Predictability of clinical wear by laboratory wear methods for the evaluation of dental restorative materials
Heintze, Siegward Dietmar

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date: 2010

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):
Chapter 7

Correlation of Wear in vivo and Six Laboratory Wear Methods
Abstract

Objective: We examined the correlation between clinical wear rates of restorative materials and enamel (TRAC Research Foundation, Provo, USA) and the results of six laboratory test methods (ACTA, Alabama (generalized, localized), Ivoclar (vertical, volumetric), Munich, OHSU (abrasion, attrition), Zurich).

Materials and methods: Individual clinical wear data were available from clinical trials that were conducted by TRAC Research Foundation (formerly CRA) together with general practitioners. For each of the n=28 materials (21 composite resins for intracoronal restorations [20 direct and 1 indirect], 5 resin materials for crowns, 1 amalgam, enamel) a minimum of 30 restorations had been placed in posterior teeth, mainly molars. The recall intervals were up to 5 years with the majority of materials (n=27) being monitored, however, only for up to 2 years. For the laboratory data, the databases MEDLINE and IADR abstracts were searched for wear data on materials which were also clinically tested by TRAC Research Foundation. Only those data for which the same test parameters (e.g. number of cycles, loading force, type of antagonist) had been published were included in the study. A different quantity of data was available for each laboratory method: Ivoclar (n=22), Zurich (n=20), Alabama (n=17), OHSU and ACTA (n=12), Munich (n=7). The clinical results were summed up in an index and a linear mixed model was fitted to the log wear measurements including the following factors: material, time (0.5, 1, 2 and 3 years), tooth (premolar/molar) and gender (male/female) as fixed effects, and patient as random effect. Relative ranks were created for each material and method; the same was performed with the clinical results.

Results: The mean age of the subjects was 40 (±12) years. The materials had been mostly applied in molars (81%) and 95% of the intracoronal restorations were Class II restorations. The mean number of individual wear data per material was 25 (range 14-42). The mean coefficient of variation of clinical wear data was 53%. The global correlation Spearman r \textit{in vitro} versus \textit{in vivo} was 0.61. The only significant correlation was reached by OHSU (abrasion) with a Spearman r of 0.86 (p=0.001). Zurich, ACTA, Alabama generalized wear and Ivoclar (volume) had correlation coefficients between 0.3 and 0.4. For Zurich, Alabama generalized wear and Munich, the correlation coefficient improved if only composites for direct use were taken into consideration; the improvement, however, is not clinically relevant. The combination of different laboratory methods did not significantly improve the correlation.

Significance: As the correlation between laboratory wear methods and clinical wear was only moderate, weak or non-existent, the clinical wear database of this study can be used to further improve wear devices and methods.
Introduction
In restorative dentistry, the wear resistance of restorative materials is mainly evaluated with laboratory methods that use an electric device that puts a counterpart object (e.g. stylus) into contact with the material that is to be tested. Some devices simulate chewing movements that occur in the mouth [1,2] and some methods use an abrasive medium [3-5]. The methods differ with regard to the applied force, the material type and movement of the counterpart, the number of cycles, etc. [1]. In 2001 the International Standard Organization ISO published a technical specification “Wear by two- and/or three-body contact” describing 8 laboratory methods without, however, giving further comments on them [6]. None of these methods are in fact validated and the devices used are often not qualified for that purpose [1]. Furthermore, discrepancies in wear rates of the same material using the same wear method and wear parameters were found.

For contemporary resin materials wear of posterior restorations seems to be no longer a clinical problem, also over a period of ten years or more [7,8] because it is restricted to contact areas and does not affect the entire restoration with complete loss of the anatomy as it was the case with the early composite resins of the 80ties. Wear can be measured on replicas but may not be detected during clinical inspection. Therefore, the issue of wear should not be over-estimated. However, wear is important if materials with new monomers and/or fillers are developed and their wear behaviour cannot be deduced from similar products. In this case, laboratory methods may help to asses the wear resistance before the material is evaluated in a clinical trial, whose results will be present only years after the beginning of the study.

Ideally, the laboratory wear rates of specific materials should reflect the wear that is measured in vivo with the same materials. Some methods claim to have a high correlation between in vitro and in vivo wear, but these correlations are only based on a limited number of restorations with mostly 2-3 materials [9-11]. In addition, they are linearly extrapolated from 6-month data [11], the clinical wear was measured with inadequate and inaccurate methods such as the Moffa-Lugassy Scale [3,4,10] or the correlation analysis was carried out with pooled clinical data from studies with different study protocols [12].

