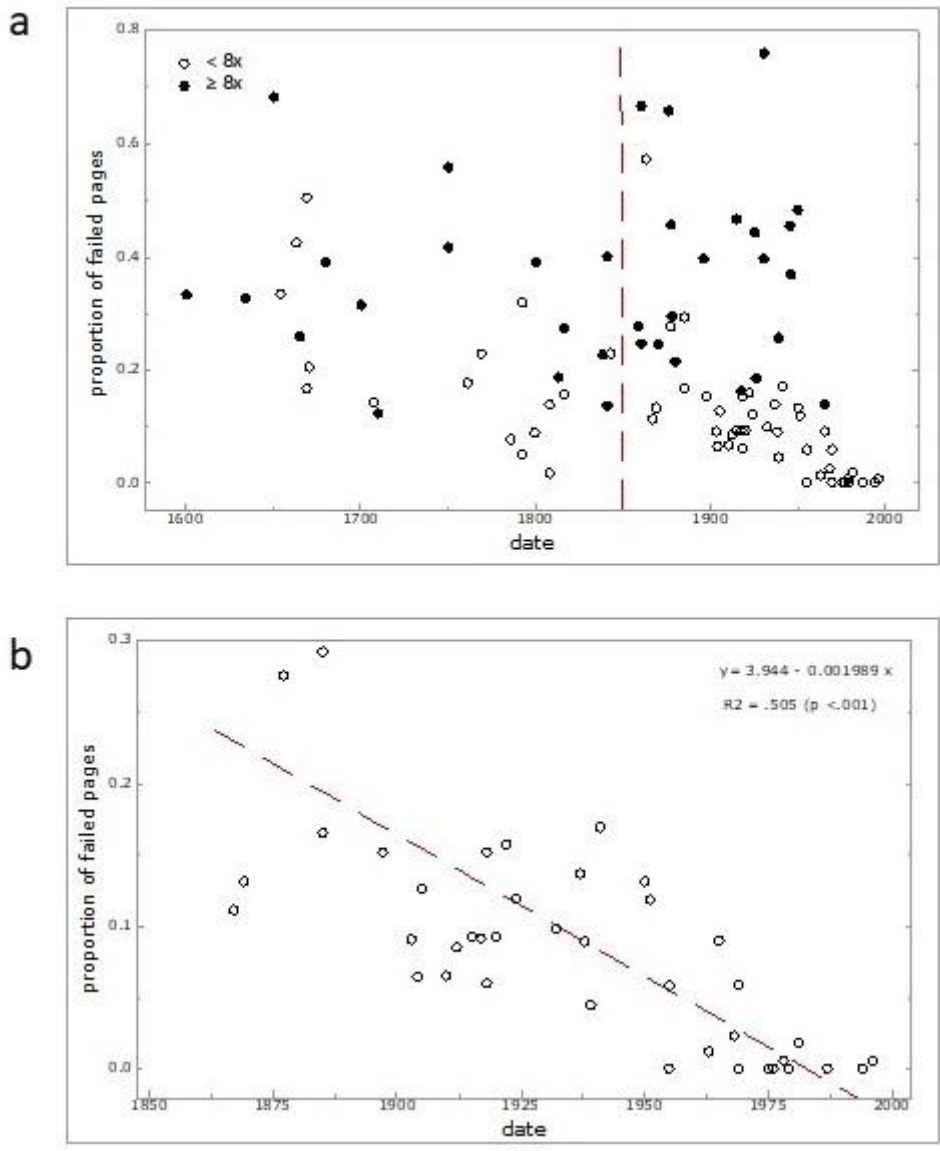
Fig. 4

a Scatterplot showing the percentage of failed sheets of inventory numbers dating from 1600 to 2000. The vertical dashed line represents the transition from rag to wood-based fibre paper.

b The same plot, with linear regression, only including inventory numbers dating post 1850 that have been accessed less than 8 times since 1998. One outlier is not included. Both plots show inventory numbers containing more than 100 sheets.

