Assessing Nursing Homes Quality Indicators' between-Provider Variability and Reliability: A Cross-Sectional Study Using ICCs and Rankability

Lauriane Favez [1] , **Franziska Zúñiga** [1,]* , **Narayan Sharma** [1] , **Catherine Blatter** [1] and **Michael Simon** [1,2]

1   Institute of Nursing Science, University of Basel, Bernoullistrasse 28, 4056 Basel, Switzerland;
   lauriane.favez@unibas.ch (L.F.); narayan.sharma@unibas.ch (N.S.); catherine.blatter@unibas.ch (C.B.);
   m.simon@unibas.ch (M.S.)
2   Nursing and Midwifery Research Unit, Inselspital Bern University Hospital, Freiburgstrasse,
   3010 Bern, Switzerland
*   Correspondence: franziska.zuniga@unibas.ch; Tel.: +41-61-207-09-13

check for updates

Abstract: Nursing home quality indicators are often used to publicly report the quality of nursing home care. In Switzerland, six national nursing home quality indicators covering four clinical domains (polypharmacy, pain, use of physical restraints and weight loss) were recently developed. To allow for meaningful comparisons, these indicators must reliably show differences in quality of care levels between nursing homes. This study's objectives were to assess nursing home quality indicators' between-provider variability and reliability using intraclass correlations and rankability. This approach has not yet been used in long-term care contexts but presents methodological advantages. This cross-sectional multicenter study uses data of 11,412 residents from a convenience sample of 152 Swiss nursing homes. After calculating intraclass correlation 1 (ICC1) and rankability, we describe between-provider variability for each quality indicator using empirical Bayes estimate-based caterpillar plots. To assess reliability, we used intraclass correlation 2 (ICC2). Overall, ICC1 values were high, ranging from 0.068 (95% confidence interval (CI) 0.047–0.086) for polypharmacy to 0.396 (95% CI 0.297–0.474) for physical restraints, with quality indicator caterpillar plots showing sufficient between-provider variability. However, testing for rankability produced mixed results, with low figures for two indicators (0.144 for polypharmacy; 0.471 for self-reported pain) and moderate to high figures for the four others (from 0.692 for observed pain to 0.976 for physical restraints). High ICC2 figures, ranging from 0.896 (95% CI 0.852–0.917) (self-reported pain) to 0.990 (95% CI 0.985–0.993) (physical restraints), indicated good reliability for all six quality indicators. Intraclass correlations and rankability can be used to assess nursing home quality indicators' between-provider variability and reliability. The six selected quality indicators reliably distinguish care differences between nursing homes and can be recommended for use, although the variability of two—polypharmacy and self-reported pain—is substantially chance-driven, limiting their utility.

Keywords: nursing homes; long-term care; benchmarking; quality indicators; health care; quality of health care

1. Introduction

Quality indicators are used worldwide to monitor, assess and report the quality of care provided in healthcare settings by measuring specific health care structures (e.g., staffing), processes (e.g., patient referrals) or outcomes (e.g., nosocomial infections) that reflect quality of care [1,2]. Healthcare providers can use them for continuous quality monitoring or for benchmarking, i.e., to compare healthcare

providers, to measure quality against accepted standards or to measure developments over time. Benchmarking these indicators allows evaluation and comparison of healthcare providers' quality of care levels. However, concerns have been expressed regarding the value of quality indicators and publicly reported benchmarking [3–6]. Therefore, to ensure that quality indicators provide useful information, they need to be evaluated by criteria including but not limited to validity, feasibility and relevance. Particularly in the context of publicly reported benchmarking, quality indicators have to reliably assess differences in quality of care between healthcare providers. Quality indicators should thus be able to show (1) between-provider variability and (2) reliability. Between-provider variability refers to the quality indicator's ability to indicate differences in quality of care beyond chance, i.e., to identify higher-performing and lower-performing healthcare providers [7]. The quality indicator's reliability is its capacity to accurately and consistently measure the particular quality it indicates [8,9]. If both characteristics apply to the quality indicators used, they can be used for benchmarking, which has the potential to support the maintenance and improvement of quality of care [10].

In several countries (e.g., the United States, Australia, Canada), nursing home quality indicators have been measured and reported publicly for some time [11]. Quality indicators cover a wide variety of themes, most commonly physical restraints, falls, pressure ulcers and weight loss. They tend to be assessed either with routinely used instruments (e.g., Resident Assessment Instrument-Minimum Data Set (RAI-MDS) in the United States and Canada) or through specific data collections (e.g., the National Aged Care Mandatory Quality Indicator Program in Australia). Depending on the country, the results may be reported to the nursing home administration, regionally and/or nationally [12–15]. In Switzerland, despite legislation providing legal bases for the measurement and public reporting of quality indicators in nursing homes since 1994, their measurement at the national level started only in 2019, with the first quality indicator results still not published as of October 2020 [16]. Their selection and development included a review and expert consultations, considering a variety of criteria (e.g., relevance, feasibility, reproducibility) [17,18]. In 2016, based on a broad stakeholder consultation, the first set of six quality indicators indicating percentages of specific health processes or outcomes among nursing home residents were selected: the percentage of residents with polypharmacy (one quality indicator), experiencing pain (two quality indicators), being subjected to physical restraints (two quality indicators) and with weight loss (one quality indicator). The details of the selection process are reported in Appendix A. After the selection was made, we conducted an analysis to evaluate whether the six quality indicators' between-provider variability and reliability were adequate for national and publicly reported benchmarking.

Methods for assessing between-provider variability include intraclass correlation 1 (ICC1) and rankability. To assess each indicator's capacity to differentiate between facilities, we used the ICC1, which reflects "the proportion of variance that is accounted for by the group level" [19]. While this method has been used in other settings to evaluate quality indicators' between-provider variability (e.g., hospital quality indicators), it has not yet been used in long-term care contexts [20]. A second metric to assess between-provider variability is rankability, i.e., "the part of heterogeneity between . . . clinics [or nursing homes] that is due to true differences" [21]. Finally, it is also important to assess the group mean via intraclass correlation 2 (ICC2), which describes the reliability of each quality indicator [22]. Therefore, this study aims to use ICC1 and rankability to report on the six selected nursing home quality indicators' between-provider variability and ICC2 to report on their reliability.