To date, no effort has been undertaken to systematically compare data of laboratory wear methods with clinical wear data. Such an approach has been hampered by the
fact that the peer-reviewed dental journals contain very few clinical wear data which were generated by applying a reliable and accurate quantitative wear method [9,13,14]. In the 90ties, the Technologies in Restorative and Caries (TRAC) Research Foundation (formerly Clinical Research Associates CRA) clinically evaluated many resin materials for both direct and indirect Class II restorations and for indirect crowns and fixed dental prostheses.

The purpose of the present study was to correlate these data with wear data published in the literature gained from the most common and current laboratory wear methods, which tested the same materials. Four of these six selected methods are also part of the ISO Technical Specification [6]. A round-robin test that involved five of the six wear methods examined in the present study revealed that the results from one wear method to another were hardly comparable [15]. The reason for that is that these laboratory wear methods follow different tribological concepts [1]. It can therefore be expected that the correlation to clinical wear is also different. As the clinical wear process is a complex mechanism with different tribological phenomena occurring at the same time, it can be assumed that there is no single laboratory wear method capable of showing a good correlation with clinical wear. However, if two or three methods that come very close to the clinical wear are combined, the correlation may be improved.

The following hypotheses were examined:

1. There is a method-dependent correlation between laboratory wear methods and the clinical wear rates.
2. If two or three wear methods are combined, the correlation can be improved.

Materials and methods

**Wear in vivo**

Clinical wear data were derived from clinical studies, which the TRAC Research Foundation (formerly CRA) in Provo, USA, organized and carried out together with general practitioners. Some of the results were previously reported at AADR/IADR conferences [16-21].

Wear was mainly analyzed in big molar restorations, sometimes including the build-up of a cusp. A minimum of 30 restorations was placed in 30 subjects. The drop-out
rate after 2-3 years was about 10%. For composite materials for indirect use, full-coverage crowns were fabricated in especially selected dental laboratories.

A non-surface-contact method was devised to determine the in vivo wear of dental restorations by quantifying the focal distance of a microscope (Olympus BH-2, Olympus Optical Co. Ltd., Japan) in sharp focus at 200x on the surface of an epoxy replica of the restored tooth [22,23] (Figure 1). Custom software provided the following parameters for each sample placed onto the test apparatus: establishment of a zero elevation from which all measurements were made, reference points defining the perimeter of the surfaces to be measured, precise movement of an x-y stage, determination of the point of sharpest focus, and acquisition of x-y-z data every 50 microns over the pre-defined occlusal table of the replica of the restored tooth. The resulting data matrices made after initial placement and any time subsequent were aligned first visually using conserved points present at both periods and then were subjected to best fit analyses using an algorithm. The difference between two matrices made at two different points in time was reported as volumetric and linear measurements. Visual display of colour coded data showed the exact locations of the differences in topography between the two epoxy replicas under comparison. Verification tests made on this system showed an accuracy of ±3 µm.

The mean vertical loss of the whole restoration surface was calculated after cutting the restoration from the surrounding tooth. Marginal overhangs of the material were not taken into the analysis.

Figure 1: Equipment for 3-D wear measurements. The whole system consisted of three measuring microscopes A, B, and C plus the accessory hardware for each. Microscope A had oculars to allow for a human to set up the dies on special blocks. Microscopes B and C had no oculars and were designed to receive the blocks and perform the x-y-z measurements, which ultimately resulted in the matrix that formed the occlusal map of each tooth surface.
Standards run with every sample validated the trouble-free, accurate operation of the system and electronic alarms signalled any malfunction. Although accuracy of the system was very good, the time required for data acquisition was poor (6-8 hours for one molar tooth, depending on overall size).

The individual data of each subject were available, including age, gender, type of tooth restored, number of restored surfaces, and wear after each recall interval.

**Wear in vitro**

The following laboratory wear methods were selected according to their significance and frequency of citation in the literature [1].

1. Academic Center for Dentistry Amsterdam (ACTA) machine and wear method developed by de Gee et al [5].
2. Alabama wear simulator and wear method by Leinfelder & Suzuki [4].
3. Ivoclar wear method developed at Ivoclar Vivadent using the Willytec wear simulator [1].
4. Munich wear method using the Munich Artificial Mouth simulator [24].
5. Oregon Health and Science University (OHSU) oral wear simulator and wear method by Condon & Ferracane [3].

The wear methods with their typical parameters are listed in Table 1.