At this point it is important to note that the obtained percentages of failed sheets are directly related to the threshold chosen to define failure in this study. As we were

interested to know at which point mechanical degradation occurs, a threshold was chosen that allowed us to record the smallest visually observable evidence of mechanical damage. Other thresholds for failure will therefore result in different percentages of the outcome variable. In addition, we need to point out that the descriptive statistics showed that the sample of this case-study consisted mostly of good quality paper. Of the assessed sheets of paper in this study, only 14% showed a DP lower than 800, 7% showed a pH lower than 5 and 8% was identified as ground wood paper (paper containing more than 75 mg/g lignin). Recent research has shown that it is important to distinguish between papers with a DP lower or higher than 800 when mechanical degradation is studied (4). Hence, the results presented in this paper are representative of the group of collections containing papers with a DP higher than 800. Collections with brittle paper are not part of this research and will be addressed in future work.

It is also important to point out that one of the sources of the scatter in the data was found to be related to the thickness of the paper stack. The data indicates that the thickness of the inventory numbers is relevant when inventory numbers are thicker than 2.5 cm (approx. equivalent to inventory numbers containing more than 100 sheets)(13) . Therefore, the plots shown in this section only include inventory numbers containing more than 100 sheets, where the effect of use is clearer.

3.4 Comparison groups

The richness of the data collected during an observational study is partly due to the possibility of obtaining more detailed information from the same dataset depending on how the observations are grouped. For example, to investigate whether differences in the outcome variable are observable between groups with particular physical and chemical properties.

Tensile strength and DP, both indicators of the mechanical properties of the paper, are the two factors most correlated with failure amongst all the studied properties. As above,

this was determined with a PCA that explains 68% of the variance (Fig. 5). The analysis also showed that pH, an indicator of the rate of chemical degradation (10) is not directly correlated to failure, but to DP; the lower the pH, the lower the DP.
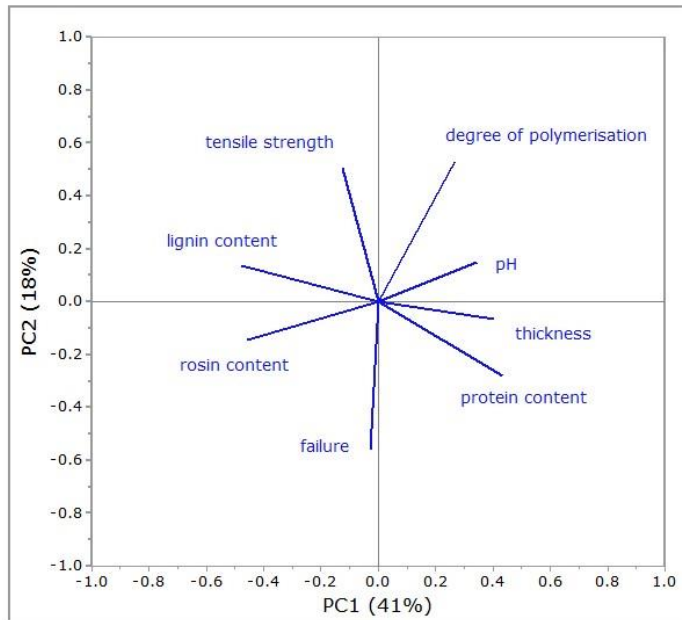

Fig. 5

Loadings of the PCA, showing the correlations between failure and chemical, physical and mechanical characteristics of single paper sheets, including rag, bleached pulp and ground wood paper. The % variance explained by the principal components is in parentheses; n= 394.