2. Materials and Methods

2.1. Design and Sample

This multicenter pilot study used routine resident data from a convenience sample of 152 nursing homes located across Switzerland's three major language regions (German, French, Italian). In 2017, 1561 Swiss nursing homes provided both medical care and social services to more than 157,716 older adults. Around 80% of these people are long-term residents, with an average length of stay

of 2.5 years [23]. Inclusion criteria at the nursing home level were for each facility to be licensed as a nursing home and to have agreed to work with a specific version of the assessment instrument including the items needed for the six national quality indicator measurements (Appendix B for further information). At the resident level, all long-term residents residing in the nursing home at the date of the data export were included.

2.2. Variables and Measurements

Each resident's birth year (YYYY format), admission date to the nursing home (YYYYMMDD format), sex (male/female) and care level (scale of 1–12) were used. Care level—calculated based on an assessment performed by the nursing home staff—was allocated a number from 1 to 12, with each higher number representing an additional 20 min of care time per day. We calculated length of stay in days from admission to data export day, and residents' age as the difference in years between birth year and year of the assessment. We also collected variables specifying each patient's depressive (depression rating scale; DRS) and cognitive status (cognitive performance scale; CPS). The DRS is calculated on a scale from 0 to 14: scores of 3 or above indicate evidence of minor to major depression [24]. The CPS is calculated on a scale from 0 ("intact") to 6 ("very severe impairment") [25]. The size of the nursing home (number of beds) was also included.

We used the following variables to calculate the six quality indicators, all of which were collected for the recall period "in the last 7 days": number of active ingredients taken; frequency and intensity of self-reported and observed pain; frequency of trunk fixation use or seating that prevents the residents from rising; frequency of bedrail use; and percentage of weight loss during the last 30 or 180 days. We also used a variable specifying whether the resident's latest assessment was that at admission (yes/no) and whether the nursing home staff evaluated that he or she had a life expectancy of under 6 months (yes/no/information not collected). In case of use of a physical restraint, we also used a variable specifying the context in which the measure was applied (use of the measure on the request or in agreement with a resident capable of judgment/use of the measure on a resident incapable of judgment/context not yet clarified).

The six selected quality indicators were defined as follows: *polypharmacy* is the percentage of residents who took 9 or more active ingredients over the last 7 days. The cut-off value of 9 is in line with other international measures of polypharmacy in nursing homes [26]. *Pain* is measured by 2 quality indicators: *self-reported pain* is the percentage of residents with daily moderate or higher pain intensity or those with nondaily very strong pain intensity in the last 7 days. *Observed pain* is the percentage of residents who showed daily moderate or higher pain intensity or those who showed nondaily very strong pain intensity in the last 7 days. *Physical restraint* is also measured through 2 quality indicators. The first measures the percentage of residents with daily fixation of the trunk or with seating that prevented them from rising in the last 7 days; the second measures the percentage of residents with daily use of bedrails or other devices on all open sides of their bed so that they could not leave the bed independently in the last 7 days. The sixth quality indicator is *weight loss*, measuring the percentage of residents with weight loss of ≥5% in the last 30 days or of ≥10% in the last 180 days.

Quality indicators are described in a numerator/denominator format. Their results are expressed as a rate for each nursing home (e.g., percentage of residents with observed pain in a specific nursing home). The numerator includes all residents for whom the outcome of interest (e.g., pain) is indicated; the denominator includes all residents except those who fit predetermined exclusion criteria. For the self-reported pain quality indicator, for example, residents were excluded if they did not give a valid answer to questions related to pain frequency or intensity. For both physical restraint quality indicators, residents capable of judgement who either requested or agreed to the measure were excluded. For the weight loss quality indicator, we applied two exclusion criteria: resident's life expectancy estimated by the staff to be under 6 months or current assessment of the resident is the admission assessment. Definitions, numerators, denominators, items measured and exclusion criteria for all quality indicators are displayed in Table 1.

Int. J. Environ. Res. Public Health **2020**, *17*, 9249

4 of 15

Table 1. Description of the six Swiss quality indicators.

Theme	Definition	Numerator	Denominator	Items Measured	Exclusion Criteria
Polypharmacy	Percentage of residents who took 9 or more active ingredients in the last 7 days	All residents who had taken 9 or more active ingredients in the last 7 days	All long-term care residents	Number of active ingredients in the last 7 days	No exclusion criteria
Self-reported pain	Percentage of residents with daily moderate or higher pain intensity or residents with nondaily very strong pain intensity in the last 7 days	All residents who reported the following pain in the last 7 days: - Daily moderate, strong or very strong, unbearable pain OR - Nondaily very strong, unbearable pain	All long-term care residents, excluding those who did not give a valid answer regarding frequency or intensity of self-reported pain	Frequency and intensity of self-reported pain in the last 7 days	No valid answer to questions on frequency OR intensity of self-reported pain
Observed pain	Percentage of residents who showed daily moderate or higher pain intensity or residents who showed nondaily very strong pain intensity in the last 7 days	All residents where the following pain was observed in the last 7 days: - Daily moderate, strong or very strong, unbearable pain OR - Nondaily very strong, unbearable pain	All long-term care residents	Frequency and intensity of observed pain in the last 7 days	No exclusion criteria
Physical restraint, trunk fixation or seating that prevents the resident from rising	Percentage of residents with daily fixation of the trunk or with seating that prevented the resident from rising in the last 7 days	All residents who had daily in the last 7 days: - Trunk fixation OR - Seating that prevents the resident from rising	All long-term residents, excluding those who wanted or agreed to the use of this measure	Frequency of use in the last 7 days and context of the measure	Residents capable of judgment who either requested or agreed to the measure

Int. J. Environ. Res. Public Health **2020**, *17*, 9249

5 of 15

Table 1. *Cont.*

Theme	Definition	Numerator	Denominator	Items Measured	Exclusion Criteria
Physical restraint, bedrails	Percentage of residents with daily use of bedrails or other devices on all open sides of the bed that did not allow the resident to leave the bed independently in the last 7 days	Residents with daily application of bedrails or other devices on all open sides of the bed, which does not allow the resident to leave the bed independently	All long-term residents, excluding those who requested or agreed to the use of this measure	Frequency of use in the last 7 days and context of the measure	Residents capable of judgment who either requested or agreed to this measure
Weight loss	Percentage of residents with weight loss of 5% or more in the last 30 days or of 10% or more in the last 180 days	Residents with a weight loss of 5% or more in the last 30 days or 10% or more in the last 180 days	All residents, excluding those with a life expectancy estimated by the staff as lower than 6 months or residents who were last assessed at admission to the nursing home	Weight loss of 5% or more in the last 30 days or of 10% or more in last 180 days	Residents with: - Life expectancy under 6 months - Last assessment at admission