Four variables were used in conjunction with the Zurich wear method: Phase 1 after 120,000 loading cycles, phase 2 after 240,000 cycles, phase 3 after 640,000 cycles and phase 4 after 1,200,000 cycles. These values were claimed to correspond to six months, one year, 2.7 years and 5 years of clinical wear, respectively [11]. For the Alabama method two variables had been included in the analysis: generalized and localized wear; for the OHSU method two different wear zones were generated: abrasion and attrition. For the Ivoclar method the vertical as well as the volumetric attrition was taken into consideration. A single variable was available for ACTA and Munich (vertical loss after completion of all cycles)
<table>
<thead>
<tr>
<th>Wear method</th>
<th>Force</th>
<th>Force actuator</th>
<th>Lateral movement</th>
<th>Stylus material</th>
<th>Abrasive medium</th>
<th>Number of cycles</th>
<th>Outcome variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACTA</td>
<td>15 N</td>
<td>Spring</td>
<td>wheel runs against specimens</td>
<td>metal wheel, slip rate 15%</td>
<td>Millet seed shells /rice</td>
<td>200,000</td>
<td>Vertical wear (µm)</td>
</tr>
<tr>
<td>Alabama</td>
<td>75 N</td>
<td>Spring</td>
<td>30° rotation no</td>
<td>polyacetal (Ø 2.5 mm) polyacetal (Ø 6.5 mm)</td>
<td>PMMA beads</td>
<td>400,000</td>
<td>Localized wear (µm) Generalized wear (µm)</td>
</tr>
<tr>
<td>Ivoclar</td>
<td>5 kg</td>
<td>Dead weight driven with stepping motor</td>
<td>0.7 mm</td>
<td>leucite reinforced ceramic (Empress) (Ø 2.4 mm)</td>
<td>water</td>
<td>120,000</td>
<td>Vertical wear (µm) Volumetric wear (10^7 µm^3)</td>
</tr>
<tr>
<td>Munich</td>
<td>5 kg</td>
<td>Dead weight driven with pneumatic cylinder</td>
<td>8 mm</td>
<td>Aluminium oxide ceramic (Degussit) (Ø 5 mm)</td>
<td>water</td>
<td>50,000</td>
<td>Vertical wear (µm)</td>
</tr>
<tr>
<td>OHSU</td>
<td>20 N</td>
<td>Abrasion 80 N attrition Electro-magnetic</td>
<td>6 mm</td>
<td>Enamel (Ø 10 mm) Poppy seeds/ PMMA beads</td>
<td>50,000</td>
<td>Mean vertical wear (µm) Abrasion Attrition</td>
<td></td>
</tr>
<tr>
<td>Zurich</td>
<td>49 N</td>
<td>Electro-magnetic</td>
<td>Sliding movement of specimen</td>
<td>Enamel (Ø 5 mm)</td>
<td>water</td>
<td>120,000 240,000 640,000 1,200,000</td>
<td>Vertical wear (µm)</td>
</tr>
</tbody>
</table>

Table 1: Six laboratory wear methods and their wear generating parameters.
A systematic search was carried out by using the databases of MEDLINE (PUBMED, since 1966) and IADR (abstracts, starting 1984). The search period was March 2008. The search terms were “wear” and/or “abrasion”, “in vitro” and the name or acronym of the respective wear method. The wear data were only used if the identical wear method parameters listed in Table 1 were applied. As far as enamel wear is concerned, the test parameters were different to those applied for composite or amalgam materials for the Alabama and OHSU method. The studies included in the comparative analysis are listed in Table 2.

<table>
<thead>
<tr>
<th>ACTA</th>
<th>Alabama</th>
<th>Ivoclar</th>
<th>Munich</th>
<th>OHSU</th>
<th>Zurich</th>
</tr>
</thead>
<tbody>
<tr>
<td>[10,15,35-37]</td>
<td>[4,9,38-45]</td>
<td>[15,46]</td>
<td>[47]</td>
<td>[3,48-54]</td>
<td>[15,55,56]</td>
</tr>
</tbody>
</table>

Table 2: References from where the wear data were retrieved for the in vitro/in vivo comparison.

Although the test parameters are identical, the wear data may be influenced by the fact that different research institutes carried out the tests. Therefore, the studies were grouped into two groups: PUB 1=from the same institute, namely the ones who first published the method; PUB 2=other institutes.

The materials were classified into 4 categories:
CAT 1: composites for intra-coronal restorations (direct and indirect)
CAT 2: composites for crowns and bridges
CAT 3: amalgam
CAT 4: enamel.

**Statistical analysis**

In order to summarize the information of clinical measurements, a single index was created for each material. Using an average value of the measurements for each material, however, would not provide a fair comparison between the materials because of the unbalanced design and the increase of wear over time. Instead, a linear mixed model with a fixed material effect, a fixed time effect and a random patient effect was considered as follows:

\[
Y_{ijk} = m_i + t_j + p_k + \epsilon_{ijk}
\]

where \(Y_{ijk}\) denotes the logarithm of in vivo measurement of material \(i\) at time \(j\) for patient \(k\), \(m_i\) is the effect of material \(i\), \(t_j\) is the effect of time \(j\), \(p_k\) is the effect of patient \(k\) and \(\epsilon_{ijk}\) is an error term; \(p_k\) and \(\epsilon_{ijk}\) were assumed to be independent and normally distributed. The age of the patients could not be considered, because this information was missing for about
43% of the patients. A logarithm transformation was necessary to reach normality. Calculating \( \exp(\text{effect})-1 \) provided us with an estimate of the percentage of median wear increase on the original scale. Estimates of material effects \( m \), were then used as a summarizing \textit{in vivo} index, to be correlated with the various \textit{in vitro} variables. Further covariates such as gender of the patients or tooth type were also included in the model. Linear mixed models were calculated using the lme routine implemented in the package nlme (version 3.1-83) from the free statistical package R.