However, more detail on the relevance of the chemical properties can be obtained if the observations are grouped according to whether pages have been folded or not. Of the assessed sheets dated prior to 1900 (mostly rag paper), 95% are bifolios and/or have been kept folded, compared to 57% of the folded sheets found dated after 1900 (mostly bleached pulp paper). Of the assessed folded sheets, 55% had failed by the fold and, for 33% of this group, the fold is the only part that has torn. If we take papers dating after 1900 as an example, then the mean difference of tensile strength for failed and not failed sheets when no fold is present is 10.3 (95% CI 6.55 to 14.04, *df* = 97, *t*= 5.46 *p*<

0.001), suggesting that failure tends to occur when the tensile strength is low (tensile strength around 40 N). However, for sheets that have been folded, the difference between the two groups is not statistically significant (p 0.052), indicating that there is little evidence that failure is correlated to low values of tensile strength, and therefore failure will occur in the high region of tensile strength values as well (tensile strength of 50 N). Hence, the presence of folds could explain partly why failure has occurred for papers which would usually be graded in the low risk group for mechanical degradation, based on the high DP and tensile strength values.

The importance of taking into account the presence of folds when analysing the effect of the chemical and mechanical properties of paper is a good example of the relevance of the study of documents in the real setting of the collections, where exposure variables are present that cannot be easily added to the design of experimental studies.

Another example of the potential of the analysis comparing groups is illustrated by the analysis of the failure of the bindings. The PCA shows that mechanical degradation can be explained by the exposure time (age) and the number of times that the object has been requested in the reading room (Fig. 6). 78% and 71% of the bindings from the 17th and 18th century respectively have failed, while 56% of the assessed bindings dated from the 19th century show signs of degradation, and the percentage of failure drops to 30% for 20th century bindings (χ^2 (3, n= 174) = 22.03, p <0.001). Regarding the failure percentage related to the reading room, 48% of the bindings which have been accessed fewer than 8 times in the reading room failed. The percentage of failure rises to 69% for highly accessed bindings (χ^2 (1, n = 174) = 7.70, p 0.005). The results of the chi-squared tests show that the differences observed between binding types and number of requests in the reading room are statistically significant.
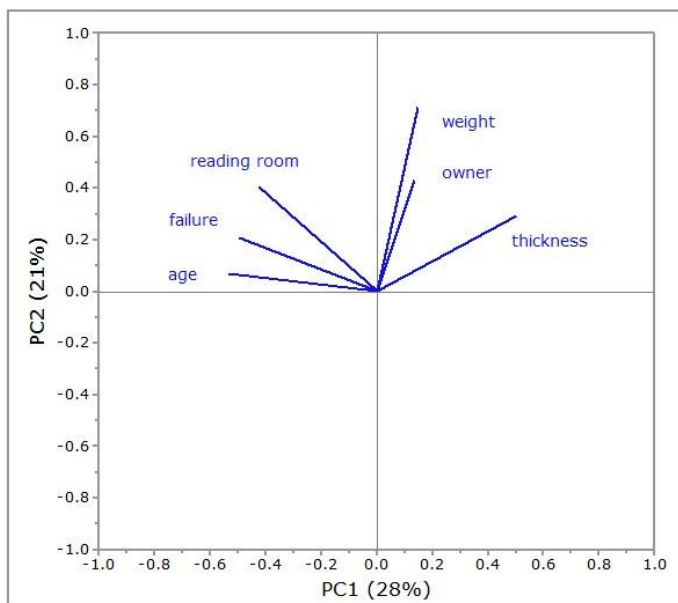
Fig. 6

Loading of the PCA, showing the correlations between failure, number of requests in the reading room, prior owner, age, weight and thickness of bindings dating from 1600 to 2000. The % variance explained by the principal components is in parentheses; n= 174.