2.3. Data Collection

Resident data were collected by the nursing homes via updated versions of routinely used resident assessments instruments between July 2016 and August 2017. All quality indicator information was obtained through routine data collection processes already in place, including observations (e.g., physical restraints) or conversations with residents (e.g., self-reported pain). At the time of the study in Switzerland, three assessment instruments were in use: (1) Nursing Home Resident Assessment Instrument (RAI-NH), (2) the Planification Informatisée des Soins Infirmiers Requis (PLASIR/PLEX) (computerized planning of required nursing care) and (3) the BewohnerInnen-Einstufungs-und Abrechnungssystem (BESA) (residents classification and billing system) [27]. Further information on these instruments is available in Appendix B. Each assessment instrument provider had to recruit a minimum of 50 nursing homes to ensure the sample would have an equal number of nursing homes working with each instrument. The goal was to have a total study sample consisting of minimum 10% of all Swiss nursing homes. To ensure that all residents present in each nursing home at the time of the data export were assessed at least once, data were collected for a minimum of 6 months in each nursing home.

2.4. Statistical Analysis

We examined the data on each quality indicator for completeness, plausibility and missing values. Missing data were dealt with by listwise deletion; for each quality indicator, the number of valid residents depended on the exclusion criteria. We also computed resident characteristics and prevalence rates for the six studied quality indicators. Risk adjustment for the indicators was assessed in a preparatory study using hierarchical multiple regression models, with the Akaike information criteria and odds ratios assessed for each quality indicator. All indicators were risk-adjusted for the resident's cognitive performance and care dependency with additional adjustment for polymedication with age and for both pain indicators with depression and gender. Risk adjustment variables are provided in Table 2. Statistical analyses were conducted by N.S., confirmed by C.B. and supervised by M.S., who has a track record of statistical analyses in healthcare quality measurement.

Table 2. Risk adjustment variables, prevalence rates and missing values for the six Swiss quality indicators.

Theme	Risk Adjustment Variables	Prevalence Rate, Mean %, SD [1]	Missing, % (n)
Polypharmacy	- CPS [2] - care level - age	43.0 (12.9)	0.0 (0)
Self-reported pain	- CPS - care level - depression - gender	19.7 (11.8)	13.4 (1525)
Observed pain	- CPS - care level - depression - gender	14.9 (10.4)	0.7 (81)
Physical restraint, trunk fixation or seating that prevents the resident from rising	- CPS - care level	3.4 (5.2)	0.0 (0)
Physical restraint, bedrails	- CPS - care level	13.0 (11.3)	1.6 (132)
Weight loss	- CPS - care level	7.9 (6.8)	0.1 (2)

Abbreviations: [1] SD: standard deviation, [2] CPS: cognitive performance scale.

2.4.1. Between-Provider Variability: ICC1 and Rankability

To assess each quality indicator's capacity to distinguish between providers, we computed ICC1, caterpillar plots and rankability. The ICC1 shows the proportion of variation in the quality indicator that is attributed to the group level [28]. In this context, ICC1 values typically range from 0.0 to 0.3, where values over 0.05 indicate relevant between-provider variability [19,29,30]. The ICC1 is the ratio of variance among providers (*VG*) over the total variance, i.e., the group variance (*VG*) and the within-group or residual variance (*VR*). As we analyze binary outcomes, VR is the latent scale variance of the logit model $\pi^2/3$, leading to the following equation: $ICC1 = VG / (VG + \pi^2/3)$ [29,31].

2.4.2. Reliability: ICC2

Variances for the ICC1 were calculated using a conditional generalized linear mixed model with 95% confidence intervals to assess the uncertainty of the estimate [32]. Additionally, to check the distribution of nursing home estimates and explore between-provider variability visually, we computed caterpillar plots based on empirical Bayes estimates with 95% confidence intervals [33,34]. Each quality indicator's caterpillar plot shows that indicator's estimate for each nursing home (e.g., weight loss), as well as whether it deviates positively or negatively from the grand mean across all nursing homes. We calculated ICC1 figures with the rptR package in R (Version 3.6.6., R Core Team, 2020) [31,34]. We also explored quality indicators' rankability, i.e., the part of variability between nursing homes measured by quality indicators that results from true differences in quality of care [35]. High rankability for a particular indicator allows performance ranking for that indicator, e.g., polymedication [20]. Rankability (ϱ) is defined as: $\varrho = VG/(VG + median(s^2))$, with *median*($s^2$) indicating the variance of the individual facility effect estimates from a fixed effect regression model. Rankability (range: 0–100%) refers to observed differences that might result from quality of care disparities and is classed as low (<50%), moderate (50–75%) or high (>75%) [28,35]. Finally, we assessed the six quality indicators' group mean reliability via ICC2. ICC2 is the ratio of group variance to total variance/*k*, where *k* is the number of nursing homes, i.e., $ICC2 = VG/((VG + \pi^2/3) \times (1/k))$ [19].

ICC1 and ICC2 are generally interdependent: the higher a quality indicator's ICC1, the higher its ICC2. The ICC2 typically ranges from 0.6 to 1.0, with values closer to 1 indicating higher measurement reliability.

2.5. Data Management and Ethical Considerations

At the end of the data collection period, each instrument developer anonymized all resident-level data and transferred all records to the Swiss Federal Office of Public Health, which pseudonymized them at the nursing home level. The study data were then transferred to the research group, which carried out the analysis. Data protection and confidentiality were ensured during every phase of the study. The Ethics Committee of Northwest and Central Switzerland declared that according to Swiss legislation, the study did not require ethical clearance (EKNZ Req-2017-00052).

3. Results

3.1. Sample and Quality Indicators Description

A total of 152 nursing homes participated in the study (56 for RAI-NH, 46 for PLAISIR/PLEX, 50 for BESA; mean size: 102.3 beds (standard deviation (SD): 51.2)). These housed 11,412 residents (mean age: 86.1 years (SD: 8.36); 72.8% female). The median length of stay was 859 days (interquartile range (IQR): 375–1646 days), with a median care level of 6.0 (IQR: 4–9). The mean prevalence of each quality indicator and missing item data per quality indicator are displayed in Table 2.