In order to be closer to a normal distribution, \textit{in vitro} variables were log-transformed, as suggested in a previous publication where data of different laboratory methods were compared [15]. To facilitate comparison, the log-values \( x \) of each variable were further transformed into relative ranks, calculated as \( 1+(n-1)(x-m)/(M-m) \), where \( n \) was the number of materials considered (\( n=28 \)), \( m \) was the minimum (mean) and \( M \) was the maximum \textit{in vitro} value (mean). The scale of each variable was hence transformed into a “rank scale” from 1 to \( K \) (with all variables being directly comparable), while the relative differences between the materials was taken into account. Relative ranks were also useful for defining an \textit{in vitro} index summarizing \textit{in vitro} performance by considering an average of the available relative ranks (noted AARR) over the different \textit{in vitro} variables. Note that this procedure could have been problematic in the present context since it was difficult to obtain a reliable estimate of the values \( m \) and \( M \) of the best and worst material for each variable (which would have been observed if measurements had been available for all 28 materials) because various materials were missing. However, we tried various techniques to impute values to the missing materials and this did not change the correlations between clinical and laboratory wear data.

To evaluate whether the combination of different laboratory methods improves the correlation with \textit{in vivo} data, stepwise linear regression modelling was applied (SPSS release 16.0).

\textbf{Results}

The individual clinical wear data of 28 materials (including enamel) for which also \textit{in vitro} data of one or more laboratory wear methods were present were included in the analysis. The mean age of the subjects for whom age data were available (\( n=835 \)) was 40 (±12) years and there were almost on average twice as many females as males in each material study (944 females, 540 males). The materials were mostly applied in molars (1199 molar and 285 premolar restorations). Almost the same number of restorations was present in the maxilla and the mandible (766 versus 718). Furthermore, 96% of the intracoronal
restorations were Class II restorations and only about 4% were Class I restorations; 2-surface restorations were more frequent than 3- or 4-surface restorations (567 2-surface, 358 3-surface, 212 4-surface); replicas of 253 full-coverage crowns were available for analysis. The mean number of individual wear data per material was 25 (range 14-42) per material group. The mean coefficient of variation of wear data (1 year) was 53.7% (±14.8). The number of materials for which in vitro data were available was 22 for Ivoclar, 20 for Zurich, 17/14 for Alabama generalized/localized wear, 12/13 for OHSU abrasion/attrition, 12 for ACTA and 7 for Munich. In vivo wear data were available after six months (19 materials), one year (28 materials), two years (27 materials) and three years (9 materials). We did not consider in vivo wear data after more than three years because they were missing for almost all materials. Among the 28 materials, 21 belonged to CAT 1 (intracoronal composites), 5 to CAT 2 (composites for full coverage crowns: Artglass, belleGlass, Sculpture, Solidex, Targis), 1 to CAT 3 (amalgam: Dispersalloy) and 1 to CAT 4 (enamel).

**Analysis of in vivo measurements**

The average in vivo wear measurements along a time axis are shown in the upper panel of Figure 2, each line representing one material. It is noticeable that both the level of wear and the between-material variability were monotonically increasing over time. The lower panel of Figure 3 shows the average in vivo wear measurements calculated on the log-scale, where the between-material variability had been stabilized, which facilitated data modeling. Note in particular the nearly parallel increase of wear between the materials, an indication that the effect of time was the same for each material, and that a ranking of materials would be approximately the same at each point in time.
A linear mixed model was fitted to the log wear measurements, using the following factors: material (28 levels), time (4 levels: 0.5, 1, 2 and 3 year), tooth (2 levels: premolar and molar) and gender (2 levels) as fixed effects and patient as random-effect. The effect of time was of course strongly significant (p<0.0001), with the percentage of median wear increase being estimated to \( \exp(0.62) - 1 = 86\% \) between six months and one year, to \( \exp(0.49) - 1 = 63\% \) between one year and two years, and \( \exp(0.37) - 1=45\% \) between two and three years. Overall, the factor ‘material’ was significant (p<0.0001). In line with Figure 2, the best material with the lowest wear was Dispersalloy, followed by Heliomolar, Z250 and SureFil, whereas the worst material was Nulite F, followed by Targis, Prodigy and Estilux. To summarize the \textit{in vivo} performance of the different materials, we calculated relative ranks from the estimated coefficients of the factor ‘material’ in the linear mixed model, which are provided in the last column of Table 1.
### Correlation between in vitro and in vivo wear