However, when the different types of bindings are studied individually, then other factors emerge that are also correlated to failure (Fig. 7). In the case of the laced-in bindings, age is clearly correlated to failure (Fig. 7a), but the PCA also shows that public bindings show a higher percentage of failure compared to the bindings found in collections formed by private owners, probably indicating that the public bindings have been more intensively used through the years. The PCA conducted on other types of bindings highlights the factors that are more relevant to the failure for that particular type of binding.
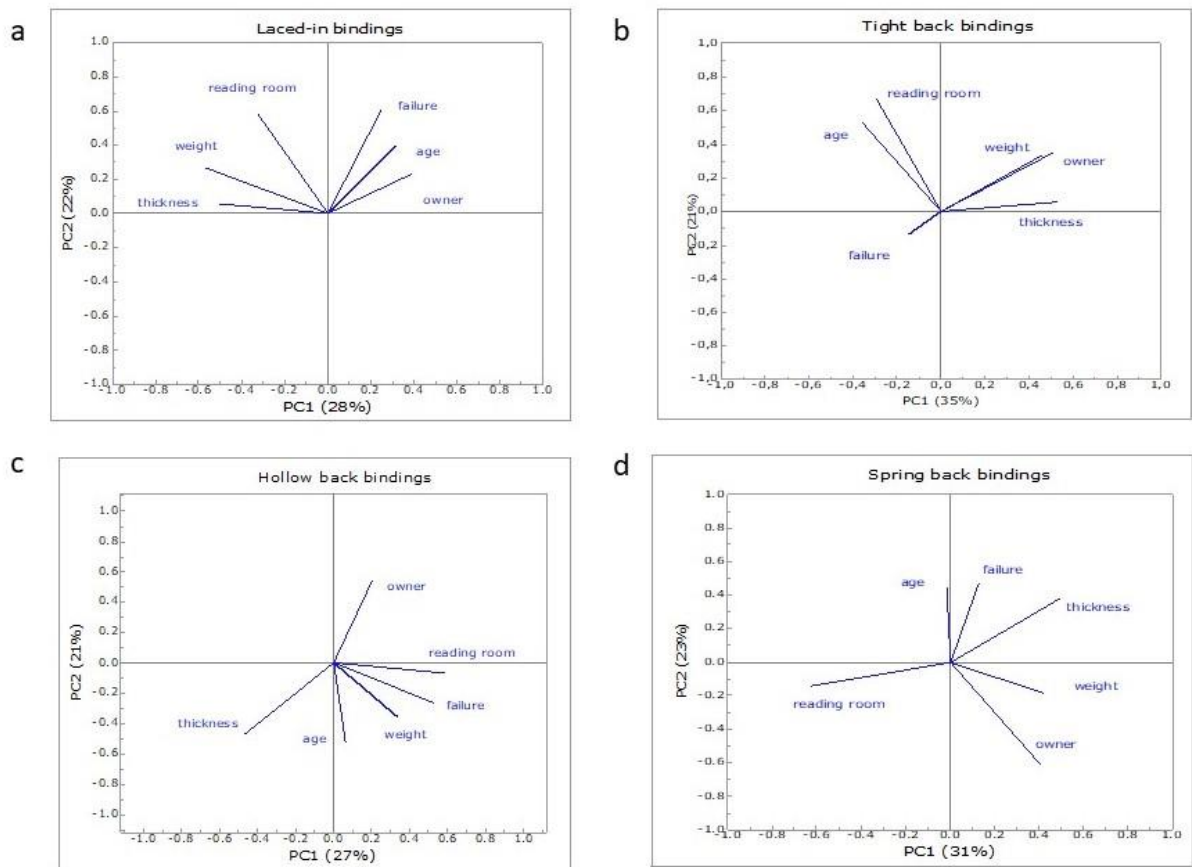
Fig. 7

Loading of the PCA, showing the correlations between failure, number of requests in the reading room, prior owner, age, weight and thickness of different types of bindings. The % variance explained by the principal components is in parentheses. a: n=65; b: n=42; c: n=15; d: n=38.

3.5 Confounding

When interpreting the results of a cross-sectional study, an important aspect to take into account is confounding; a third variable that correlates with two studied variables, therefore distorting the interpretation of their association (14). Although control of confounding can be done at the design stage (e.g. stratification), the analyses stage must also include the search for variables that might confound the results.