3.2. Between-Provider Variability: ICC1 and Rankability

The between-provider variability of this study's six selected quality indicators was relatively high: all ICC1 values were above 0.05, ranging from 0.068 (polypharmacy) to 0.396 (physical restraint, trunk fixation or seating that prevents the resident from rising). Our caterpillar plots illustrate that each quality indicator can discriminate sufficiently between the higher- and lower-performing nursing homes (Figure 1). On the one hand, we were able to identify better-performing nursing homes, i.e., those housing submean proportions of residents with the indicator result; e.g., 17 nursing homes had significantly fewer residents with polypharmacy. Only in the case of the *physical restraint, trunk fixation or seating that prevents the resident from rising* quality indicator was it not possible to differentiate higher-performing nursing homes, as many facilities had no such cases, resulting in a low mean. On the other hand, we were able to identify lower-performing nursing homes, i.e., those having significantly higher proportions of residents with the quality indicator result than the mean; e.g., 13 had a higher percentage of residents with weight loss. Rankability values ranged from low, at 0.144 (polypharmacy) and 0.471 (self-reported pain); to moderate, at 0.692 (observed pain) and 0.720 (weight loss); to high, at 0.865 (physical restraint, bedrails) and 0.976 (physical restraint, trunk fixation or seating that prevents the resident from rising). ICC1 and rankability results are provided in Table 3. Unadjusted results are provided in Appendix C.
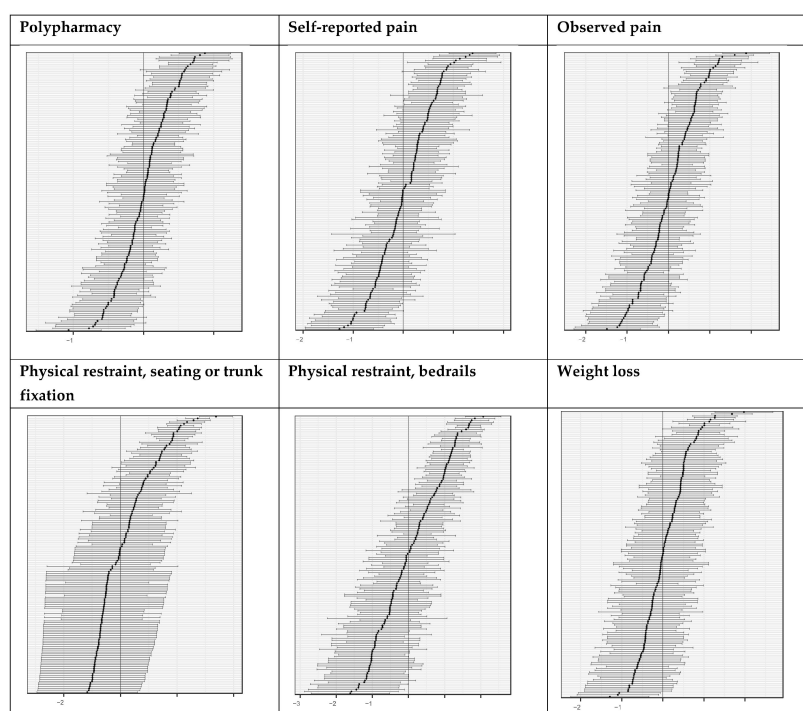
Figure 1. Caterpillar plots based on empirical Bayes estimates of the six quality indicators. In each plot, the horizontal lines represent the nursing homes: the dot in the middle of each line represents the percentage of nursing home residents to whom the quality indicator applies; the whiskers represent the 95% confidence interval (CI). The vertical line represents the (standardized, i.e., centered to 0) sample mean of the specified quality indicator. If the whiskers (i.e., the CI) do not touch the vertical line (i.e., the mean), the result of the nursing home in the quality indicator specified (e.g., the percentage of residents with weight loss in this nursing home) differs significantly from the sample mean (e.g., the mean percentage of residents with weigh loss across all nursing homes). Lower-performing nursing homes with quality indicator values above the mean, e.g., more residents with weight loss, are on the top-right side of the plot. Those with quality indicator values below the mean, e.g., fewer residents with weight loss, are on the negative (lower left) side of the plot. If the CI touches the sample mean, the nursing home's result does not differ significantly from the mean.

Table 3. Risk-adjusted results of intraclass correlation 1 (ICC1), intraclass correlation 2 (ICC2) and rankability of the six quality indicators.

Theme	ICC1 [1] (95% CI [2])	ICC2 [3] (95% CI)	Rankability (ϱ)
Polypharmacy	0.068 (0.047–0.086)	0.917 (0.889–0.935)	0.144
Self-reported pain	0.134 (0.104–0.166)	0.896 (0.852–0.917)	0.471
Observed pain	0.223 (0.131–0.325)	0.941 (0.879–0.965)	0.692
Physical restraint, trunk fixation or seating that prevents the resident from rising	0.396 (0.297–0.474)	0.990 (0.985–0.993)	0.976
Physical restraint, bedrails	0.371 (0.297–0.425)	0.989 (0.984–0.991)	0.865
Weight loss	0.137 (0.085–0.180)	0.899 (0.856–0.922)	0.720

Abbreviations: [1] ICC1: intraclass correlation 1, [2] CI: confidence interval, [3] ICC2: intraclass correlation 2.

3.3. Reliability: ICC2

The reliability results of the six quality indicators were high according to usual standards: ICC2 ranged from 0.896 (self-reported pain) to 0.990 (physical restraint, trunk fixation or seating that prevents the resident from rising). ICC2 results are found in Table 2. Unadjusted results are provided in Appendix C.

4. Discussion

For each of six selected quality indicators, this study uses ICC1, rankability and ICC2 to evaluate two important properties: between-provider variability and reliability. Our results show that four of the six quality indicators (observed pain, physical restraint, trunk fixation or seating that prevents the resident from rising and bedrails, malnutrition) have high ICC1, moderate to high rankability and high ICC2 values. This indicates respectively that between-nursing home variability was high and that these four quality indicators were generally reliable. Two indicators—polypharmacy and self-reported pain—also showed high reliability and variability beyond chance, however to a lesser extent, which makes them less ideal for comparing nursing homes. These two quality indicators represent similar challenges for nursing homes. Even with focused efforts, polypharmacy is difficult to tackle for facilities and reducing the polypharmacy rate might be complex because of structural circumstances (e.g., physician system, legal regulations). Similarly, lowering the percentage of residents with pain is complex (i.e., difficulties in and possibilities for treating chronic painful conditions) for nursing homes. Therefore, from a measurement viewpoint, while four quality indicators can be recommended without hesitation for publicly reported benchmarking, two do not fully achieve this status.