Relative ranks for 9 wear variables are given in Table 3.

| Material               | in vitro | in vitro |
|------------------------|----------|----------|-----------|
|                        | 1925.4   | 25.7     | 25.1      |
|                        | 5.1      | 15.4     | 20.3      |
|                        | 24.4     | 13.8     | 19.8      |
|                        | 17.4     | 20.9     | 14.5      |
|                        | 1.1      | 14.7     | 6.8       |
|                        | 26.6     | 15.5     | 13.1      |
|                        | 28.0     | 17.2     | 6.6       |
|                        | 25.5     | 17.8     | 29.9      |
|                        | 25.0     | 23.4     | 26.3      |
|                        | 20.9     | 24.7     | 17.5      |
|                        | 16.4     | 20.4     | 25.8      |
|                        | 10.9     | 16.3     | 16.5      |
|                        | 15.4     | 18.4     | 15.4      |
|                        | 17.2     | 20.4     | 20.2      |
|                        | 1.1      | 10.6     | 7.4       |
|                        | 13.4     | 27.4     | 24.2      |
|                        | 24.5     | 25.4     | 24.2      |
|                        | 23.7     | 17.9     | 28.0      |
|                        | 16.6     | 18.3     | 23.3      |
|                        | 14.5     | 13       | 23.8      |
|                        | 20.7     | 5.5      | 8.6       |
|                        | 2.9      | 1.4      | 20.5      |
|                        | 6.9      | 12.2     | 10.7      |
|                        | 24.5     | 12.5     | 10       |
|                        | 14.9     | 16.8     | 5.1       |
|                        | 3.8      | 28       | 7.5       |

Table 3: Relative ranks (RR) characterizing in vitro performance for 28 materials according to six methods (9 variables). Average of available relative ranks (AARR) summarizing in vitro performance over six methods is also provided, as well as relative ranks characterizing in vivo performance calculated via a linear mixed model. The first 21 materials are direct composite materials (CAT 1).

For the Zurich method, we considered an average of the different Zurich time variables (corresponding to 0.5-year-, 1-year-, 2.7-year- and 5-year equivalents). Scatterplots of these relative ranks against the relative ranks of in vivo performance are shown in Figure 3.
Figure 3: Relative ranks characterizing in vitro performance according to 6 methods (9 variables) versus relative ranks characterizing in vivo performance calculated via a linear mixed model, together with Spearman correlation r and associated p-value (in parentheses, restricted to materials of CAT 1). Each dot represents one material. Composites for intracoronal restorations, composite crowns, amalgam and enamel are represented in red, blue, green and black, respectively.

Spearman correlations between in vitro and in vivo performance, together with associated p-values, are provided. The correlation was negative for the Alabama localized wear method, and was between 0.29 and 0.44 for the ACTA method, the Alabama generalized wear method, the Ivoclar method (vertical and volumetric) as well as for the Zurich and Munich method, with the latter being calculated from only 7 materials. On the other hand, the correlation was significantly positive for the OHSU method, with the highest correlation being found for OHSU abrasion (r=0.86, p=0.001). For the Zurich and Alabama generalized methods, the coefficient significantly improved if only composites for intracoronal Class II restorations were included versus all materials (enamel, amalgam, and composites for full-coverage crowns) (see Figure 3). Also, the correlation slightly
increased for the Alabama localized and generalized wear methods when only the results for materials published by the same group were included (from –0.14 to 0.09 and from 0.40 to 0.49), while the correlation slightly decreased for the ACTA method (from 0.43 to 0.36). For the Zurich variables, the correlation slightly decreased over time (r=0.49 at 0.5 year- and at 1-year equivalent, r=0.41 at 2.7 year- and r=0.37 at 5-year equivalent).

![Figure 4: Average of available relative ranks (AARR) summarizing in vitro performance over six methods versus relative ranks characterizing in vivo performance calculated via a linear mixed model, together with Spearman correlation r and associated p-value (in parentheses, restricted to materials of CAT 1). Each dot represents one material. (Materials of) Composites for intracoronal restorations, composite crowns, amalgam and enamel are represented in red, blue, green and black, respectively. Ranking of materials with respect to in vitro measurements

To summarize the in vitro performance, we calculated the average of the available relative ranks (AARR) calculated over six in vitro variables, one for each method (ACTA, Alabama generalized wear, Ivoclar vertical attrition, Munich, OHSU abrasion and Zurich average). Values of AARR are provided in Table 1. The best overall in vitro performance was achieved by enamel (AARR of 7.4) followed by Dispersalloy (AARR of 8.5) and SureFil...
(AARR of 8.5). The worst in vitro performance was achieved by Nulite F and Prisma AP.H, although both materials were measured by a single method only.