For example, the PCA shown in Fig. 1 indicates that 'protection' is strongly associated with the thickness (inventory number housed in portfolios are thicker than those in

23

boxes) and the age of the documents as well (documents housed in portfolios are older). In addition, the use of boxes is intimately related to the year when the inventory was made because, usually, the collection is rehoused at the same time. In the Amsterdam City Archives, this means that the collections made available to visitors after 1990 are housed in boxes, the year when archive boxes were introduced.

In order to explore the confounding between variables, logistic regression with multiple explanatory variables can be conducted (15). In the case of the variable 'protection', the coefficients in the simple regression ('protection') differs from the multiple regression, fitting 'protection' and a second variable (e.g 'age'), which confirms the presence of confounding.

Therefore, it can be difficult to decide whether a detected difference is related to the use of boxes or the effect of other variables. However, we observed that the median percentage of failed sheets housed in portfolios is 33% for highly accessed records, while the median percentage for boxes is 26% suggesting that boxes perform slightly better as protective measure when documents are requested in the reading room. Table 3 also shows that if all observations are taken into account, the difference between boxes and portfolios is not statistically significant (t=-0.24, df=155, p 0.812). However, when thickness of the inventory numbers is taken into account, boxes seem to provide a better protection for inventory numbers containing more than 100 sheets, this difference being statistically significant (t=2.37, df=88, p 0.020).

Table 3. Results of t-tests and descriptive statistics on the percentage of failed sheets within an inventory number grouped by the type of protective enclosures, **(A)** including all observations and **(B)** including only inventory numbers containing more than 100 sheets.

Exposure variable	Outcome variable: Failed sheets (%)				95% CI for Mean Difference		t	df	p
	M	SD	SE	n					

	M	SD	SE	n	CI	t	df	p
A. Portfolio	21.9	21.2	2.2	94				
					-8.6, 6.8	-0.24	155	0.812
Box	22.9	30.2	3.2	88				
B. Portfolio(>100 sheets)	25.2	17.1	2.5	46				
					1.3, 15.5	2.37	88	0.020
Box (>100 sheets)	17.7	16.8	2.5	45				

Note. M=mean, SD=standard deviation, SE=standard error of the mean, n=number of observations, CI=confidence interval, *t*= *t*-test, df=degrees of freedom, *p*= P-value

3.6 The importance of the epidemiological approach to collection management

This research has shown that the epidemiological approach to collection surveying makes it possible to visualize the relevance of certain properties of the archive records for the occurrence of mechanical degradation. The attained knowledge is therefore directly applicable to the decision-making process within collection management.

This research has made clear, for instance, that mechanical degradation mostly accumulates when the records are used in the reading room. Therefore, wear and tear will accumulate as long as records exceed the observed division line between low and highly accessed records. For those archive records, the only effective preservation measure is to reduce their physical use in the reading room, for instance, through the provision of digital access to those collections, before wear and tear occurs. The results of this research also show that, as long as the collections are frequently used, the effect of rehousing projects from portfolios to boxes will be almost unnoticeable. On the other hand, reducing the thickness of the stack may be an effective preservation measure to reduce mechanical degradation, as a positive correlation between thickness and percentage of failure for inventory numbers thicker than 2.5 cm has been observed.

The contribution of the epidemiological approach to well-informed decisions regarding collection management is even clearer in cases where the effect of a certain variable can be quantified. For example, we were able to identify at what point wear and tear accumulates as a function of the number of accesses in the reading room. Herewith,

institutions can determine at what point handling in the reading room will result in an unacceptable level of damage, according to their preservation policy.

Furthermore, the epidemiological approach can help practitioners to identify the patterns of decay for different groups of objects and to understand how reliable they are, as this aspect is of great relevance in identifying the preservation needs of each group. In the case of bindings, for example, the PCA has revealed that, for certain types of bindings, damage occurs even if they are not intensively used, due to their thickness and weight. In the case of loose records, as papers dating after 1850 show a clear difference between low and highly accessed records, this group will benefit the most if their physical use is reduced.