Internationally, reports of nursing home quality indicators' between-provider variability are rare. To our knowledge, the study by Rantz et al. (2004), who reported a between-provider variability evaluation for 23 nursing home quality indicators used in the United States, is the only published study to do so. That study grouped nursing homes according to resident outcomes, tested the groups for significant differences and performed pairwise comparisons [36]. Of the 23 quality indicators tested, the authors concluded that only ten could distinguish the group of nursing homes with good resident outcomes from that whose corresponding outcomes—including for polymedication and weight loss—were poorer. In contrast, ICC1 and rankability provide measures to assess each indicator's ability to differentiate between facilities. ICC1 does so while addressing clustering and multiple testing, neither of which featured in the study by Rantz et al. However, Rantz et al. were by no means exceptional in this respect: our literature review could not identify a single study using ICC1, rankability or caterpillar plots based on empirical Bayes estimates to evaluate between-provider variability in the long-term care sector, although all have been used in other fields.

Among publicly reported nursing home performance figures, a small number of countries (e.g., the Netherlands) have reported the reliability of isolated quality indicators, while Germany and

the United States have published studies or reports assessing the reliability of entire nursing home quality indicator sets (respectively, of 10 and more than 100 quality indicators) [37–40]. These studies used single item-level and/or weighted kappas and percent agreement between "gold standard" nurses and nursing home nurses to assess the selected quality indicators' interrater and intrarater reliability. Reliability results varied widely between quality indicators. The use of Cohen's kappa to assess nursing home quality indicators' reliability only provides information on the reliability of individual measures. Using ICC2 allows us to acquire information on the reliability of quality indicators at the group level, which we argue is more interesting, as the facility mean is targeted rather than the reliability of the measure at the individual level [29]. While this level of reference makes the ICC2 ideal for nursing homes benchmarking, we could identify no other studies using it as a reliability measure.

Despite having widely different cut-off values, our results show that ICC1 and rankability correlated strongly: high ICC1 values were reflected by high rankability values; e.g., our highest ICC1 value, 0.396, was linked to our highest rankability value, 0.976 (for physical restraint, trunk fixation or seating that prevents the resident from rising). The same applies for low figures: our lowest ICC1, 0.068 corresponded with the lowest rankability, 0.144 (for polypharmacy). This relationship has been explored with a similar correlation for the ranking of binary hospital quality indicators [20]. Although an ICC1 of 0.05 has been regarded as the lower threshold for quality indicators, considering the rankability found in our study, this threshold might be higher, at roughly 0.15. Indeed, if both rankability and ICC1 results are higher than the threshold, there is evidence of differences in quality of care between nursing homes, and these quality indicators can therefore be recommended.

Several issues surround the use of quality indicators that have not been adequately evaluated or simply do not meet acceptable standards. Quality indicators that cannot distinguish quality of care differences are not usable to publicly benchmark healthcare providers: they can lead to the publication of erroneous information, inappropriate comparisons or misguided quality improvement efforts, i.e., resulting from nursing home administrators' or policy-makers' use of them to set quality improvement targets. Further, inaccurate benchmarking results can lead to unjustified rewards or sanctions both by governments and by other stakeholders, particularly residents' families [3,41]. Conversely, regular, accurate reporting on meaningful quality indicators contributes to accountability and transparency in the healthcare system [10]. Therefore, ensuring nursing home quality indicators' can identify between-provider variability and reliability for benchmarking is a highly important step in their evaluation.

Although quality indicators can be extremely useful to identify quality improvement targets at the nursing home and policy levels, they cannot be used without considering the context. For example, as each quality indicator shows only one very limited aspect of a healthcare provider's care, no single quality indicator can be used to characterize providers' overall quality of care. Instead, sets of reliable quality indicators can show nursing home administrators their facilities' rankings compared to other providers, thereby allowing them to identify, prioritize and allocate resources to quality improvement targets. However, while quality indicators are excellent tools for comparing quantifiable outcomes, they do not identify poor results' underlying problems, indicate whether results are clinically meaningful or guide nursing homes regarding their improvement (i.e., regarding which specific factors require action or at which levels) [42,43]. Perhaps most importantly, while well-developed and well-evaluated quality indicators can provide valuable information to nursing homes, nursing home administrators often lack the skills, knowledge, leadership or professional and organizational capacities to put that information to good use [43]. Therefore, nursing homes need to work towards developing feasible strategies to identify and act on genuine quality improvement efforts based on quality indicator results.

Despite this pilot study's large sample size, which includes around 10% of all Swiss nursing homes (1561 nursing homes in Switzerland in 2017), the included nursing homes' mean bed count (102 beds) was somewhat higher than the Swiss average (62 beds) [23]. Even based on this rather homogeneous sample in comparison to the full sample of Swiss nursing homes, all six selected quality

indicators showed between-provider variability, indicating they could be used in Swiss nursing homes. The indicator of self-reported pain had 13.4% missing values due to respondents not wanting or not being able to answer, which precludes a proportion of nursing home residents from being represented by this indicator. Therefore, it is important to evaluate both self-reported and observed pain, since the latter includes all residents. We report ICC1 and ICC2 in this study. While the value of the ICC1 is unchallenged in assessing between-provider variability, the ICC2 has generally been less frequently used and more recent literature has identified difficulties of the ICC2 when ICC1 values are very low [44]. This is not the case in our study; however, the generally high ICC2 values might indicate less sensitivity. The rankability scores seem to provide a more nuanced picture.