The AARR in vitro index was useful to investigate further questions in an economical way. The global in vitro performance of materials of CAT 1 and CAT 2 were, for example, compared and this comparison showed that the average AARR was 19.3 for the former and 14.0 for the latter. In a Mann-Whitney test, this difference was almost significant (p=0.05). The Spearman correlation coefficient between clinical and overall laboratory wear was 0.61 (p=0.0008) (see Figure 4).

When different laboratory methods were combined with each other, the correlation between in vitro and in vivo relative ranks of the materials was not significantly improved, which was shown by the stepwise regression analysis (Table 4).

<table>
<thead>
<tr>
<th>Method</th>
<th>Adjusted R^2</th>
<th>Number of materials</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>OHSU abrasion</td>
<td>0.68</td>
<td>12</td>
<td>0.001</td>
</tr>
<tr>
<td>Alabama generalized</td>
<td>0.42</td>
<td>17</td>
<td>0.005</td>
</tr>
<tr>
<td>Ivoclar</td>
<td>0.31</td>
<td>22</td>
<td>0.007</td>
</tr>
<tr>
<td>Munich</td>
<td>0.26</td>
<td>7</td>
<td>0.24</td>
</tr>
<tr>
<td>ACTA</td>
<td>0.14</td>
<td>12</td>
<td>0.22</td>
</tr>
<tr>
<td>Alabama localized</td>
<td>0.01</td>
<td>14</td>
<td>0.76</td>
</tr>
<tr>
<td>OHSU attrition</td>
<td>0.36</td>
<td>13</td>
<td>0.03</td>
</tr>
<tr>
<td>Ivoclar + Alabama + OHSU</td>
<td>0.79</td>
<td>7</td>
<td>0.15</td>
</tr>
<tr>
<td>Ivoclar + OHSU ab</td>
<td>0.68</td>
<td>9</td>
<td>0.03</td>
</tr>
<tr>
<td>Ivoclar + Alabama generalized</td>
<td>0.42</td>
<td>14</td>
<td>0.05</td>
</tr>
<tr>
<td>Ivoclar + Zurich</td>
<td>0.43</td>
<td>17</td>
<td>0.02</td>
</tr>
<tr>
<td>OHSU abrasion + Alabama</td>
<td>0.72</td>
<td>8</td>
<td>0.04</td>
</tr>
</tbody>
</table>

Table 4: Stepwise regression analysis of different combinations of wear methods with adjusted R^2 and p-value.

**Discussion**

This is the first study which systematically evaluated the relationship between laboratory data on wear and clinical wear rates. In the past, some efforts were made to correlate in vitro and in vivo data, but those studies were marred by a number of shortcomings related to an inadequate methodology of in vivo wear measurements, inadequate statistical methods and/or a limited number of materials [4,9-11]. From the early 70ties to the present date, the clinical wear of restorative materials was subjectively evaluated directly on the
patient together with other clinical parameters using (modified) USPHS criteria. USPHS criteria, however, are by far not an adequate tool to measure wear [26,27]. In the 80ties and early 90ties, clinical wear used to be quantified by comparing cast replicas with a set of standards, known as scales. Two evaluators would compare the replica with the standards by means of loupes and assign a wear value. The scales were based on the concept that material loss at the restoration margin was indicative of the loss of material over the entire restoration surface. Mainly three different scales were propagated at that time. (1) The Leinfelder scale [28], which used 6 calibrated die stone standards from clinical restorations, exhibiting approximately 100 μm to 500 μm of occlusal loss; (2) the Moffa-Lugassy (M-L) scale [29], which used 18 standard dies with cylindrical incremental defects ranging from 25 μm to 1000 μm; (3) the Vivadent scale (modification of M-L scale developed by V. Rheinberger, Ivoclar Vivadent, Schaan), which used tooth-sized dies with restoration-like incremental defects. All scales involved problems with internal and external validity. By applying the standard scales, the agreement among different evaluators can vary tremendously, especially amongst inexperienced evaluators [30]; the Vivadent scale, however, consistently achieved the highest level of agreement. Yet, even the Vivadent scale was shown to lacking in accuracy, as it was proven that the actual wear is systematically underestimated when the results obtained with the Vivadent scale were compared with those obtained with sophisticated laser equipment [31].

By using the Leinfelder scale, the correlation coefficient Spearman r between the clinical wear results and those generated with the Alabama method was 0.81 as calculated by the author based on original data [4]. High correlation coefficients (>0.8) were also reported for the OHSU method in conjunction with clinical studies that used different methods to quantify the wear in vivo [3]; However, the correlations were based on only 2 to 3 resin materials.