The findings of this study rely strongly on the available metadata about the collections and the data collected during the surveys. On the one hand, this research has highlighted the importance of recording metadata about the collections in order to study the degradation processes. In the case of mechanical failure, for example, the number of requests in the reading room has proved to be of great importance. The recording of this type of data is an aspect that should be well-embedded in the work process within the institutions. On the other hand, this approach helps practitioners to determine which data should be collected during collection surveys due to their relevance to the preservation of the collections.

In addition, the epidemiological data obtained in this type of surveys can be used to develop demographic curves specific for the collection (13). Demographic curves, based on damage functions for wear and tear (4) and chemical degradation (5), can be developed to explore the effect of different environmental conditions, e.g. reduction of temperature and/or relative humidity, on the long term preservation of the collections. In the case of the Amsterdam City Archives, the demographic plot showed that loss of fitness could be more than halved by decreasing storage temperature by 2°C and relative humidity by 10% (13).

The sample size in this study consisted of 360 inventory numbers, which is comparable to sample size required to conduct (condition) surveys in large populations, with a confidence level of 95% and confidence intervals of ±5% when reporting percentages of damages (16). This sample size permitted us to detect large and medium effect size between two groups, as 52 samples (26 observations for each group) and 130 observations (65 observations for each group) are respectively needed.

The type of data collected in this study is also comparable to the data collected during condition surveys. However, as quantification is an important aspect of the epidemiological approach, data need to be collected in a more precise way. For example, in order to quantify the percentage of failed sheets within an inventory number, each single sheet needs to be assessed. This is more time consuming than grading the condition of the whole inventory number in one of the pre-defined condition categories. However, if the outcome variable is not numerical, but categorical, as in the example of the bindings, then the assessment of the outcome variable is comparable to grading in condition surveys.

4. Conclusion

For the purpose of this study, a new approach to collection surveying is proposed based on the principles of epidemiology. These provide the study design and the data analysis needed to study the difference between groups of objects that differ regarding their level of exposure, in this case understood as the factors that might affect mechanical failure. Our research showed that the designed observational study provided usable data to identify the most relevant factors involved in the mechanical degradation process in the actual setting of the archive collections.

The clear division between outcome variables, related to the assessment of mechanical failure, and exposure variables, understood as the properties of the records, together with the classification of failure causes provided by reliability engineering proved to be a

useful framework when collecting data for the aim of this study. By grouping the data differently and conducting different exploratory data analyses we were able to visualize and quantify the effect of the assessed variables.

Data analysis showed that most of the damage that we now observe is related to the intensive use of records in the reading room. In addition, it was also possible to determine at which rate the wear and tear accumulates, as the data showed a positive linear correlation between the percentage of failure and the number of times that the documents were accessed by visitors. Furthermore, the relevance of other factors, such as the thickness of the inventory number and the age of the documents, was also identified in order to predict failure more accurately.

A limitation of this type of observational study is that it fails to inform us whether the observed correlations are causal. However, we showed that patterns of decay can be identified for groups of objects with certain levels of exposure. The PCA proved to be a good method to identify these patterns, as shown by the study of the failure of bindings. For each type of binding, the most relevant factors correlated to the occurrence of failure, which tended to be related to design and use, were identified.

Although the designed observational study provided rich data for the aim of this study, this approach also revealed several limitations, mostly related to the large scatter of the data and the sample size. Firstly, only a large and medium effect-size (effect side index d = 0.50 and 0.80) could reliably be detected. In order to detect statistically smaller differences between the groups of interest, the sample size should be increased. Secondly, in order to take confounding into account, observations should be studied in groups, implying the need for a larger sample size.

In terms of the resources needed to carry out an epidemiological cross-sectional study in archives and libraries, these do not differ substantially from the ones needed for

condition surveys. However, the epidemiological approach has the potential to provide the evidence needed to help practitioners to understand the vulnerability of the collections and to evaluate the effectiveness of the preservation measures in a more precise and accurate.