5. Conclusions

Based on the six selected nursing home quality indicators' ICC1, rankability and ICC2 values, we determined that all six quality indicators can reliably distinguish differences in quality of care between nursing homes, although two operate at a lower level. Even though all are suitable as quality indicators for benchmarking and public reporting, for two of them, the observed variability is substantially driven by chance, limiting their utility. Still, they can serve nursing homes to assess their quality in this area and initiate quality improvement projects where needed. This pilot study showed that both ICCs and rankability are meaningful methods both to evaluate nursing home quality indicators' between-provider variability and reliability and to validate them. Assessing quality indicators' measurement properties is an essential step towards building sets of quality indicators that are useful in nursing home practice, policy and research. Public reporting of quality indicators increases transparency of the quality of care provided in nursing homes and provides an assessment of the national system. In practice, such indicators allow nursing homes to compare themselves with other facilities. For lower-performing nursing homes, this can be a starting point in identifying domains where quality improvement might be needed. Identifying higher-performing nursing homes helps to identify best care practices in these domains and enable learning from them. Regular evaluation of nursing home quality indicators, including between-provider variability and reliability, should be carried out and reported in all applicable contexts.

Appendix A

Selection Process of the Four Themes of the Swiss Quality Indicators

The selection of the four themes covered by the six Swiss quality indicators was based on a multistep method-guided process started in 2008 [17]. This process was managed by a Steering Committee under the leadership of the Swiss association of nursing homes (CURAVIVA), with the participation of the Swiss Federal Office of Public Health, the Swiss Conference of Cantonal Health Directors and the Swiss Federal Statistical Office. The first step consisted of a literature review on the themes measured by quality indicators in the nursing home sector at the international, national and cantonal levels [45–48]. Based on this review, the Steering Committee selected five themes: physical

restraint, weight loss, behavioral and psychological symptoms of dementia, medication (polypharmacy and antipsychotics) and pain.

The following phase consisted of the development of definitions, the development of evidence-based measurements and the operationalization of the measurements and answer options for the five selected themes. According to the RAND/UCLA methodology, development was supported by a multiexpert consultation for each theme chosen [49]. In 2014, the results of this development and of the quality indicator definitions at the time were sent to the appropriate offices of the Swiss federal government, cantons, nursing home associations, professional societies and associations, insurance companies, assessment instruments and other experts for consultation [50]. The Steering Committee then selected the 6 quality indicators with the highest acceptance rate (based on the national consultation) and prepared them for national measurement. The six selected quality indicators cover 4 themes: physical restraint (2 quality indicators), weight loss, polypharmacy and pain (2 quality indicators). Criticism of the nonselected themes included doubts about their validity as quality indicators (e.g., antipsychotic use) or the fear that some themes, if used, could send misleading signals (e.g., behavioral and psychological symptoms of dementia). The topics chosen show similarities to the quality indicator set recommended at roughly the same time in Germany [51]. Furthermore, based on a literature review, the Steering Committee also recommended variables to be tested for exclusion criteria and risk adjustment. Switzerland's approach is to start with a limited set of well-evaluated quality indicators, then to strengthen and complete this set over time with additional quality indicators.

Appendix B

Table A1. Overview of the assessment instruments used in the pilot study.

Assessment Instrument [1]	Nursing Home Resident Assessment Instrument	Planification Informatisée des Soins Infirmiers Requis [2]	BewohnerInnen-Einstufungs-und Abrechnungssystem [3]
Abbreviation	RAI-NH	PLAISIR/PLEX	BESA
Distributor in Switzerland	Q-Sys	Eros	BESAcare
Language availability	German, French, Italian	French	German, French, Italian
QI variables integration	Updated version of the instrument	Additional module	Updated version of the instrument
Data collection by	Healthcare staff	Healthcare staff or external evaluators (choice of each NH)	Healthcare staff
Start of the data collection (month)	July 2016	July 2016	July 2016
Data export (month)	August 2017	February 2017	August 2017

[1] At the time of the study, only three assessment instruments were in use by Swiss nursing homes. Each nursing home has the right to choose a preferred assessment instrument. These are used to carry out routine resident data collection to help evaluate aspects of residents' needs and care (e.g., amount of care needed, cognitive functions, mobility). One such assessment must be carried out at each resident's admission, then at least once every 6 months. The data gathered serve as a basis for care planning and health insurance claims [27]. [2] Computerized planning of required nursing care. [3] Residents classification and billing system.

Int. J. Environ. Res. Public Health **2020**, *17*, 9249

13 of 15

Appendix C

Table A2. Unadjusted results of intraclass correlation 1 (ICC1), intraclass correlation 2 (ICC2) and rankability of the six quality indicators.

Theme	ICC1 (95% CI)	ICC2 (95% CI)	Rankability (ϱ)
Polypharmacy	0.055 (0.037–0.068)	0.898 (0.865–0.918)	0.120
Self-reported pain	0.119 (0.087–0.149)	0.953 (0.931–0.962)	0.437
Observed pain	0.147 (0.113–0.177)	0.963 (0.949–0.971)	0.575
Physical restraint, trunk fixation or seating that prevents the resident from rising	0.343 (0.235–0.405)	0.988 (0.980–0.991)	0.970
Physical restraint, bedrails	0.245 (0.197–0.286)	0.980 (0.973–0.983)	0.783
Weight loss	0.135 (0.095–0.165)	0.959 (0.941–0.969)	0.715

Abbreviations: ICC1: intraclass correlation 1, CI: confidence interval, ICC2: intraclass correlation 2.