For the ACTA method, in vitro/in vivo comparisons were carried out using the clinical data of 7 different studies, which used different methods to quantify wear at different time intervals; the correlation coefficients were between 0.81 and 0.99 [10]. However, the number of materials was again only between 2 to 4. Performing a comparative analysis based on such a low number of materials or even on as few as two, is meaningless. A similar attempt using ACTA results was made by another study, which reported correlation coefficients between 0.88 and 0.92 when the results of 7 composite resins were compared [12]. In this study, the laboratory results were related to pooled clinical results, which came from different studies.
In another comparative study, no correlation was found between the results obtained with a laboratory method using a slider-on-disc wear apparatus on 7 composite resins materials and an amalgam and the results obtained with the same materials in 6 posterior restorations per material and following them up for 6 months *in vivo* [32]. The *in vivo* replicas were measured with a modified profilometer, which measured the material loss in the central fossa by defining profiles relative to reference points drilled in the cusps [33].

In the present study, the clinical wear was determined by a single research institute only, following always the same protocol for the placement of the restorations and the wear quantification on replicas. It can therefore be assumed that the wear rate of the different materials can be compared to a certain degree. The wear was determined on replicas with a method that uses a computerized microscope with narrow focus adjustment. The accuracy of the method is adequate for the research purpose.

It cannot be excluded that the wear rates might have been different for some materials, if the material had been tested in another population group. In the present study, generally one material was placed per test person. A common method to exclude the patient factor is to place restorations of different materials in the same patient in what is known as the split-mouth design. However, the number of materials that can be inserted is limited to 2 to 4. For the linear mixed model variables, such as the number of wear values per material, gender and age of the test person were included.

The variability of the clinical wear results was very high; the mean coefficient of variation was 53%. This means that the test person himself or herself plays a major role in the clinical wear performance of a material. However, differences in the anatomical shape of the restorations, which may vary from one test person to another or from one operator to another, can also render a restoration more prone to occlusal wear than another restoration. For 17 of the 28 materials, the mean clinical wear after two years was in a relatively narrow range between 70 µm and 130 µm (see Figure 2); due to the high variability of clinical results, these materials are statistically not significantly different, if ANOVA with a post-hoc test to adjust for multiple comparisons is applied.

The high correlation coefficient published in the literature for certain methods was only confirmed for the OHSU method, not, however for the Alabama and ACTA methods. A possible reason for this is that those studies included only few materials, which may not be representative of the whole diversity of the materials considered in the present study. In general, in the present study the correlation between the different laboratory methods was weak or even non-existent for most wear methods, except for the OHSU method, for which the Spearman r coefficient was 0.86 for abrasion and 0.66 for attrition. Some of these
correlations increased and some decreased when only composites for intracoronal Class II restorations versus all materials (enamel, amalgam, and composites for indirect use) were tested. The improvement, however, is not clinically relevant. Furthermore, it has to be taken into account that the selection of materials was not identical in all comparisons of laboratory methods and clinical wear data for the comparison of the laboratory methods with the clinical wear data. If one e.g. includes for the Ivoclar method only those materials that were also included for the OHSU abrasion method, the correlation coefficient improves for the Ivoclar method. On the other hand, if one includes only those materials that were evaluated with all laboratory methods, one ends up with only 2 materials (Heliomolar and Z100), which were evaluated by all methods.

Although there are some similarities between the OHSU and the Alabama method, such as the use of an abrasive slurry, the relatively low force between the specimen and the stylus and the big diameter of the stylus (OHSU 10 mm, Alabama 6.5 mm), only the OHSU method showed a strong correlation with the clinical wear index. For the clinical wear the entire occlusal surface is measured including attrition zones (occlusal contact areas, OCA) and contact-free areas (CFA). Therefore, the clinical wear rate represents the mean wear across the entire occlusal restoration and it is conceivable that the OHSU abrasion method comes close to what is measured in vivo.

When the wear of individual composite materials is considered, the microfilled material Heliomolar (light-curing) demonstrated the lowest clinical wear, followed by Z250 and the packable material SureFil. The worst material was Nulite F, which contains large fillers, followed by Prodigy, Targis; materials like Tetric Ceram, Herculite, and Pertac Hybrid were in between (see Figure 4). None of the materials reached the low wear rates of the amalgam Dispersalloy.