Although this research focussed solely on archive collections, this study provides an example of how epidemiology and reliability analysis can be applied in heritage science, opening up a broad set of applications within the heritage sector and bridging the growing interest in obtaining data from the actual setting of the collections within the field of heritage science.

List of abbreviations

DP: degree of polymerisation

PCA: principal component analysis

CI: confidence intervals

Declarations

Authors' contributions

CD, JG, TF and MS jointly developed the concept of this work. CD carried out the survey and all co-authors contributed to interpretation. CD wrote the manuscript and JG, TF and MS made comments and suggestions to improve it. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on request.

Endnotes

[1] Fig.1 includes the two first principal components which explain 60% of the variance in the data. Although in this plot 'age' appears orthogonal to failure, when PC3 and PC4 are investigated, this variable shows a strong association to failure, confirming the results of the regression.

References

1. Williams JIM. NHSS Report 1. The role of science in the management of the UK ' s heritage. 2009. p. 11.

2. Gurnagul N, Howard RC, Zou X, Uesaka T, Page DH. The mechanical permanence of paper: a literature review. J Pulp Pap Sci. 1993;19:160–6.

3. Zou X, Uesaka T, Gurnagul N. Prediction of paper permanence by accelerated

aging. Part I: Kinetic analysis of the aging process. Cellulose. 1996;3:243–67.

4. Strlič M, Grossi CM, Dillon C, Bell N, Fouseki K, Brimblecombe P, et al. Damage function for historic paper. Part II: Wear and tear. Herit Sci 2015 Nov 26;3(1):36.

5. Strlič M, Grossi CM, Dillon C, Bell N, Fouseki K, Brimblecombe P, et al. Damage function for historic paper. Part III: Isochrones and demography of collections. Herit Sci 2015 Dec 23;3(1):40.

6. Duran-Casablancas C, Grau-Bové J, Strlič M. Accumulation of wear and tear in archival and library collections. Part I: Exploring the concepts of reliability and epidemiology. Herit Sci 2019; 7:10.

7. Bhopal RS. Concepts of epidemiology. Integrating the ideas, theories, principles, and methods of epidemiology. 3 ed. Oxford: Oxford University Press; 2016. p. 4.

8. Porck HJ, Smit WJT. Schade-inventarisatie bij krantenmateriaal. Een onderzoek naar de zuurgraad en kwaliteit van ingebonden kranten in vijf Nederlandse bibliotheken [Newsprint survey. A study of the acidity and quality of bound newsprints in five Dutch libraries]. Open. 1990; 22:160–2.

9. Clapp VW. The story of permanent/durable book-paper, 1115-1970. Restaurator. 1972;1(s3): 1-51.

10. Strlič M, Kolar J, editors. Ageing and stabilisation of paper. Ljubljana: National and University Library; 2005. Chapter 1.

11. Lichtblau D, Strlič M, Trafela T, Kolar J, Anders M. Determination of mechanical properties of historical paper based on NIR spectroscopy and chemometrics – a new instrument. Appl Phys A. 2008 Apr 1;92(1):191–5.

12. Cohen J. A Power Prime. Psychol Bull. 1992;112(1):155–9.

13. Duran C, Grau-Bové J, Fearn T, Strlič M. Data mining in collections: From epidemiology to demography. In: ICOM-CC 18th Triennial Conference, Copenhagen. 2017.

14. Kirkwood BR, Sterne JAC. Essential Medical Statistics. 2nd ed. Oxford: Blackwell Science; 2003. p. 179.

15. Daly LE, Bourke GJ. Interpretation and uses of medical statistics. 5th ed. Oxford: Blackwell Science; 2007. Chapter 11.

16. Instituut Collectie Nederland. De steekproef als hulpmiddel bij collectiebeheer [Sampling as a tool for collection management].ICN-Informatie. 2000;4:1–4.