References

1. Organisation for Economic Co-operation and Development. *Improving Value in Health Care*; OECD Publishing: Paris, France, 2010.
2. Mainz, J. Defining and classifying clinical indicators for quality improvement. *Int. J. Qual. Health Care* **2003**, *15*, 523–530. [CrossRef] [PubMed]
3. Lilford, R.; MMohammed, A.; Spiegelhalter, D.; Thomson, R. Use and misuse of process and outcome data in managing performance of acute medical care: Avoiding institutional stigma. *Lancet* **2004**, *363*, 1147–1154. [CrossRef]
4. Anderson, J.; Hackman, M.; Burnich, J.; Gurgiolo, T.R. Determining hospital performance based on rank ordering: Is it appropriate? *Am. J. Med. Qual.* **2007**, *22*, 177–185. [CrossRef] [PubMed]
5. Mor, V. Defining and measuring quality outcomes in long-term care. *J. Am. Med. Dir. Assoc.* **2007**, *8* (Suppl. 2), e129–e137. [CrossRef]
6. Arling, G.; Kane, R.L.; Lewis, T.; Mueller, C. Future development of nursing home quality indicators. *Gerontologist* **2005**, *45*, 147–156. [CrossRef]
7. Van Dishoeck, A.M.; CLooman, W.N.; Lier, E.C.M.; Mackenbach, J.P.; Steyerberg, E.W. Displaying random variation in comparing hospital performance. *BMJ Qual. Saf.* **2011**, *20*, 651–657. [CrossRef]
8. Normand, S.-L.T.; Glickman, M.E.; Gatsonis, C.A. Statistical Methods for Profiling Providers of Medical Care: Issues and Applications. *J. Am. Stat. Assoc.* **1997**, *92*, 803–814. [CrossRef]
9. Normand, S.-L.; Shahian, T.D.M. Statistical and clinical aspects of hospital outcomes profiling. *Stat. Sci.* **2007**, *22*, 206–226. [CrossRef]
10. Burke, R.E.; Werner, R.M. Quality measurement and nursing homes: Measuring what matters. *BMJ Qual. Saf.* **2019**, *28*, 520–523. [CrossRef]
11. Du Moulin, M.F.; van Haastregt, J.C.; Hamers, J.P. Monitoring quality of care in nursing homes and making information available for the general public: State of the art. *Patient Educ. Couns.* **2010**, *78*, 288–296. [CrossRef]
12. Ibrahim, J.E.; Chadwick, L.; MacPhail, A.; McAuliffe, L.; Koch, S.; Wells, Y. Use of quality indicators in nursing homes in Victoria, Australia: A cross-sectional descriptive survey. *J. Aging Health* **2014**, *26*, 824–840. [CrossRef] [PubMed]
13. Zimmerman, D.R. Improving nursing home quality of care through outcomes data: The MDS quality indicators. *Int. J. Geriatr. Psychiatry* **2003**, *18*, 250–257. [CrossRef] [PubMed]
14. Canadian Institute for Health Information. *When a Nursing Home Is Home: How Do Canadian Nursing Homes Measure up on Quality?* CIHI: Ottawa, ON, Canada, 2013.
15. VLAAMS Indicatoren Project. Vlaams Indicatorenproject Woonzorgcentra: Kwaliteitsindicatoren 2019 Deel 1 [Flemish Indicators Project for Residential Care Centres: Quality Indicators 2019 Part 1]. 2020. Available online: https://www.zorgkwaliteit.be/sites/default/files/imports/wzc_20192.pdf (accessed on 19 April 2020).

16. Article 59a of the Swiss Federal Act of 18 March 1994 on Health Assurance (KVG/LAmal; RS 832.10). 2017. Available online: https://www.admin.ch/opc/fr/classified-compilation/19940073/index.html (accessed on 14 November 2020).

17. Zúñiga, F.; Blatter, C.; Wicki, R.; Simon, M. Fragebogenerhebung zur Zuverlässigkeit der Daten und Nützlichkeit aus Sicht der Anwender [National quality indicators in Swiss nursing homes: Questionnaire survey on data reliability and users' view on the usefulness]. *Zeitschrift fur Gerontologie und Geriatrie* **2019**, *52*, 730–736.

18. Bundesamt für Statistik. *Bundesamt für Statistik, Schnittstellenkonzept Medizinische Qualitätsindikatoren Pflegeheime, Version 1.3 [Interface Concept of the Medical Quality Indicators in Nursing Homes, Version 1.3]*; University of Basel: Basel, Switzerland, 2019.

19. Snijders, T.A.; Bosker, R.J. *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*, 2nd ed.; SAGE: Los Angeles, CA, USA; London, UK, 2012.

20. Austin, P.C.; Ceyisakar, I.E.; Steyerberg, E.W.; Lingsma, H.F.; de Mheen, P.J.M. Ranking hospital performance based on individual indicators: Can we increase reliability by creating composite indicators? *BMC Med. Res. Methodol.* **2019**, *19*, 131. [CrossRef] [PubMed]

21. Lingsma, H.F.; Eijkemans, M.J.; Steyerberg, E.W. Incorporating natural variation into IVF clinic league tables: The Expected Rank. *BMC Med. Res. Methodol.* **2009**, *9*, 53. [CrossRef]

22. Bliese, P.D. Within-group agreement, non-independence, and reliability: Implications for data aggregation and analysis. In *Multilevel Theory, Research, and Methods in Organizations: FOUNDATIONS, Extensions, and New Directions*; Jossey-Bass: San Francisco, CA, USA, 2000; pp. 349–381.

23. Bundesamt für Statistik. *Statistik der Sozialmedizinischen Institutionen 2017-Definitive Standardtabellen [Statistics of Socio-Medical Institutions 2017-Definitive Standard Tables]*; Bundesamt für Statistik: Neuchâtel, Switzerland, 2017.

24. Anderson, R.L.; Buckwalter, K.C.; Buchanan, J.L.; Maas, M.L.; Imhof, S.L. Validity and reliability of the Minimum Data Set Depression Rating Scale (MDSDRS) for older adults in nursing homes. *Age Ageing* **2003**, *32*, 435–438. [CrossRef]

25. Paquay, L.; de Lepeleire, J.; Schoenmakers, B.; Ylieff, M.; Buntinx, O.F.F. Comparison of the diagnostic accuracy of the Cognitive Performance Scale (Minimum Data Set) and the Mini-Mental State Exam for the detection of cognitive impairment in nursing home residents. *Int. J. Geriatr. Psychiatry* **2007**, *22*, 286–293. [CrossRef]

26. Jokanovic, N.; Tan, E.C.; Dooley, M.J.; Kirkpatrick, C.M.; Bell, J.S. Prevalence and factors associated with polypharmacy in long-term care facilities: A systematic review. *J. Am. Med. Dir. Assoc.* **2015**, *16*, 535e1. [CrossRef]

27. Vettori, A.; von Stokar, T.; Petry, C.; Britt, D.; Saxer, H.G.S. *Mindestanforderungen für Pflegebedarfserfassungssysteme (Schlussbericht) [Minimum Requirements for Care Needs Assessment Systems]*; Final Report; Bundesamt für Gesundheit (BAG): Neuchâtel, Switzerland, 2017.