The laboratory methods showed quite different results when they were compared to the clinical wear. Heliomolar produced high wear rates in conjunction with the ACTA and Alabama methods, while it showed low rates when subjected to the Ivoclar, Munich and OHSU abrasion methods (see Table 3). When all laboratory methods were summed up, the lowest available relative ranks (AARR) index was 7.4 for enamel, followed by Dispersalloy (8.5) and SureFil (8.5), illustrating the fact that the various methods did not agree with each other. On the opposite side of the ranking, the highest AARR was obtained for Nulite F (28), Prisma AP.H (28), Adaptic 2 (25.1), Occlusion (24.9) and Tetric Ceram (24.7); the first three materials were, however, only tested by 1 or 2 wear methods. These materials contain fillers larger than 2 µm. For both the clinical trials and some laboratory tests, there was a tendency for microfilled composites to show lower wear rates.
compared to composites with large fillers. Large fillers can easily be removed and act as an abrasive medium, which further accelerates the wear of the material. However, there are materials such as Z100 and Tetric Ceram whose filler particles have a similar dimension but showed different wear results \textit{in vivo} and \textit{in vitro} for the Zurich method but not for the Ivoclar or ACTA method.

There are obviously many other material-related parameters that come into play if a manufacturer intends to develop a material with high wear resistance. First of all, the size of the fillers reflects only the mean particle size. In most materials, however, the distribution of the filler particles does not correspond to a normal distribution but rather to a distorted one. Therefore, the mean particle size does not give valid information about the amount of small, medium-sized and large particles. This, however, may affect the resins` resistance to wear-induced attrition. Other important factors are the shape of the filler particles, the distance between them, the composition of the resin matrix, the chemical link between inorganic fillers and the resin matrix and the conversion rate after polymerization [34].

Other factors worth considering are related to the wear methods themselves. Many of these methods involve devices that are not qualified for the purpose to generate wear in a standardized and reproducible way [1]. The force which the actuators exert on the specimens is not regulated and controlled; the same holds true for the distance at which the stylus slides over the specimen. For the methods that use an abrasive medium (ACTA, Alabama, OHSU), problems arise with regard to the homogenous distribution of particles between stylus specimen, thickening of the slurry because of the evaporation of water and the impossibility to standardize organic components like millet or poppy seeds. A systematic review produced evidence that many of these methods lack reproducibility, as the wear data of the same composite material tested by the same authors using the same wear method and the same wear parameters differed as much as 72% from one publication to another [1]. Therefore, the present study investigated the fact whether the correlation with \textit{in vivo} data differs if only data from the same research group are taken or if all available data that used the same wear method and parameters but are published by different research groups are analysed. There were, however, only slight influences and for three methods (Ivoclar, Munich, Zurich) all data came from the same research group or even the same publication.

The combination of the results of different methods did not significantly improve the correlation. Based on the limited amount of data available to us, we found no evidence that it would be worth to install different wear methods in the laboratory. Another question
arises: How should the results of a certain material be interpreted if they differ so largely from one method to the other?

One of the conclusions of the present study is that only one laboratory method showed a sufficient correlation with the clinical wear data. The results of all the other methods were only poorly correlated with the clinical data.

Even if laboratory wear methods do have limitations, they may be able to roughly categorize a new material as to whether it will probably exhibit a high or low wear rate. Furthermore, the between-subject variability of wear rates is very high and limits also the meaningfulness of clinical wear data, especially if the mean differences between different materials are small. However, there are materials with relatively low wear and those with high wear. Therefore, laboratory methods should be able to identify those materials and the method should be able to show low wear of a standard material that also had low wear in clinical trials, e.g. as is the case for amalgam. As the correlation between laboratory wear methods and clinical wear was only moderate, weak or non-existent, the clinical wear database of this study can be used to further improve wear devices and methods.

Summary and conclusions
- Estimates of random effects in a linear mixed model provided a fair summary of in vivo performance of materials, taking into account the unbalanced design, while using the whole information available. The average of available relative ranks over the various methods provided a useful summary of laboratory wear. These indices could be used to test hypotheses in an efficient way.
- The correlation between the various laboratory methods and the in vivo results was different for the various methods. As a whole, however, the correlation was significant only for one method (OHSU abrasion) and weak or non-existent for the other methods. Therefore, the first two hypotheses were partially confirmed. The global correlation in vitro versus in vivo was 0.61. The high correlations published for certain methods were not confirmed in the present study.
- For the Zurich, Alabama generalized wear and Munich methods, the correlation coefficient improved if only composites for direct use were taken into consideration; the improvement, however, is not clinically relevant.
- The combination of different laboratory methods did not significantly improve the correlation, thus rejecting the second hypothesis.
- There is still a need for clinical wear studies as composites with new monomers that reduce polymerization shrinkage or provide biological effects (antibacterial and/or remineralization enhancing effects) or composites with new fillers may have wear patterns that completely differ from the wear of the composites used thus far.

ACKNOWLEDGEMENTS
The authors are very grateful to PD Dr. Till U. Goehring, Zurich, for additional information about the results of the Zurich wear testing.

References