28. Van Dishoeck, A.M.; Lingsma, H.F.; Mackenbach, J.P.; Steyerberg, E.W. Random variation and rankability of hospitals using outcome indicators. *BMJ Qual. Saf.* **2011**, *20*, 869–874. [CrossRef]

29. Nakagawa, S.; Schielzeth, H. Repeatability for Gaussian and non-Gaussian data: A practical guide for biologists. *Biol. Rev. Camb. Philos. Soc.* **2010**, *85*, 935–956. [CrossRef]

30. LeBreton, J.J.; Senter, J.L. Answers to 20 Questions About Interrater Reliability and Interrater Agreement. *Organ. Res. Methods* **2008**, *11*, 815–852. [CrossRef]

31. Stoffel, M.A.; Nakagawa, S.; Schielzeth, H. An Introduction to Repeatability Estimation with rptR. 2013. Available online: https://cran.r-project.org/web/packages/rptR/vignettes/rptR.html) (accessed on 19 April 2020).

32. Stoffel, M.A.; Nakagawa, S.; Goslee, H.S.S. rptR: Repeatability estimation and variance decomposition by generalized linear mixed-effects models. *Methods Ecol. Evol.* **2017**, *8*, 1639–1644. [CrossRef]

33. Steyerberg, E.W. *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating*; Springer: New York, NY, USA, 2009.

34. R Development Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2019.

35. Vos, E.L.; Lingsma, H.F.; Jager, A.; Schreuder, K.; Spronk, P.; Peeters, M.V.; Koppert, S.S.L.B. Effect of Case-Mix and Random Variation on Breast Cancer Care Quality Indicators and Their Rankability. *Value Health* **2020**, *23*, 1191–1199. [CrossRef] [PubMed]

36. Rantz, M.J.; Hicks, L.; Petroski, G.F.; Madsen, R.W.; Mehr, D.R.; Conn, V.M.; Zwygart-Staffacher, M.; Maas, M. Stability and sensitivity of nursing home quality indicators. *J. Gerontol. A Biol. Sci. Med. Sci.* **2004**, *59*, 79–82. [CrossRef] [PubMed]

37. UBC GmbH. Modellhafte pilotierung von indikatoren in der stationären pflege (MoPIP) (SV14-9015). In *Abschlussbericht zum Forschungsprojekt [Model-Based Piloting of Indicators in Inpatient care (MoPIP) (SV14-9015)]*; Final Report on the Research Project; Universität Bremen: Bremen, Germany, 2017.

38. Morris, J.N.; Moore, T.; Jones, R.; Mor, V.; Angelelli, J.; Berg, K.; Hale, C.; Morris, S.; Rennison, K.M.M.M. *Validation of Long-Term and Post-Acute Care Quality Indicators*; Final Report; Brown University: Rockville, MD, USA; Abt Associates Inc.: Cambridge, MA, USA, 2003.

39. Bates-Jensen, B.M.; Simmons, S.F.; Alessi, J.F.S.C. Evaluating the accuracy of minimum data set bed-mobility ratings against independent performance assessments: Systematic error and directions for improvement. *Gerontologist* **2005**, *45*, 731–738. [CrossRef]

40. Bours, G.J.; Halfens, R.J.; Haalboom, M.L.J.R. The development of a national registration form to measure the prevalence of pressure ulcers in The Netherlands. *Ostomy Wound Manag.* **1999**, *45*, 28–33.

41. Mor, V. Improving the quality of long-term care with better information. *Milbank Q.* **2005**, *83*, 333–364. [CrossRef]

42. Fung, V.; Schmittdiel, J.A.; Fireman, B.; Meer, A.; Thomas, S.; Smider, N.; Selby, J.H.J.V. Meaningful variation in performance: A systematic literature review. *Med. Care* **2010**, *48*, 140–148. [CrossRef]

43. Berwick, D.M.; James, B.; Coye, M.J. Connections between quality measurement and improvement. *Med. Care* **2003**, *41*, I:30–I:38. [CrossRef]

44. Shieh, G. Choosing the best index for the average score intraclass correlation coefficient. *Behav. Res. Methods* **2016**, *48*, 994–1003. [CrossRef]

45. Berner Fachhochschule. *Nachweis der Qualitäts- und Leistungsfähigkeit für die Stationären Leistungserbringer in der Langzeitpflege [Proof of Quality and Efficiency for Inpatient Service Providers in long-Term Care]*; Berner Fachhochschule: Bern, Switzerland, 2010.

46. Wingenfeld, K.; Kleina, T.; Franz, S.; Engel, D.; Mehlan, S.; Engel, H. *Entwicklung und Erprobung von Instrumenten zur Beurteilung der Ergebnisqualität in der Stationären Altenhilfe [Development and Testing of Instruments for Assessing the Quality of Outcomes in Inpatient Care for the Elderly]*; Bundesministerium für Gesundheit: Bielefeld, Germany; Köln, Germany, 2011.

47. National Quality Forum. *National Voluntary Consensus Standards for Nursing Homes: A Consensus Report*; National Quality Forum: Washington, DC, USA, 2011.

48. Leichsenring, K. Ergebnisorientierte Qualitätsindikatoren für Alters-und Pflegeheime [Outcome-oriented quality indicators for care and nursing homes]. *Soziale Sicherheit CHSS* **2012**, *19*, 26–31.

49. Fitch, K.; Bernstein, S.J.; Aguilar, M.D.; Burnand, B.; LaCalle, J.R.; Lazaro, P.; van het Loo, M.; McDonnell, J.; Vader, J.; Kahan, J.P. *The RAND/UCLA Appropriateness Method User's Manual*; RAND Corporation: Santa Monica, CA, USA, 2001.

50. Zúñiga, F.; Schneider, P.; Gehrlach, C.T. *Grolimund. Medizinische Qualitäts-Indikatoren. Schlussbericht zu Handen von Curaviva Schweiz [Medical Quality Indicators]*; Final Report at the Attention of CURAVIVA Switzerland; University of Basel: Basel, Switzerland, 2014.

51. Wingenfeld, K.; Stegbauer, C.; Willms, G.; Woitzik, C.V.R. *Entwicklung der Instrumente und Verfahren für Qualitätsprüfungen nach §§ 114ff. SGB XI und die Qualitätsdarstellung nach § 115 Abs. 1a SGB XI in der Stationären Pflege. aQua, Bielefeld (Abschlussbericht: Darstellung der Konzeptionen für das neue Prüfverfahren und die Qualitätsdarstellung) [Development of Instruments and Procedures for Quality Inspections According to §§ 114 ff. SGB XI and Quality Presentation According to § 115 para. 1a SGB XI in Inpatient Nursing care. aQua, Bielefeld]*; Final report: Presentation of the Concepts for the new Inspection Procedures and Quality Presentation; Universität Bielefeld: Bielefeld, Germany, 2018.